

Information-theory based surrogate marker evaluation from several
randomized clinical trials with continuous true and binary surrogate endpoints

Peer-reviewed author version

ASSAM NKOUIBERT, Pryseley; TILAHUN ESHETE, Abel; ALONSO ABAD, Ariel &
MOLENBERGHS, Geert (2007) Information-theory based surrogate marker
evaluation from several randomized clinical trials with continuous true and binary
surrogate endpoints. In: CLINICAL TRIALS, 4(6). p. 587-597.

DOI: 10.1177/1740774507084979

Handle: <http://hdl.handle.net/1942/7966>

Information-theory Based Surrogate Marker Evaluation from Several Randomized Clinical Trials with Continuous True and Binary Surrogate Endpoints, Using R

Running Title: Surrogacy with Mixed Endpoints

**Assam Pryseley, MSc Abel Tilahun, MSc Ariel Alonso, PhD
Geert Molenberghs, PhD**

Hasselt University, Center for Statistics, Agoralaan 1, B3590 Diepenbeek, Belgium

The authors gratefully acknowledge the financial support from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

Abstract

Background. Surrogate endpoints potentially reduce the duration and/or increase the amount of information available in a study, thereby diminishing patient burden and cost. They may also increase the effectiveness and reliability of research, through beneficial impact on non-compliance and missingness.

Purpose and Methods. In this paper, we review the meta-analytic approach of Buyse *et al* (2000) and its extension to mixed continuous and binary endpoints by Molenberghs Geys, and Buyse (2001). An information-theoretic alternative, based on Alonso and Molenberghs (2007) is proposed. The method is evaluated using simulations and application to data from an ophthalmologic trial, with lines of vision lost at six months as candidate surrogate endpoints for lines of vision lost at twelve months. The method is implemented as an R function.

Results. The information-theoretic approach is based on solid theory, easy to apply, and enjoys elegant properties. While the information-theoretic approach appears to be somewhat biased downwards, this is due to fact that it operates at explicitly observed outcomes, without the need for unobserved, latent scales. This is a desirable property.

Conclusions. Since the meta-analytic framework can be computationally extremely expensive, the information-theoretic approach of Alonso and Molenberghs (2007) is a viable alternative. For the ophthalmologic case study, the conclusion is that the lines of vision lost at six month does have some, but not overwhelming promise as a surrogate endpoint.

Some Key Words: Surrogate endpoint; meta-analysis; information theory; ophthalmology.

1 Introduction

Evaluation of new treatments (Z) is based on the observation of clinically meaningful variables that are sensitive to detect treatment effects, which we shall refer to as the true endpoints (T). However, such variables might increase the complexity and/or the duration of a clinical trial, either because they are costly, difficult to measure, require a long follow up time, or require a large sample size due to low incidence of the event. These problems might be avoided through replacing the true endpoints by other ones, measured earlier or in a more convenient fashion, termed surrogate endpoints (S ; Ellenberg and Hamilton 1989). For some chronic diseases, survival is the ultimate endpoint which clinicians would like to affect but cannot always observe due to the prolonged period of follow-up required. Thus, intermediate endpoints must be considered, for instance, using CD4 blood count and viral load in so-called highly active antiretroviral therapies (HAART), rather than conventional clinical events and overall survival in HIV positive subjects (DeGruttola *et al* 1993).

It is evident from the above that candidate surrogate endpoints should be associated with the true endpoint. However, an association is not sufficient for using the former as a surrogate endpoint (Fleming and DeMets 1996). A surrogate for a true endpoint is an endpoint that can be used in lieu of the true endpoint to assess treatment benefits. That is, the effect of the treatment on the surrogate endpoint should reliably predict the effect of the treatment on the true endpoint. Consequently, formal methods for validating or evaluating candidate surrogate endpoints are required. The following paragraph presents a brief overview of some statistical methods developed for the evaluation of surrogate endpoints. It should be noted that, although a quantitative evaluation is important, it is by no means the only component in the decision process. Several parties, such as the regulatory agencies and the industry developing a medical product, are involved (Burzykowski, Molenberghs, and Buyse 2004).

Several formal methods for evaluating surrogate endpoints have been proposed. A landmark paper by Prentice (1989) proposed a formal definition of surrogate endpoints, together with a set of criteria for validating surrogate endpoints. Freedman *et al* (1992) and Buyse and Molenberghs (1998) supplemented Prentice's proposal with the estimation paradigm, introducing measures which can be used to evaluate surrogate endpoints. The above proposals are all based on a single trial, thus, the lack of replication at the treatment level hampers inferences about the prediction of treatment effect on the true endpoint

based on its effect on the surrogate. This led to the proposal of a meta-analytic approach by several authors, e.g, Albert *et al* (1998), Daniels and Hughes (1997) using a Bayesian approach, Buyse *et al* (2000) and Gail *et al* (2000), using the theory of linear mixed-effects models and related hierarchical modeling frameworks. This paper is situated within the philosophy of Buyse *et al* (2000).

The meta-analytic approach proposed by Buyse *et al* (2000) has been formulated originally for two continuous, normally distributed outcomes. It has been extended to handle mixed continuous and binary endpoints by Molenberghs Geys, and Buyse (2001), as well as other sets of outcome types. A review can be found in Burzykowski, Molenberghs, and Buyse (2005). In this paper we review an extension to mixed continuous and binary outcomes, together with validation based on a so-called *Information Theoretic Approach* (ITA), introduced by Alonso and Molenberghs (2007). This paper provides the first application of ITA to the mixed continuous-discrete case.

Data from a motivating clinical trial are introduced in Section 2. Section 3 reviews the meta-analytic approach and the extensions needed in this paper. The case study is analyzed in Section 5. A simulation study is undertaken to investigate the performance of ITA, as well as the operational characteristics of various ways to derive confidence interval for measures of surrogacy. Also, the assumption of linearity, inherent in the models, is put to the test. The simulation study is reported in Section 4. Calculations are based on a generic R function, written by the first author. Details of the function are given in the appendix; the interested reader can download the function from the authors' web site.

2 The Age-related Macular Degeneration Study (ARMD)

The mixed continuous-binary case is encountered in data from a randomized multi-center clinical trial, involving patients with age-related macular degeneration, a condition in which patients progressively lose vision. Overall, 190 patients from 42 centers participated in the trial. Patients' visual acuity was assessed using standardized vision charts displaying lines of five letters of decreasing size, which patients had to read from top to bottom. The visual acuity was measured by the number of letters correctly read. Each line with at least four letters correctly read is called one 'line of vision'. The surrogate endpoint S is the binary variable indicating the loss of at least two lines of vision at 6 months while the true endpoint T is the change in visual acuity at 1 year. The binary indicator for treatment is set to $Z = -1$

for placebo and $Z = 1$ for interferon- α . In the analysis, the centers in which patients were treated will be considered as units of analysis. Six out of 42 centers participating in the trial enrolled patients only to one of the two treatment arms. These centers were excluded from further consideration. A total of 36 centers were thus available for analysis, with a number of individual patients per center ranging from 2 to 18 (183 patients overall).

3 Validation Methods

We first briefly review the meta-analytic framework in the case of two normally distributed endpoints, and then move on to its extension to the case of mixed binary continuous endpoints. In view of the computational demands placed on fitting the models, evaluation of surrogate endpoints using ITA is presented as an attainable alternative.

3.1 Two Normally Distributed Endpoints

As it is the basis for what follows, it is useful to briefly review the random-effects model used by Buyse *et al* (2000) for surrogate endpoint validation in multiple randomized trials in the case of two normally distributed endpoints. We refer to this paper for additional details. It rests upon a joint regression model for S and T :

$$S_{ij} = \mu_S + m_{Si} + \alpha Z_{ij} + a_i Z_{ij} + \varepsilon_{Sij}, \quad (1)$$

$$T_{ij} = \mu_T + m_{Ti} + \beta Z_{ij} + b_i Z_{ij} + \varepsilon_{Tij}, \quad (2)$$

where $i = 1, \dots, N$ refers to the trials and $j = 1, \dots, n_i$ to the subjects within trials. Parameters μ_S and μ_T are fixed intercepts, α and β are the fixed effects of treatment Z on the endpoints, m_{Si} and m_{Ti} are random intercepts, and a_i and b_i are the random effects of treatment Z on the endpoints in trial i . The vector of random effects $(m_{Si}, m_{Ti}, a_i, b_i)$ is assumed to be mean-zero normally distributed with covariance matrix

$$D = \begin{pmatrix} d_{SS} & d_{ST} & d_{Sa} & d_{Sb} \\ & d_{TT} & d_{Ta} & d_{Tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix}. \quad (3)$$

The error terms ε_{Sij} and ε_{Tij} are correlated and assumed to be mean-zero normally distributed with covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix}. \quad (4)$$

A measure to assess the quality of a surrogate at the trial level is given by the coefficient of determination

$$R_{\text{trial}}^2 = R_{b_i | m_{S_i}, a_i}^2 = \frac{\begin{pmatrix} d_{sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{sb} \\ d_{ab} \end{pmatrix}}{d_{bb}}. \quad (5)$$

This coefficient measures how precise the effect of treatment on the true endpoint can be predicted if the treatment effect on the surrogate has been observed in a new trial. It is unitless and ranges in the unit interval if D is positive definite, two desirable features for its interpretation.

A measure to assess the quality of surrogate at the individual level is given by the coefficient of determination

$$R_{\text{indiv}}^2 = \frac{\sigma_{ST}^2}{\sigma_{SS}\sigma_{TT}}, \quad (6)$$

which is the squared correlation between S and T after adjusting for the treatment and trial effects. A surrogate endpoint is valid if it is both trial-level valid ($R_{\text{trial}}^2 \approx 1$) and individual level valid ($R_{\text{indiv}}^2 \approx 1$). Expert opinion and confidence limits may help in determining how close these values should be close to 1. As mentioned before in Section 1, these statistical measures should be complimented with decisions from a regulatory agency and other parties, when deciding whether an endpoint is an appropriate surrogate for a true endpoint.

Although the meta-analytic approach is theoretically appealing, it faces some practical computational problems. This is especially true in situations with extreme correlation and/or when the number of trials or number of patients per trial is small. Tibaldi *et al* (2003) circumvented this issue by introducing a number of simplified modeling strategies that are less computationally demanding. Tilahun *et al* (2007) highlighted some computational issues and further illustrated these through a simulation study. They also presented SAS macros and R functions for the implementation of both the meta-analytic approach and the simplified strategies, for normally distributed endpoints.

3.2 Methods for Combined Binary and Normally Distributed Endpoints

Statistical problem where various outcomes of a combined nature are observed are common, especially with normally distributed outcomes on the one hand and binary or categorical outcomes on the other hand. Emphasis may be on the determination of the entire joint distribution of both outcomes or on specific aspects, such as the association in general or correlation in particular between both outcomes. Burzykowski, Molenberghs, and Buyse (2005) review extensions of the meta-analytic approach, ranging over continuous, binary, ordinal, time-to-event, and longitudinally measured outcomes. Here, we focus on the combination of continuous and binary outcomes.

In this section, we start with a bivariate non-hierarchical setting, which can always be expressed as the product of a marginal distribution of one of the responses and the conditional distribution of the remaining response given the former one. The main problem with this approach is that no easy expressions for the association between both endpoints are available. Thus, we opt for a symmetric treatment of both endpoints. We focus on the case where the true endpoint is continuous and the surrogate is binary, the reverse case being entirely similar.

Generalized linear mixed models for endpoints of different data types are challenging (Molenberghs and Verbeke 2005). Hence, we concentrate on two-stage fixed-effects models. In the first stage, let \tilde{S}_{ij} be a latent variable of which S_{ij} is the dichotomized version. A bivariate normal model for \tilde{S}_{ij} and T_{ij} is given by (Molenberghs, Geys, and Buyse 2001):

$$\tilde{S}_{ij} = \mu_{Si} + \alpha_i Z_{ij} + \varepsilon_{Sij}, \quad (7)$$

$$T_{ij} = \mu_{Ti} + \beta_i Z_{ij} + \varepsilon_{Tij}, \quad (8)$$

where μ_{Si} and μ_{Ti} are trial-specific intercepts, α_i and β_i are trial-specific effects of treatment Z_{ij} on the endpoints in trial i , and ε_{Si} and ε_{Ti} are correlated error terms, assumed to be zero-mean normally distributed with covariance matrix

$$\Sigma = \begin{pmatrix} \frac{1}{(1-\rho^2)} & \frac{\rho\sigma}{\sqrt{(1-\rho^2)}} \\ \frac{\rho\sigma}{\sqrt{(1-\rho^2)}} & \sigma \end{pmatrix}. \quad (9)$$

The variance of \tilde{S}_{ij} is chosen for computational reasons. Using a probit formulation like Molenberghs Geys, and Buyse (2001) and owing to the replication at the trial level, we can impose a distribution on

the trial-specific parameters. At the second stage, we assume

$$\begin{pmatrix} \mu_{Si} \\ \mu_{Ti} \\ \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_T \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} m_{Si} \\ m_{Ti} \\ a_i \\ b_i \end{pmatrix}, \quad (10)$$

where the second term on the right hand of (10) is assumed to follow a zero-mean normal distribution with dispersion matrix (3). Measures to assess the quality of the surrogate both at the trial and individual level are obtained as in (5) and (6). Interpretation of these measures and decision making follows the logic laid out in Section 3.1.

3.3 Information Theoretic Approach

Information theory is a branch of the mathematical theory of probability and statistical systems of observations. Its mathematical root is connected with the idea of disorder and entropy used in thermodynamics and statistical mechanics. Shannon (1948) can be considered the definitive, classic work on information theory. Entropy associated with a probability distribution can be defined as:

Definition 1 *Let Y be discrete random variable taking the values $\{k_1, k_2, \dots, k_m\}$ and with probability function $P(Y = k_i) = p_i$, then the entropy of Y , denoted by $H(Y)$, is defined as*

$$H(Y) = \sum_i p_i \log \left(\frac{1}{p_i} \right). \quad (11)$$

Fisher's (1925) measure of the amount of information supplied by data about an unknown parameter is well known to statisticians. Information and entropy are opposite concepts, the information content of a single event A can be defined as $I(A) = \ln P_A$ and the entropy of A is just the negative of this. That is, $H(A) = -I(A)$. The statistical concept of entropy, which tries to quantify the "epistemic" uncertainty or lack of knowledge implied by a certain distribution, plays a key role in ITA to surrogate marker validation.

Since extension of the initial meta-analytic framework to non-normal settings required a different modeling paradigm for each new combination of endpoints, a disparate collection of individual-level validation measures has emerged, even though the trial-level is mostly described by R^2 -type measures (Burzykowski, Molenberghs, and Buyse 2005). This had lead Alonso and Molenberghs (2007) to cast

validation in the ITA, thereby offering at the same time simplicity and unification. Based on ITA, Alonso and Molenberghs (2007) define surrogacy at the individual level as:

Definition 2 *Let T and S be the true and surrogate endpoint respectively. We will call S a good surrogate for T at the individual level if a “large” amount of uncertainty about T , measured by entropy based parameters, is reduced when S is known, or equivalently, if the lack of knowledge about the true endpoint is significantly reduced when the surrogate endpoint is known.*

For the general setting, Alonso *et al* (2005) and Alonso and Molenberghs (2007) considered the following generalized linear models

$$g_T\{E(T_{ij})\} = \mu_{Ti} + \beta_i Z_{ij}, \quad (12)$$

$$g_T\{E(T_{ij}|S_{ij})\} = \theta_{0i} + \theta_{1i} Z_{ij} + \theta_{2i} S_{ij}, \quad (13)$$

where g_T is an appropriate link function, μ_{Ti} are the trial-specific intercepts and β_i are trial-specific effects of treatment Z on the true endpoint in trial i . θ_{0i} and θ_{1i} are trial-specific intercepts and effects of treatment on the true endpoint when the surrogate endpoint is known. Note that (12) and (13) can be readily extended to incorporate more complex settings. Other extensions, such as non-linearity between S_{ij} and $g_T\{E(T_{ij})\}$ are possible. We assume a linear relationship between S_{ij} and $g_T\{E(T_{ij})\}$, but consider extensions of (12) and (13) in the light of simplified modeling strategy, as presented by Tibaldi *et al* (2003). They suggested several simplifications for the case of continuous true and surrogate endpoints. They have introduced the concept of three possible dimensions along which simplifications can be made: the trial, endpoint, and measurement error dimensions. Here, we adapt their ideas to ITA. We consider only the trial and measurement error dimensions, as the endpoint dimension is not applicable to univariate models.

The trial dimension provides a choice between treating the trial-specific effects as fixed or random. The former is often chosen out of necessity, when the latter is too challenging. If the trial-specific effects are chosen fixed, then (12) and (13) are used to validate the surrogate endpoint. On the other hand, if the trial-specific effects are considered random, we extend (12) and (13) to appropriate generalized linear mixed-effects models

$$g_T\{E(T_{ij})\} = \mu_T + m_{Ti} + \beta Z_{ij} + b_i Z_{ij}, \quad (14)$$

$$g_T\{E(T_{ij}|S_{ij})\} = \theta_0 + c_{Ti} + \theta_1 Z_{ij} + a_i Z_{ij} + \theta_{2i} S_{ij}, \quad (15)$$

where μ_T and β are a fixed intercept and treatment effect on the true endpoint, while m_{Ti} and b_i are a random intercept and treatment effects on the true endpoint. θ_0 and θ_1 are a fixed intercept and treatment effect on the true endpoint when the surrogate is known, and c_{Ti} and a_i are a random intercept and treatment effects on the true endpoint when the surrogate is known.

It is often the case in practice that different trials in meta-analysis have different sizes. Since univariate models are used to evaluate surrogacy in the information theoretic approach, there is a need to adjust for the heterogeneity in information content between trial-specific contributions. This is the target of the choices along the so-called measurement error dimension. One way to account for a variable amount of information per trial is by weighting the contributions according to trial size, thus giving rise to a weighted linear regression models, particularly when estimating measures for trial-level surrogacy.

Let us turn to the so-called *likelihood reduction factor* (LRF). Observe that, in the case where the true endpoint is continuous and normally distributed, (12) and (13) reduce to normal regression models and (14) and (15) reduce to linear mixed models. On the other hand, when the true endpoint is binary, (12) and (13) reduce to logistic regression models. Alonso and Molenberghs (2007) used the LRF to evaluate individual level surrogacy, which is obtained by

$$LRF = 1 - \frac{1}{N} \sum_i \exp\left(-\frac{G_i^2}{n_i}\right), \quad (16)$$

where G_i^2 denotes the log-likelihood ratio test statistic to compare (12) and (13) or (14) and (15) within trial i . The authors showed that LRF is a consistent estimator for R_h^2 . Alonso *et al* (2005) showed that LRF reduces to R_h^2 for normally distributed endpoints. For discrete outcomes, Alonso and Molenberghs (2007) showed that LRF, for non-Gaussian outcomes, has an upper bound that can generally be smaller than 1, i.e., $R_h^2 \leq 1 - e^{-2H(T)}$. Thus, for discrete true endpoints, it would be more meaningful to use the quantity

$$\tilde{R}_h^2 = \frac{R_h^2}{1 - e^{-2H(T)}}, \quad (17)$$

that takes value 1 when both endpoints are deterministically related.

Alonso and Molenberghs (2007) developed asymptotic confidence intervals for R_h^2 , based on the idea of Kent (1983), to build confidence intervals for $2I(T, S)$. Let $\hat{a} = 2n\hat{I}(T, S)$, where n is the number

of patients. Define $\kappa_{1:\alpha}(\mathbf{a})$ and $\delta_{1:\alpha}(\mathbf{a})$ by $P(\chi_2^1(\kappa_{1:\alpha}(\mathbf{a})) \geq a) = \alpha$ and $P(\chi_2^1(\delta_{1:\alpha}(\mathbf{a})) \leq a) = \alpha$. If $P(\chi_2^1(0) \geq a) = \alpha$ then we set $\kappa_{1:\alpha}(\mathbf{a}) = 0$. A conservative two-sided $1 - \alpha$ asymptotic confidence interval for R_h^2 is

$$\sum_i \alpha_i [n_i^{-1} \kappa_{1:\alpha}^i(\hat{\mathbf{a}}), n_i^{-1} \delta_{1:\alpha}^i(\hat{\mathbf{a}})], \quad (18)$$

where $1 - \alpha_i$ is the Bonferroni confidence level for the trial intervals (Alonso and Molenberghs 2007). This asymptotic interval has considerable computational advantage with respect to the bootstrap approach used by Alonso *et al* (2005). Although ITA involves substantial mathematics, its implementation in practice is fairly straightforward and less computer-intensive than the meta-analytic approach. This is a direct consequence of the fact that the models used in the former are univariate models, which can be fitted using any standard regression software. However, the performance of this approach has not been studied in the mixed continuous and binary endpoint settings. In the next section, insight into the performance of this approach, together with that of the asymptotic interval, is offered through a simulation study.

4 A Simulation Study

We first lay out the design of the simulation study, whereafter the results are described.

4.1 Design of the Simulation Study

Due to the computational difficulties encountered in practice with the bivariate random-effects models required for the meta-analytic approach by Buyse *et al* (2000), ITA becomes an interesting option to consider in practice. However, as stated earlier, the performance of the later has not been investigated in the mixed continuous binary setting, and is the focus of this section. Here, we outline the procedures followed in generating the data used for simulation. The data were generated based on model (1)–(2). Choices made are $\mu_S = 0.5$, $\mu_T = 0.45$, $\alpha = 0.05$, and $\beta = 0.03$. Values assumed for the covariance matrices are:

$$\Sigma = \begin{pmatrix} 3 & 2.4 \\ & 3 \end{pmatrix}, \quad D = \begin{pmatrix} 3 & 2.4 & 0 & 0 \\ & 3 & 0 & 0 \\ & & 3 & 2.84605 \\ & & & 3 \end{pmatrix}.$$

After generating continuous outcomes based on the above models, a binary surrogate is obtained by dichotomizing the resulting continuous surrogate using the fixed intercept as cut-off point. The dichotomized surrogate takes value 1 if the corresponding continuous surrogate is greater than μ_S and zero otherwise. The above model assumes trial-level and individual-level R^2 values of 0.9 and 0.64, respectively, at the continuous scale. It is important to note that this value of the individual-level R^2 is the squared correlation between the latent unobservable continuous surrogate endpoint and the observable true endpoints. However, the situation is totally different at the trial-level. Based on (7) and (8), Alonso *et al* (2005) showed that the relationship between the treatment effects on the latent-continuous and observed-binary surrogate endpoints is linear. Hence, the value of the trial-level R^2 (0.9) is valid both for the latent and observed surrogate.

The number of trials was fixed to either 5, 10, 20 or 30. There were 2 sets of trial sizes used, the first set consists of 10, 20, 40 or 60, which we term *small trial size*. The second set consists of 100, 150, 200 or 300, termed *large trial size*. A full combination of the number of trials and trial sizes was obtained. In each case, 100 runs were performed, assuming either models (12) and (13) or (14) and (15).

Apart from the primary objectives to investigate the performance of ITA as well as comparing the percentile intervals with the asymptotic interval by Alonso and Molenberghs (2007), there are two secondary objectives. The first is to investigate the impact of alternative link functions, at the individual-level, on the performance of ITA. Thus, both probit and logit link functions were implemented in all settings. Second, both linear and non-linear (splines) functions were considered, at the trial-level, to explore the assumption of linearity between treatment effects. Results are shown for the probit link at the individual-level and linear function at the trial-level. Histograms are used to depict results of the secondary objectives.

4.2 Simulation Results

Tables 1 and 2 show results of the individual-level and trial-level R^2 measures, respectively, for large trial sizes based on models (14) and (15). The tables have columns indicating the number of trials, the trial size, median R^2 , percentile bootstrap and asymptotic 95% confidence intervals.

ITA yields estimates of surrogacy at the individual-level, bounded above by 0.3. Hence, the approach

yields estimates substantially lower than the value assumed when generating the datasets, 0.64. This phenomenon is observed in all settings considered in the simulation study. However, it should be noted that the value of 0.64 is the individual-level surrogacy at the latent scale, whereas ITA estimates assess the individual-level surrogacy at the observed scale. Also, it is expected that dichotomizing a continuous variable leads to information loss, which would imply that results obtained from the continuous and discrete version should not generally be expected to be in agreement with each other.

Unlike the individual level, Alonso *et al* (2002) showed that the trial-level surrogacy at the latent scale translates equally to the observed scale. For small trial sizes ITA tends to underestimate the trial-level surrogacy. Nevertheless, the models perform considerably well for large trial sizes. The mixed-effect models, (14) and (15), outperform the fixed-effect models, (12) and (13), in all simulation settings considered. However, the mixed-models had some convergence issues, which were not encountered with the fixed-effect models. Even so, the percentage of non-convergence is smaller than 10% within each simulation setting. Generally, increasing the number of trials has little effect on the surrogacy measures, although increasing the trial size appears to yield better estimates for the surrogacy measures. Also, it is not advisable to use a very small number of trials, as it may overestimate or not provide enough data points to reliably assess the trial-level surrogacy.

The 95% asymptotic intervals are tighter than the 95% percentile bootstrap intervals for all simulation settings considered. The discrepancy between these intervals reduces with increases in the number of trials and trial sizes. Further, the choice of an appropriate link function appears to have little influence on the results. Figure 1 shows a plot of magnitude of pooled differences of the trial-level estimates between the logit and probit links, for the fixed-effect models with large sample size. Observe that more than 97% of the samples have differences below 0.1. Also, almost identical results were obtained in each sample when the spline and linear functions were considered at the trial level. This is supported by Figure 1, as more than 93% of the samples have differences inferior to 0.04.

5 Analysis of the ARMD Data

The case study introduced in Section 2 will now be analyzed. For these data, it is natural to consider center as the unit of analysis, as the data comes from a multi-center trial. There were 36 centers, each

treating between 2 and 18 patients. The two-stage meta-analytic approach and the corresponding ITA models, described in Section 3, have been applied to this dataset and results displayed in Table 9.

Extension of the meta-analytic approach to the mixed continuous and binary endpoints, using two-stage fixed-effects model yields $R_{\text{indiv}}^2=0.42$ (s.e. 0.13) and $R_{\text{trial}}^2=0.44$ (s.e. 0.09). Thus, the loss of at least two lines of vision at 6 months is a relatively poor surrogate for visual acuity at 1 year, a conclusion in synchrony with the one reached by *et al* (2000) at the continuous level.

At the individual level, ITA yield estimates of R_{indiv}^2 ranging from 0.2319 to 0.2735. It should be noted that we do not have information about the degree of under-estimation of R_{indiv}^2 by ITA at the observed scaled. As mentioned earlier, research on this issue is still ongoing. Nevertheless, the very low values obtained indicate that the loss of at least two lines of vision at 6 months may not be a good surrogate for visual acuity at 1 year, at the individual level.

ITA yields estimates of R_{trial}^2 ranging from 0.3211 to 0.4864. This indicates that the loss of at least two lines of vision at 6 months does not seem to be a very good surrogate for visual acuity at 1 year, at the trial level. It should be noted that the size of the largest unit of analysis (center) was only 18, though. Thus, there may be a considerable degree of under-estimation on the estimates of R_{trial}^2 .

There appears to be no difference between the probit and logit link functions on these data. Also, the line and spline models yield similar results, indicating that the linearity assumption at the trial level may be a plausible one. Furthermore, the mixed models generally have higher estimates for surrogacy measures than the fixed models, hence, exhibiting a lower degree of underestimation.

6 Discussion

In this paper, we reviewed the meta-analytic strategy of Buyse *et al* (2000), its extension to mixed binary and continuous endpoints, and the information theoretic approach for validating surrogate endpoints. Combination of the latter with combined-type outcomes is novel. The meta-analytic approach and its extension are mathematically appealing, but encounter practical and/or computational issues. The information theoretic approach involves substantial mathematics yet it is more practically feasible than the meta-analytic approach as it depends on simple univariate models.

The paper primarily investigated the performance of the ITA for combined continuous and binary endpoints, particularly continuous surrogate and binary true endpoints, through a simulation study. Generally, this approach underestimates the measures of surrogacy. The underestimation reduces with increase in both the number of trials and trial sizes. However, the simulation study showed that the degree of underestimation is higher with very small trial sizes, even for large number of trials.

The model proposed by Alonso *et al* (2005) for a general setting, which we referred to as fixed-effects models, was outperformed by its extension to generalized linear mixed models, which we refer to as mixed models. Quite similar results were obtained by extending the linear relationship between the true and surrogate endpoints to non-linear, spline-based models, at the trial level. Thus, it may be reasonable to assume a linear relationship between the treatment effects on the true and surrogate endpoints.

Asymptotic confidence intervals for surrogacy measures (R_{indiv}^2 and R_{trial}^2) developed by Alonso and Molenberghs (2007) performed better than bootstrap confidence intervals used by Alonso *et al* (2005). The asymptotic confidence intervals are computationally advantageous and are tighter than the bootstrap confidence intervals.

The choice of link function appears to have little influence on the estimates of the surrogacy measures. Particularly, the logit and probit link functions gave similar estimates in all settings considered in the simulation study. This is also supported by the fact that these link functions gave almost identical estimates when applied to the motivational case study. These findings are not surprising in view of their well-known relationship.

The meta-analytic strategy for evaluating surrogacy faces computational problems, the information theoretic approach encounters less computational problems than the former and may be of more practical use. However, the latter is biased downwards. It is advisable to reserve the use of ITA for larger trial sizes. Also, the extended generalized linear mixed models are recommended.

References

Albert JM, Ioannidis JPA, Reichelderfer P, Conway B, Coombs RW, Crane L, Demasi R, Dixon DO, Flandre P, Hughes MD, Kalish LA, Lartnz K, Lin D, Marschner IC, Muñoz A, Murray J, Neaton J, Pettinelli C, Rida W, Taylor JMG, and Welles SL Statistical issues for HIV surrogate endpoints:

- point and counterpoint. *Stat. Med.* 1998; 17: 2435–2462.
- 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* 2004b; 60: 845–853.
- Alonso A, Geys H, Molenberghs G, and Vangeneugden T. Investigating the criterion validity of psychiatric symptom scales using surrogate marker validation methodology. *J. Biopharm. Stat.* 2002; 12: 161–178.
- Alonso A. and Molenberghs G. Surrogate marker evaluation from an information theory perspective. *Biometrics* 2007; 63: 180–186.
- Alonso A, Molenberghs G, Burzykowski T, Renard D, Geys H, Shkedy Z, Tibaldi F, Cortiñas J, and Buyse M. (Prentice’s approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics* 2004a; 60: 724–728.
- Alonso A, Molenberghs G, Geys, H, and Buyse M. A unifying approach for surrogate marker validation based on Prentice’s criteria. *Stat. Med.* 2005; 25: 205–211.
- Burzykowski T, Molenberghs G, and Buyse M. The validation of surrogate endpoints using data from randomized clinical trials: a case-study in advanced colorectal cancer. *J. Roy. Stat. Soc. A* 2004; 167: 103–124.
- Burzykowski T, Molenberghs G, and Buyse M. (2005). *The Evaluation of Surrogate Endpoints*. New York: Springer.
- Buyse M. and Molenberghs G. The validation of surrogate endpoints in randomized experiments. *Biometrics* 1998; 54: 1014–1029.
- Buyse M, Molenberghs G, Burzykowski T, Renard D, and Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000; 1: 49–67.
- Daniels MJ and Hughes MD Meta-analysis for the evaluation of potential surrogate markers. *Stat. in Med.* 1997; 16: 1515–1527.

- DeGruttola V, Wulfsohn M, Fishk MA, and Tsiatis, AA Modelling the relationship between survival and CD4 lymphocytes in patients with AIDS and AIDS-related complex. *J. Acq. Imm. Def. Synd.* 1993; 6: 359–365.
- Ellenberg SS and Hamilton JM Surrogate endpoints in clinical trials: cancer. *Stat. Med.* 1989; 8: 405–413.
- Fisher R Theory of statistical estimation. *Proc. Cam. Phil. Soc.* 1925; 22: 700–725.
- Fleming TR and DeMets DL Surrogate endpoints in clinical trials: are we being misled? *Ann. Internal Med.* 1996; 125: 605–613.
- Freedman LS, Graubard BI, and Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat. Med.* 1992; 11: 167–178.
- Gail MH, Pfeiffer R, van Houwelingen HC, Carroll RJ. On meta-analytic assessment of surrogate outcomes. *Biostatistics* 2000; 1: 231–246.
- Kent J. Information gain and a general measure of correlation. *Biometrika* 1993; 70: 163–173.
- Molenberghs G, Geys, H, and Buyse M Evaluation of surrogate end-points in randomized experiments with mixed discrete and continuous outcomes. *Stat. Med.* 2001; 20: 3023–3038.
- Molenberghs G and Verbeke G (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. *Stat. Med.* 1989; 8: 431–440.
- Shannon C. A mathematical theory of communication. *Bell System Technical Journal* 1948; 27: 379–423 and 623–656.
- Tibaldi FS, Cortiñas Abrahantes J, Molenberghs G, Renard D, Burzykowski T, Buyse M, Parmar M, Stijnen T, and Wolfinger R. Simplified hierarchical linear models for the evaluation of surrogate endpoints. *J. Stat. Comp. Simul.* 2003; 73: 643–658.
- Tilahun A, Assam P, Alonso A, and Molenberghs G. Flexible surrogate marker evaluation from several randomized clinical trials with continuous endpoints, using R and SAS. *Comp. Stat. Data Anal.* 2007; 51: 4152–4163.

A Appendix

We briefly outline the use of the R function we developed.

`conbin` is an R function that performs surrogate marker evaluation in the case of mixed continuous and binary endpoints, using the information theoretic approach as proposed by Alonso and Molenberghs (2007), and laid out in this paper. The function is called as follows:

```
conbin(resp, endp, treat, units, subject,  
       surtype="binary", mixed=TRUE, link="probit", weighted=FALSE)
```

The arguments are

resp: the response variable holding the values for both the true and surrogate endpoints, for each subject in the data. It should be sorted by endpoints (see below).

endp: endpoint indicator, should be coded as: surrogate endpoint= -1 and true endpoint= 1 .

treat: treatment indicator: Standard treatment is coded as -1 and new treatment as 1 .

units: the unit of analysis, e.g, trial, center, investigator, ...

subject: a variable identifying the various subjects/patients in the data.

surtype: character string specifying the type of the surrogate endpoint. A binary or continuous surrogate endpoint is specified by "binary" or "continuous," respectively. Note that a binary surrogate endpoint implies a continuous true endpoint and vice versa.

mixed: a logical value. "TRUE" indicates that generalized linear mixed models should be used, generalized linear models are used otherwise. The default is "TRUE."

link: a specification for the model link function of the binary endpoint. Appropriate links are "probit," "logit," "log," and "cloglog" (complementary log-log). The default is "probit."

weighted: a logical value indicating whether weighted normal regression should be performed for the estimation of trial level association (R-square). The default is "FALSE."

Details: This function evaluates surrogacy for mixed continuous and binary endpoints, when 2 treatment groups are used. It is necessary that treatment and endpoint be coded as 1 and -1 . A consequence of the 0/1 coding is that the group coded as 0 is assumed to have a smaller variance than the group coded as 1. On the other hand the $-1/+1$ coding leads to equal variance in both groups.

Value: An object of class "conbin," representing the surrogate endpoint evaluation for mixed continuous binary endpoints, using the information theoretic approach. Generic functions, such as print, plot, and summary, are equipped with ways to display the results of the fit.

Table 1: Simulation study. Univariate mixed-effects model for large trial sizes, individual-level surrogacy.

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	100	0.1537	(0.0600;0.2512)	(0.1099;0.2095)
2	5	150	0.1580	(0.0816;0.2706)	(0.1245;0.2073)
3	5	200	0.1649	(0.0826;0.2719)	(0.1309;0.2030)
4	5	300	0.1553	(0.0821;0.2680)	(0.1341;0.1926)
5	10	100	0.1652	(0.0974;0.2299)	(0.1269;0.1993)
6	10	150	0.1610	(0.0875;0.2427)	(0.1335;0.1927)
7	10	200	0.1651	(0.0801;0.2362)	(0.1375;0.1887)
8	10	300	0.1640	(0.0961;0.2353)	(0.1408;0.1827)
9	20	100	0.1624	(0.1119;0.2107)	(0.1356;0.1870)
10	20	150	0.1607	(0.1094;0.2106)	(0.1403;0.1823)
11	20	200	0.1596	(0.1162;0.2055)	(0.1433;0.1797)
12	20	300	0.1590	(0.1103;0.2069)	(0.1459;0.1756)
13	30	100	0.1633	(0.1203;0.1967)	(0.1417;0.1839)
14	30	150	0.1592	(0.1159;0.1992)	(0.1434;0.1777)
15	30	200	0.1578	(0.1134;0.1953)	(0.1440;0.1736)
16	30	300	0.1601	(0.1215;0.1961)	(0.1473;0.1715)

Table 2: Simulation study. Univariate mixed-effects model for large trial sizes, trial-level surrogacy.

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	100	0.8159	(0.0752;0.9804)	(0.3016;0.9360)
2	5	150	0.8323	(0.0864;0.9823)	(0.3305;0.9481)
3	5	200	0.8619	(0.1321;0.9870)	(0.3545;0.9594)
4	5	300	0.8807	(0.3362;0.9884)	(0.4123;0.9761)
5	10	100	0.7696	(0.0043;0.9211)	(0.3614;0.8849)
6	10	150	0.7975	(0.0144;0.9417)	(0.3905;0.8899)
7	10	200	0.7973	(0.1295;0.9422)	(0.3876;0.9031)
8	10	300	0.7870	(0.0374;0.9502)	(0.4019;0.8984)
9	20	100	0.7628	(0.1500;0.8841)	(0.4583;0.8701)
10	20	150	0.8060	(0.4690;0.9018)	(0.5229;0.9073)
11	20	200	0.8075	(0.2821;0.9092)	(0.5171;0.8975)
12	20	300	0.8289	(0.2745;0.9170)	(0.5626;0.9177)
13	30	100	0.7770	(0.2732;0.9041)	(0.5440;0.8706)
14	30	150	0.7835	(0.4593;0.8963)	(0.5672;0.8892)
15	30	200	0.7976	(0.5068;0.9130)	(0.5823;0.8941)
16	30	300	0.8158	(0.5118;0.9234)	(0.6072;0.9024)

Table 3: *Simulation study. Univariate fixed-effects model for large trial sizes, individual-level surrogacy.*

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	100	0.1511	(0.0575;0.2483)	(0.1068;0.2054)
2	5	150	0.1553	(0.0803;0.2688)	(0.1222;0.2045)
3	5	200	0.1659	(0.0796;0.2696)	(0.1293;0.2011)
4	5	300	0.1542	(0.0811;0.2663)	(0.1330;0.1913)
5	10	100	0.1609	(0.0939;0.2294)	(0.1242;0.1961)
6	10	150	0.1586	(0.0861;0.2410)	(0.1318;0.1907)
7	10	200	0.1626	(0.0790;0.2366)	(0.1362;0.1873)
8	10	300	0.1634	(0.0952;0.2343)	(0.1400;0.1817)
9	20	100	0.1588	(0.1086;0.2089)	(0.1326;0.1836)
10	20	150	0.1576	(0.1071;0.2095)	(0.1383;0.1801)
11	20	200	0.1581	(0.1144;0.2045)	(0.1418;0.1781)
12	20	300	0.1578	(0.1092;0.2061)	(0.1448;0.1745)
13	30	100	0.1602	(0.1159;0.1943)	(0.1387;0.1806)
14	30	150	0.1569	(0.1135;0.1973)	(0.1414;0.1755)
15	30	200	0.1563	(0.1120;0.1934)	(0.1425;0.1720)
16	30	300	0.1590	(0.1202;0.1952)	(0.1463;0.1705)

Table 4: *Simulation study. Univariate fixed-effects model for large trial sizes, trial-level surrogacy.*

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	100	0.8134	(0.0133;0.9852)	(0.2938;0.9274)
2	5	150	0.8004	(0.1640;0.9796)	(0.2884;0.9560)
3	5	200	0.8342	(0.2109;0.9906)	(0.3237;0.9576)
4	5	300	0.8167	(0.1457;0.9827)	(0.3027;0.9564)
5	10	100	0.7350	(0.2608;0.9163)	(0.3272;0.9118)
6	10	150	0.7797	(0.1995;0.9308)	(0.3619;0.9177)
7	10	200	0.7668	(0.1808;0.9478)	(0.3693;0.9166)
8	10	300	0.7755	(0.1742;0.9066)	(0.3645;0.9129)
9	20	100	0.7075	(0.4191;0.8659)	(0.4117;0.8692)
10	20	150	0.7283	(0.4732;0.8865)	(0.4548;0.8903)
11	20	200	0.7399	(0.4631;0.8786)	(0.4549;0.8893)
12	20	300	0.7543	(0.4872;0.8850)	(0.4713;0.8960)
13	30	100	0.7145	(0.4588;0.8440)	(0.4730;0.8516)
14	30	150	0.7200	(0.4596;0.8530)	(0.4818;0.8545)
15	30	200	0.7223	(0.5393;0.8673)	(0.4956;0.8643)
16	30	300	0.7492	(0.5445;0.8747)	(0.5225;0.8754)

Table 5: *Simulation study. Univariate fixed-effects model for small trial sizes, individual-level surrogacy.*

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	10	0.1501	(0.0003;0.3844)	(0.0497;0.3338)
2	5	20	0.1443	(0.0255;0.3379)	(0.0670;0.2777)
3	5	40	0.1589	(0.0552;0.3060)	(0.0974;0.2560)
4	5	60	0.1607	(0.0471;0.2810)	(0.0983;0.2251)
5	10	10	0.1503	(0.0520;0.3253)	(0.0679;0.2846)
6	10	20	0.1676	(0.0855;0.2697)	(0.0945;0.2559)
7	10	40	0.1608	(0.0737;0.2506)	(0.1068;0.2196)
8	10	60	0.1576	(0.0903;0.2436)	(0.1152;0.2077)
9	20	10	0.1558	(0.0627;0.2696)	(0.0878;0.2451)
10	20	20	0.1601	(0.0907;0.2138)	(0.1067;0.2203)
11	20	40	0.1623	(0.1055;0.2223)	(0.1216;0.2024)
12	20	60	0.1619	(0.1099;0.2113)	(0.1286;0.1949)
13	30	10	0.1572	(0.0916;0.2443)	(0.0986;0.2287)
14	30	20	0.1634	(0.0994;0.2297)	(0.1207;0.2148)
15	30	40	0.1588	(0.1019;0.2114)	(0.1283;0.1945)
16	30	60	0.1641	(0.1197;0.2096)	(0.1373;0.1919)

Table 6: *Simulation study. Univariate fixed-effects model for small trial sizes, trial-level surrogacy.*

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	10	0.4767	(0.0010;0.9480)	(0.1236;0.8055)
2	5	20	0.5559	(0.0293;0.9429)	(0.1364;0.8745)
3	5	40	0.6232	(0.0800;0.9593)	(0.1429;0.9012)
4	5	60	0.5870	(0.0267;0.9536)	(0.1333;0.8739)
5	10	10	0.3973	(0.0333;0.7484)	(0.0870;0.7478)
6	10	20	0.4644	(0.0567;0.8768)	(0.1371;0.7872)
7	10	40	0.4659	(0.0143;0.8405)	(0.1297;0.7898)
8	10	60	0.5583	(0.0857;0.8743)	(0.1767;0.8338)
9	20	10	0.4139	(0.0649;0.7116)	(0.1483;0.6732)
10	20	20	0.4745	(0.0820;0.7412)	(0.1853;0.7194)
11	20	40	0.4979	(0.1206;0.8008)	(0.2218;0.7467)
12	20	60	0.5251	(0.1257;0.7265)	(0.2155;0.7424)
13	30	10	0.4141	(0.1317;0.6158)	(0.1712;0.6360)
14	30	20	0.4522	(0.2472;0.6932)	(0.2231;0.6853)
15	30	40	0.4855	(0.1755;0.6816)	(0.2367;0.6964)
16	30	60	0.5092	(0.1934;0.6993)	(0.2570;0.7143)

Table 7: Simulation study. Univariate mixed-effects model for small trial sizes, individual-level surrogacy.

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	10	0.1595	(0.0235;0.3929)	(0.0596;0.3667)
2	5	20	0.1594	(0.0404;0.3332)	(0.0736;0.2920)
3	5	40	0.1669	(0.0567;0.2979)	(0.1008;0.2612)
4	5	60	0.1635	(0.0475;0.2784)	(0.1011;0.2291)
5	10	10	0.1712	(0.0553;0.3263)	(0.0789;0.3054)
6	10	20	0.1777	(0.0924;0.2682)	(0.1007;0.2652)
7	10	40	0.1633	(0.0805;0.2515)	(0.1106;0.2246)
8	10	60	0.1598	(0.0968;0.2401)	(0.1178;0.2110)
9	20	10	0.1773	(0.0823;0.2881)	(0.1053;0.2708)
10	20	20	0.1718	(0.1057;0.2253)	(0.1163;0.2329)
11	20	40	0.1683	(0.1099;0.2262)	(0.1262;0.2080)
12	20	60	0.1657	(0.1127;0.2138)	(0.1318;0.1986)
13	30	10	0.1799	(0.1135;0.2604)	(0.1164;0.2528)
14	30	20	0.1750	(0.1101;0.2380)	(0.1294;0.2256)
15	30	40	0.1637	(0.1070;0.2135)	(0.1330;0.1999)
16	30	60	0.1687	(0.1220;0.2137)	(0.1403;0.1952)

Table 8: Simulation study. Univariate mixed-effects model for small trial sizes, trial-level surrogacy.

sim. #	strategy		R_{indiv}^2	confidence intervals	
	# trials	# subjects		percentile	asymptotic
1	5	10	0.4173	(0.0010;0.9497)	(0.1272;0.7803)
2	5	20	0.5699	(0.0021;0.9589)	(0.1306;0.8506)
3	5	40	0.5990	(0.0136;0.9389)	(0.1306;0.8702)
4	5	60	0.5914	(0.0024;0.9395)	(0.1359;0.8416)
5	10	10	0.4024	(0.0038;0.7753)	(0.0886;0.7027)
6	10	20	0.4395	(0.0330;0.8527)	(0.1204;0.7794)
7	10	40	0.4754	(0.0047;0.8042)	(0.1250;0.7709)
8	10	60	0.5226	(0.0280;0.8562)	(0.1576;0.8047)
9	20	10	0.4212	(0.0017;0.6803)	(0.1470;0.6495)
10	20	20	0.4769	(0.0406;0.7370)	(0.1886;0.7057)
11	20	40	0.5218	(0.0373;0.7701)	(0.2167;0.7307)
12	20	60	0.5352	(0.0034;0.7554)	(0.2218;0.7201)
13	30	10	0.4268	(0.0415;0.6198)	(0.1753;0.6284)
14	30	20	0.4505	(0.0105;0.7169)	(0.2288;0.6677)
15	30	40	0.5175	(0.1498;0.7066)	(0.2636;0.7106)
16	30	60	0.5322	(0.2769;0.7543)	(0.2756;0.7267)

Table 9: Macular degeneration trial. Estimates (standard error) of the individual-level (R_{indiv}^2) and trial-level (R_{trial}^2) surrogacy analysis based on information theoretic approach.

level	type	probit link		logit Link	
		fixed	mixed	fixed	mixed
trial	line	0.3271(0.1339)	0.4864(0.1247)	0.3212(0.1339)	0.4832(0.1251)
	spline	0.3270(0.1339)	0.4864(0.1247)	0.3211(0.1339)	0.4832(0.1251)
individual		0.2319(0.1288)	0.2733(0.1323)	0.2319(0.1288)	0.2735(0.1323)

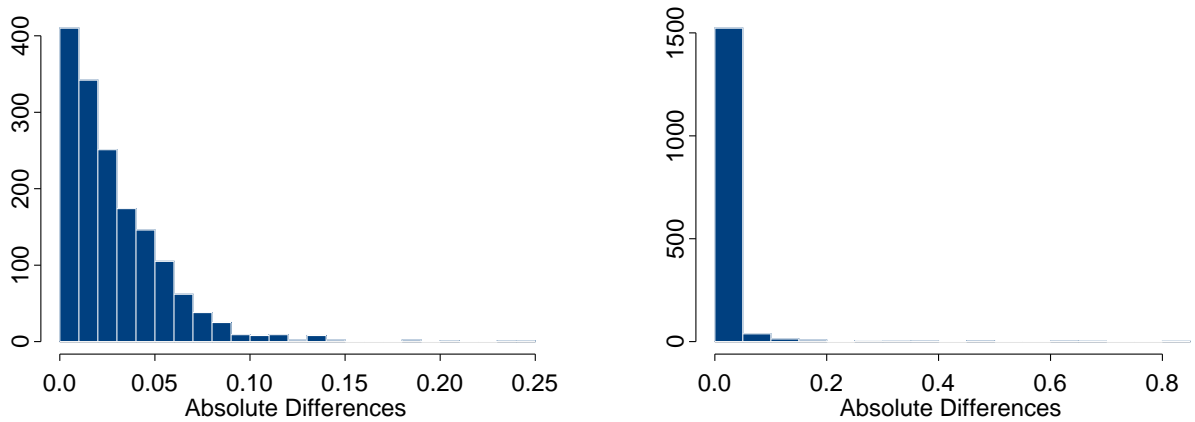


Figure 1: Frequency histogram showing absolute differences obtained for each sample. (a) Difference between probit and logit link functions. (b) Linearity assumption at trial level: linear versus spline-based.