JP Journal of Biostatistics
Volume 3, Number 2, 2009, Pages 145-169
Published Online: July 16, 2009
This paper is available online at http://www.pphmj.com

UNDERSTANDING THE NATURE OF TREATMENT EFFECTS OF MULTIPLE ENDPOINTS OF A CLINICAL TRIAL USING BOOTSTRAP

MUSHFIQUR RASHID and M. F. HUQUE

Division of Biometrics IV
Office of Biostatistics/CDER/FDA
Room WO 21-3644
10903 New Hampshire Avenue
Silver Spring, MD 20993, U. S. A.
e-mail: mushfiqur.rashid@fda.hhs.gov

Abstract

In a clinical trial, treatment effects of multiple endpoints can be either of overlapping (partially or completely) or of non-overlapping nature, in the sense that a single or a group of multiple endpoints is jointly able to explain or not able to explain a part or whole of the treatment effect of an endpoint of interest. This information is useful in assessing the total benefit of a treatment for a set of multiple endpoints of a trial. An easy-to-understand measure for this purpose is the proportion of the treatment effect (PTE) of a clinical endpoint explained by the treatment effects of other endpoint(s) of interest. Conventionally, it is estimated by the ratio of two statistics. However, this ratio estimate has been statistically challenging for some applications as it can produce a wide confidence interval beyond the $[0, 1]$ interval and even the point estimate may fall outside this interval. This article presents a bootstrapping based measure using linear models and a simple bootstrapping based r^* measure that avoid the weakness of the conventional PTE measure. The former is a

2000 Mathematics Subject Classification: 62G09, 62G35, 62-07, 62P10.

Keywords and phrases: PTE measure, multiple endpoints, association measures, surrogate endpoints, marginal model, conditional model, bootstrapping methods, meta-analysis.

Received February 4, 2009

conditional probability measure involving PTE and the latter one is an association measure between the induced treatment effects of a clinical endpoint versus the part of it that is explained by other endpoint(s) of interest. These measures can help in understanding the nature of the treatment effects of multiple endpoints of a clinical trial, for finding whether some of these treatment effects are of overlapping nature or of independent nature in adding to the treatment benefit.

1. Introduction

Randomized clinical trials generally include multiple response variables called endpoints for characterizing clinically meaningful benefits of a study intervention. Causal pathways by which an intervention induces treatment effects in endpoints in a patient population can be of different nature for different endpoints. It is of interest to learn about the nature of treatment effects found for different endpoints in a trial, whether some of these treatment effects are of overlapping or non-overlapping nature for adding to the treatment benefit. If treatment effects in two endpoints of a trial are mostly of overlapping nature, then these treatment effects for the two endpoints can be considered as corroborative, strengthening each other's result. On the other hand, these treatment effects can be typically non-overlapping, i.e., one does not explain the other. In that case, each endpoint can be of different clinical value. This can happen when the test treatment targets the two endpoints differently.

For example, consider a primary clinical endpoint and a key secondary endpoint for a confirmatory controlled clinical trial. The role the secondary endpoint may be either for strengthening the result of the primary endpoint or for additional treatment benefit. If the secondary endpoint treatment effect result explains little or none of the treatment effect of the primary endpoint, then this secondary endpoint treatment benefit may qualify for an additional benefit of the treatment in addition to that already contributed by the primary endpoint. However, if this secondary endpoint treatment effect explains most of the treatment effects of the primary endpoint, then this secondary endpoint result strengthens the result of the primary endpoint but may not qualify for an extra benefit of the treatment. Such information can be of considerable value in proper labeling of a new drug or treatment for the benefit of patients and prescribing physicians.

Some statistical methods are already available that can help in this endeavor, though these methods were developed mostly under the umbrella of evaluation of surrogate endpoints, i.e., for validating an intermediate endpoint or a surrogate

marker endpoint for the purpose of replacing a late observable clinical endpoint for accelerating the drug approval process (Chakravarty [4]). Weir and Walley [7] have reviewed some of these methods. For example, Buyse and Molenberghs [3] introduced a method based on the RE (relative efficacy) concept. Buyse et al. [2] considered the coefficient of determination R^2 and Alonso et al. [1] introduced an interesting Likelihood Reduction Factor (LRF) measure.

On the other hand, Freedman et al. [5] proposed the concept of PTE, which stands for the proportion of treatment effect of a clinical endpoint explained by other endpoint(s) of interest. They considered this other endpoint as an intermediate endpoint in the context of a surrogate endpoint. Conventionally, this PTE is estimated on taking the ratio of two statistics. The denominator of this ratio is a measure of the total treatment effect of a clinical endpoint and numerator is a measure of the treatment effect of this clinical endpoint explained by the other endpoint(s) of interest. This ratio estimate, though conceptually popular, is statistically a challenging measure as its sampling distribution can be poorly heavy tailed, and consequently for some applications, the confidence interval may be wide often exceeding the $[0, 1]$ interval, and even the point estimate may fall outside this interval. Nonetheless the PTE is an attractive and easy-to-understand concept for clinical trial applications and for clinicians, and there is a need for a different approach for evaluating PTE that can be free from these problems.

This article presents a bootstrapping based measure using linear models and a simple bootstrapping based r^* measure that avoid the weakness of the conventional PTE measure. The former is a conditional probability measure involving PTE. It is bootstrapping based and involves linear modeling for evaluating PTE that avoids the troubling concerns of the conventional ratio estimate. In this approach, we set two linear models for a given primary endpoint. The first model is the marginal model which has three terms: an intercept, a treatment effect term and the error term. This model provides a least squares estimate of the 'total treatment effect' of a primary clinical endpoint. We represent this estimate by the notation *U*. The second model is the conditional linear model of the main endpoint given the other endpoints of interest. This conditional model gives the so called the 'adjusted treatment effect' for the primary clinical endpoint. The direct difference between the two, i.e., the difference between the 'total treatment effect' of the primary endpoint and its 'adjusted treatment effect' along with the 'total treatment effect' itself, are the two key quantities for estimating how much of the treatment effect of the primary

endpoint is explained by the treatment effects of the other endpoints included in the conditional model. We denote this direct difference by the notation D. The ratio D/U, i.e., of this direct difference measure to the total treatment effect of the primary endpoint is the conventional PTE measure.

Our measure differs from this conventional ratio estimate in the sense that these two treatment effect random variables D and U are converted into a binary outcome on checking whether a value D exceeds a fraction c times the corresponding value of U or not. A conditional probability curve P_c is obtained as a function of c that is the probability that the value of D exceeds a fraction c times the corresponding value of U given that the primary endpoint treatment effect is statistically significant in the trial. A maximum value of c is then read from this curve so that $P_c \geq 1 - \gamma$ (e.g., $\gamma = 0.20$). This c then serves as a lower bound for PTE with the conditional probability of P_c associated with it. For example, if for such a $c = 0.50$ the associated $P_c = 80$ percent, then one may conclude that the chances are fairly high for concluding that the true PTE value is at least 50 percent. And, if for such a c of 0.75 the associate value of $P_c = 10$ percent, then one may conclude that chances are fairly low for the true PTE of greater than 75 percent. This lower bound is different from the conventional lower confidence interval for the PTE as it is obtained under the restriction that the treatment difference U is statistically significant at a specified significance level of α. The article proposes a bootstrapping based approach for determining this P_c curve and provides an algorithm for its computation.

The above bootstrapping method, while computing P_c for a given c, also provides a bootstrapping joint distribution of U and D which readily gives a bootstrapping based r^* correlation measure between U and D, i.e., between the induced treatment effects of a clinical endpoint versus the part of it that is explained by other endpoint(s) of interest. A larger value of r^* in the interval $[0, 1]$ indicates a larger overlap in sharing the treatment effect of the clinical endpoint by the joint treatment effect contributed by the other endpoints. In this regard, a value approaching one indicates that the other endpoints in the model can be treated as joint surrogates for the clinical endpoint, if clinically meaningful.

We illustrate the use of our resampling based measures through a clinical trial example in which the main endpoint is a key clinical endpoint. The interest is to find

out how much of the treatment effect of this clinical endpoint overlaps with the joint treatment effect of the other two endpoints. The two other endpoints in the trial are objective endpoints but individually of lesser importance. This example shows that this overlap is not small and also not large enough for replacing the clinical endpoint by these two objective endpoints. On the other hand, the extent of this overlap in this trial, though partial, suggests that positive findings in objective endpoints strengthen the positive finding of the clinical endpoint. For such a trial, it would be surprising to observe a statistically significant positive finding for the clinical endpoint without any positive finding individually or jointly for the other two endpoints. In this example, we also calculate a likelihood based measure introduced by Alonso et al. [1]. The developments in this paper for determining P_c and r^* are trial specific and assume that the treatment effect of the key clinical endpoint that is being explained is statistically significant in the trial. The methods of this article are intended towards understanding the nature of the treatment effects of multiple endpoints of a clinical trial, for finding whether some of these treatment effects intersect each other or are of non-overlapping nature in adding to the total treatment benefit.

The rest of the article is organized as follows. Section 2 discusses model setting and parameter estimation for the two-endpoint case with assumptions. Section 3 gives concepts about the total treatment effect and the "treatment effect explained". Section 4 introduces the probability curve P_C, addresses its properties and interpretation, and gives a bootstrapping based algorithm for computing it. Section 5 provides extension to more than two endpoints. Section 6 includes a bootstrapping based r^* correlation. Section 7 includes a numerical example, and finally, Section 8 makes some concluding remarks. Appendices A to C include derivations, a table and the steps of the bootstrapping algorithm for computational purpose.

2. Models and Estimation of Parameters

In this section, we discuss models and their parameter estimation for the two linear models in the setting of clinical trials. The first model provides a least squares estimate of the 'total treatment effect' of the primary clinical endpoint which we have called *U*. The second model is the conditional linear model of the primary clinical endpoint given the other endpoints of interest. This conditional model gives the so called the 'adjusted treatment effect' for the primary clinical endpoint. The direct difference *U* minus this 'adjusted treatment effect', which we have called *D*, is

the other statistic of interest. These two statistics are key statistics in estimating how much of the treatment effect of the primary clinical endpoint is explained by the treatment effects of the other endpoints included in the conditional model. In the following development, we assume that endpoints under consideration are continuous (or score) variables and the uses of linear models are appropriate.

2.1. Marginal model and estimation of its parameters

Consider a 2-arm clinical trial that compares a treatment to a control for a relevant clinical benefit of the treatment. Let T be a key clinical endpoint for this trial and S be any other endpoint of interest or a surrogate endpoint. Let Z be an indicator variable that takes value 1 if a patient is in the treated group and zero if a patient is in the control. Let the data on (T, Z, S) be $\{(T_{ij}, Z_{ij}, S_{ij}) : j = 1, ..., n_i \text{ and } i = 1, 2\}$ with n_1 patients in the treated group $(i = 1)$ and n_2 in the control group $(i = 2)$. Let T and S be both continuous (or score) variables with non-zero correlation, i.e., an increase or decrease in T is associated with a corresponding increase or decrease in S. In addition, we assume statistically significant treatment effect for T in the trial. The marginal treatment effect models for S and T can then be stated as:

$$S_{ij} = \mu_S + \beta_S Z_{ij} + \varepsilon_{ij}^{(S)}, \text{ with } E(\varepsilon_{ij}^{(S)}) = 0 \text{ and } \text{Var}(\varepsilon_{ij}^{(S)}) = \sigma_S^2, \qquad (2.1)$$

$$T_{ij} = \mu_T + \beta_T Z_{ij} + \varepsilon_{ij}^{(T)}, \text{ with } E(\varepsilon_{ij}^{(T)}) = 0 \text{ and } \text{Var}(\varepsilon_{ij}^{(T)}) = \sigma_T^2, \qquad (2.2)$$

where $j = 1, ..., n_i$, and $i = 1, 2$; $Z_{ij} = 1$ if $i = 1$ and 0 if $i = 2$; $\varepsilon_{ij}^{(S)}$ and $\varepsilon_{ij}^{(T)}$ are independently and identically distributed random variables. In addition, we assume that there exists a joint statistical distribution between T and S with covariance $\sigma_{TS} = \rho\sigma_T\sigma_S$.

With notations $\bar{S}_{i.}$ and $\bar{T}_{i.}$ as sample means of the endpoints S and T, respectively, for the ith treatment group $(i = 1, 2)$, least square unbiased estimators of β_S, β_T, σ_S^2, σ_T^2 and σ_{TS} are, respectively, given by

$$b_S = \bar{S}_{1.} - \bar{S}_{2.},$$

$$b_T = \bar{T}_{1.} - \bar{T}_{2.},$$

$$\hat{\sigma}_S^2 = S^2/(n_1 + n_2 - 2), \quad S^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(S_{ij} - \bar{S}_{i.})^2,$$

$$\hat{\sigma}_T^2 = S_T^2/(n_1 + n_2 - 2), \quad S_T^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(T_{ij} - \bar{T}_{i.})^2$$

and

$$\hat{\sigma}_{TS}^2 = S_{TS}/(n_1 + n_2 - 2), \quad S_{TS} = \sum_{i}^{2}\sum_{j=1}^{n_i}(T_{ij} - \bar{T}_{i.})(S_{ij} - \bar{S}_{i.}).$$

2.2. Conditional model and estimation of its parameters

A linear conditional model that expresses values of the endpoint T in terms of values of the endpoint S can be written as considered in (2.3). We use a bold lower case letter or symbol for a vector, a bold upper case letter for a matrix, and the symbol $'$ for the transpose of a vector or a matrix

$$T_{ij} = \beta_{20} + \beta_{21}Z_{ij} + \beta_{22}S_{ij} + \varepsilon_{ij}, \text{ with } E(\varepsilon_{ij} \mid S_{ij}) = 0 \text{ and } \mathrm{Var}(\varepsilon_{ij} \mid S_{ij}) = \tau^2, \tag{2.3}$$

where as above $j = 1, ..., n_i$, and $i = 1, 2$; $Z_{ij} = 1$ if $i = 1$ and 0 if $i = 2$. Model (2.3) can be conveniently written in matrix notation as $\boldsymbol{t} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. In this notation, $\boldsymbol{t}$ is a column vector with $\boldsymbol{t} = (T_{11}, ..., T_{1n_1}, T_{21}, ..., T_{2n_2})'$ of observations on the key endpoint T; $\boldsymbol{\beta} = (\beta_{20}, \beta_{21}, \beta_{22})'$ is a column vector of the model parameters, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, ..., \varepsilon_{n_1}, \varepsilon_{21}, ..., \varepsilon_{n_2})'$ is a column vector of errors; $\boldsymbol{X}$ is the design matrix with $(n_1 + n_2)$ rows and 3 columns. The matrix $\boldsymbol{X'X}$ is given by

$$\boldsymbol{X'X} = \begin{bmatrix} (n_1 + n_2) & n_1 & (n_1\bar{S}_{1.} + n_2\bar{S}_{2.}) \\ n_1 & n_1 & n_1\bar{S}_{1.} \\ (n_1\bar{S}_{1.} + n_2\bar{S}_{2.}) & n_1\bar{S}_{1.} & (S^2 + n_1\bar{S}_{1.}^2 + n_2\bar{S}_{2.}^2) \end{bmatrix}. \tag{2.4}$$

In addition, the matrix $\boldsymbol{X't} = (n_1\bar{T}_{1.} + n_2\bar{T}_{2.}, n_1\bar{T}_{1.}, (n_1\bar{T}_{1.}\bar{S}_{1.} + n_2\bar{T}_{2.}\bar{S}_{2.}))'$. Then the least square estimator of $\boldsymbol{\beta}$ is given by $\boldsymbol{b} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X't})$ and conditional variance-covariance matrix $\boldsymbol{V}_{\boldsymbol{T}\mid\boldsymbol{S}}(\boldsymbol{b}) = \tau^2(\boldsymbol{X'X})^{-1}$. Using the decomposition of $(\boldsymbol{X'X})^{-1}$ (see Appendix A) gives the following unbiased estimators and the estimators of their conditional variances and covariances:

$$b_{20} = \bar{T}_{2.} - b_{22}\bar{S}_{2.}, \quad V_{T|S}(b_{20}) = \tau^2[1/n_2 + \bar{S}_{2.}^2/S^2], \tag{2.5}$$

$$b_{21} = (\bar{T}_{1.} - \bar{T}_{2.}) - b_{22}(\bar{S}_{1.} - \bar{S}_{2.}),$$

$$V_{T|S}(b_{21}) = \tau^2[(1/n_1 + 1/n_2) + (\bar{S}_{1.} - \bar{S}_{2.})^2/S^2], \tag{2.6}$$

$$b_{22} = S_{TS}/S^2, \quad V_{T|S}(b_{22}) = \tau^2/S^2, \tag{2.7}$$

$$\mathrm{COV}_{T|S}(b_{20}, b_{11}) = \tau^2[-1/n_2 + \bar{S}_{2.}(\bar{S}_{1.} - \bar{S}_{2.})/S^2], \tag{2.8}$$

$$\mathrm{COV}_{T|S}(b_{20}, b_{22}) = -\tau^2[\bar{S}_{2.}/S^2], \tag{2.9}$$

$$\mathrm{COV}_{T|S}(b_{21}, b_{22}) = -\tau^2[(\bar{S}_{1.} - \bar{S}_{2.})/S^2] \tag{2.10}$$

and

$$\hat{\tau}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(T_{ij} - b_{20} - b_{21}I_{ij} - b_{22}S_{ij})^2/(n_1 + n_2 - 3). \tag{2.11}$$

3. Total Treatment Effect and Treatment Effect Explained

In this section, we introduce the concept of the total treatment effect of a key clinical endpoint *T* and the part of this total treatment effect "explained" by another endpoint *S*. An estimate of the total treatment effect of the endpoint *T* from the marginal model (2.2) is

$$U = \bar{T}_{1.} - \bar{T}_{2.} \tag{3.1}$$

and the adjusted treatment effect of this endpoint from the conditional model (2.3) is $b_{21} = (\bar{T}_{1.} - \bar{T}_{2.}) - b_{22}(\bar{S}_{1.} - \bar{S}_{2.})$. Note that the adjusted treatment effect is a measure of treatment effect of *T* that is adjusted for *S* through a linear model. The adjustment is to remove from the treatment of *T* the portion explained by *S*. Therefore, a direct measure of treatment effect of *T* that is explained by *S* is the difference between the two estimators, $\bar{T}_{1.} - \bar{T}_{2.}$ and $b_{21} = (\bar{T}_{1.} - \bar{T}_{2.}) - b_{22}(\bar{S}_{1.} - \bar{S}_{2.})$. This difference is

$$D = b_{22}(\bar{S}_{1.} - \bar{S}_{2.}). \tag{3.2}$$

Then the PTE is simply the ratio D/U. See Freedman et al. [5]. The joint sampling distribution of *D* and *U* contains information about the extent of dependency that

exists between the treatment-induced effects of endpoints *T* and *S*. Note that $U = b_T$ and $D = b_T - b_{21}$ are functions of least square estimators of their corresponding parameters in the models (2.2) and (2.3), respectively.

In Appendix B, we derive the variances and covariances for U and D. If error terms in the models (2.2) and (2.3) follow normal distribution, the joint distribution of *U* and *D* follows a bivariate normal distribution. The means $\mu_U = \beta_T$, $\mu_D = \beta_{22}\beta_S$ and variances and covariances are derived in Appendix B. In a general situation, the joint distribution between *U* and *D* can be evaluated by the bootstrapping based technique. We use this joint distribution information to compute PTE value by our approach.

It is important to note that *U* and *D* will enjoy the same sign as shown in Table B1 (see Appendix B). This situation will also hold when generating the bootstrapping based joint distribution of *D* and *U*. The estimate b_{22} will have the same sign as the estimate of ρ, because $b_{22} = S_{TS}/S^2$. Therefore, the sign of *D* will simply be product of the signs of b_{22} and $\bar{S}_{1.} - \bar{S}_{2.}$. The sign of *U* will be the same as that *D*, because, given the models (2.2) and (2.3), covariance between *U* and *D* is always positive (see Appendix B for the proof).

4. Bootstrapping Based Evaluation of PTE for a Trial

In this section, we discuss a method of evaluating the true PTE value on using a probability curve P_c which for its determination uses the joint sampling distribution of the sample statistics $U = \bar{T}_{1.} - \bar{T}_{2.}$ and $D = b_{22}(\bar{S}_{1.} - \bar{S}_{2.})$. These two statistics and their concepts are already addressed in the previous section.

Without loss of generality, we can assume that both *U* and *D* are positive, because if *U* and *D* are both negative, they can be converted both to positive numbers. For a given *c* in the interval $[0, 1]$, let P_c be the conditional probability

$$P_c = \Pr(D > cU \mid U > u_0), \text{ where } 0 < c < 1. \tag{4.1}$$

Notice that it is a non-decreasing conditional probability function in *c* given $U > u_0$. In this function, the value of u_0 is to form the statistical test of the primary endpoint for treatment efficacy. It is the critical value for rejecting the null hypothesis of H_0 of no treatment efficacy for the clinical endpoint *T*. That is,

$\Pr(U > u_0 \mid H_0) = 1 - \alpha/2$ for a 1-sided test at the significance level of $\alpha/2$. We make this condition of $U > u_0$ because, in a clinical trial, unless the treatment effect for the primary clinical endpoint is statistically significant (or at least marginally significant) there may not be any interest in explaining it through other endpoints. Note that P_c by definition (4.1) is basically a conditional probability curve for a binary event obtained on checking whether the extent of the treatment effect explained *D* exceeds a fraction *c* times the total treatment effect *U* or not.

Once this P_c curve is established (we give a bootstrapping based method for this), a value of *c* as c^* is then read on this curve such that this c^* is the maximum value of *c* satisfying $P_c \geq 1 - \gamma$ (e.g., for $\gamma = 0.10$ or 0.20). This c^* then serves as a lower bound for PTE with the conditional probability of $1 - \gamma$. If in a trial, for a value of $c = 0.50$ the corresponding value of $P_c = 80$ percent, then one may conclude that the chances are fairly high for concluding that the true PTE value is at least 50 percent. However, if in a trial, for a $c = 0.50$ the corresponding value of P_c happens to be fairly low, e.g., less than 10 percent, then one may conclude that chances are fairly low for the true PTE value of greater than 50 percent. Note that this lower bound is different from the conventional lower confidence interval for the PTE as it is derived under the condition $U > u_0$.

There is no interpretational difficulty in evaluating PTE on using the new measure P_c discussed above. Suppose that the unknown true value of PTE given endpoints *T* and *S* for a clinical trial is c_0. Then $c_0 = E(D)/E(U)$, where *E* stands for the expected value. As *D* and *U* are consistent estimators of $E(D)$ and $E(U)$, respectively, the conditional probability measure P_c will converge to a single-step function as the sample size is large. That is, $P_c = 1$ when $c \leq c_0$ and $P_c = 0$ when $c > c_0$. In this case, for any $\gamma < 1$, $c^* = c_0$ = PTE. Note that for the case when P_c is near 1 for all $c < 1$, then *S* will be the perfect surrogate of *T*. For other cases, *S* will be a partial surrogate or not even a partial surrogate of *T* depending on the properties of P_c curve for a trial.

The P_c curve for observed treatment effects of *T* and *S* can be directly computed by a numerical integration method (see Appendix D) under the

assumptions D and U follow a bivariate normal density and estimates of the covariance matrix of T and S, and that of the error variance in the model (2.3) can be treated as true values. This may be reasonable approach to follow a sufficiently large size trial. However, the P_c curve can easily be computed by the following bootstrapping technique. Original data can provide a pair of (u, d) values from equations (2.1) and (2.3). Then repeated bootstrapping N_a times with replacement from the original data, and using of (3.1) and (3.2), can give N_a paired data on U and D as $\{(u_j^*, d_j^*) : j = 1, ..., N_a\}$, asterisks indicating data points obtained by the bootstrapping method. Suppose that for this data,

$$A_1^* = \{\#\text{ of times } d_k^* > cu_k^* \text{ and } u_k^* > u_0\}/N_a,$$

$$A_2^* = \{\#\text{ of times } u_k^* > u_0\}/N_a,$$

then $P_c^* = A_1^*/A_2^*$ is a bootstrapping estimate of P_c for a given c. This algorithm is then repeated for different values of c for getting the P_c curve (see Appendix C). This algorithm also allows r^* calculations from a bootstrapped data $\{(u_j^*, d_j^*) : j = 1, ..., N_a\}$ by the usual product moment correlation coefficient formula.

5. General Case

5.1. Model specification and estimation

In this section, we discuss the extension of P_c measure to more than two endpoints. We assume that along with T there are several other endpoints (e.g., likely surrogates or other endpoints of interest) and one is interested in finding out as to how much of the treatment induced effect of T can be explained jointly by the treatment effects of these other endpoints. Let $S^{(1)}, S^{(2)}, ..., S^{(K)}$ be such K endpoints. The model (2.3) when updated takes the form

$$T_{ij} = \beta_{20} + \beta_{21}Z_{ij} + \beta_{22}S_{ij}^{(1)} + \beta_{23}S_{ij}^{(2)} + \cdots + \beta_{2,K+1}S_{ij}^{(K)} + \varepsilon_{ij}, \qquad (5.1)$$

with $E(\varepsilon_{ij}) = 0$ and $\mathrm{Var}(\varepsilon_{ij}) = \tau^2$ conditional on $S^{(1)}, S^{(2)}, ..., S^{(K)}$ for $j = 1, ..., n_i$ and $i = 1, 2$.

The design matrix $\boldsymbol{X}$ for this model is given by

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{n1} & \mathbf{1}_{n1} & \boldsymbol{s}_1^{(1)} & \boldsymbol{s}_1^{(2)} & \ldots & \boldsymbol{s}_1^{(K)} \\ \mathbf{1}_{n2} & \mathbf{0}_{n2} & \boldsymbol{s}_2^{(1)} & \boldsymbol{s}_2^{(2)} & \ldots & \boldsymbol{s}_2^{(K)} \end{bmatrix}. \tag{5.2}$$

In (5.2), $\mathbf{1}_{n1}$ and $\mathbf{1}_{n2}$ are column vectors of ones of dimensions n_1 and n_2, respectively; $\mathbf{0}_{n2}$ is a column vector of zeros of dimension n_2; $\boldsymbol{s}_i^{(k)}$ (for $k = 1, \ldots, K$ and $i = 1, 2$) are column vectors of observations for the k endpoints in the ith treatment group. Let $\boldsymbol{\beta} = (\beta_{20}, \beta_{21}, \beta_{22}, \ldots, \beta_{2,K+1})'$ be the parameter column vector and let $\boldsymbol{t} = (T_{11}, \ldots, T_{1n_1}, T_{21}, \ldots, T_{2n_2})'$ be the column vector of observations on the key endpoint T. Then the least square estimate $\boldsymbol{b} = (b_{20}, b_{21}, b_{22}, \ldots, b_{2,K+1})'$ of $\boldsymbol{\beta}$ is given by $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{t}$. It is easy to verify that the random measure D is given by

$$D = b_{22}(\bar{S}_{1.}^{(1)} - \bar{S}_{2.}^{(1)}) + b_{23}(\bar{S}_{1.}^{(2)} - \bar{S}_{2.}^{(2)}) + \cdots + b_{2,K+1}(\bar{S}_{1.}^{(K)} - \bar{S}_{2.}^{(K)}), \tag{5.3}$$

and $U = (\bar{T}_{1.} - \bar{T}_{2.})$ as before. Therefore, P_c can be calculated on using the bootstrapping algorithms of the previous section. In practice, as before, D should have the same sign as U. It is also desirable that each component $D_k = b_{2,k+1}(\bar{S}_{1.}^{(k)} - \bar{S}_{2.}^{(k)})$ (for $k = 1, \ldots, K$) should be of the same sign as that of U, otherwise, there can be difficulty in interpretation of the results. The conditional variance $\hat{\sigma}^2_{T|S}$ is given by (5.4):

$$\hat{\sigma}^2_{T|S} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\{T_{ij} - (b_{20} + b_{21}Z_{ij} + b_{22}S_{ij}^{(1)} + b_{23}S_{ij}^{(2)} + \cdots + b_{2,K+1}S_{ij}^{(K)})\}^2/n. \tag{5.4}$$

Example in Section 7 estimates P_c for the case when along with T there are two endpoints of interest are included in the conditional models to see as to how these two endpoints jointly explain the treatment effect of T. The above theory and calculations are meant for the continuous or score type endpoints.

6. Correlation Measure r^* and the Likelihood Measure LRF

6.1. Correlation measures r^*

The bootstrapping algorithm introduced in Section 4, gives a simple bootstrap based correlation measure r^* from the data $\{(u_j^*, d_j^*) : j = 1, ..., N_a\}$ when calculated by the usual product moment correlation coefficient formula. This r^* provides a simple measure of association between the induced treatment effect of a primary clinical endpoint and the part of it that is explained by other endpoints of interest. Note that $0 \le r^* \le 1$, because covariance between U and D is non-negative (see Appendix B). An r^* value close to 1 would indicate that the other endpoints in the conditional model are joint surrogates for the primary clinical endpoint. This r^* calculation can be made with or without the restriction that the treatment effect for the clinical endpoint is statistically significant in the trial at a specified significance level.

An r value similar to r^* value can also be calculated for the simple case of two endpoints without the bootstrapping technique on using the variances and covariances expressions for U and D given in Appendix B on replacing the unknown parameters by their corresponding sample estimates (see equations (B.2), (B.3) and (B.5)). The confidence interval for this r, however, can be calculated by the bootstrapping technique.

6.2. The LRF (likelihood reduction factor) measure

In this section, for comparison purposes we discuss another measure known as the LRF or the likelihood reduction factor (see Alonso et al. [1]). One can think of LRF as a sample estimate of a general measure of association between endpoints based on the information gain about the true endpoint on using the surrogate. However, we like the idea of using $\sqrt{\text{LRF}}$ instead. The measures r^* and $\sqrt{\text{LRF}}$ can be thought of as global measures for explaining the treatment effect of a clinical endpoint by other endpoints. That is, greater is the value of D relative to U, greater will be the values of these measures. On the other hand, P_c based PTE measure provides more specific information in this regard as a function of c. For example, a value of r^* or $\sqrt{\text{LRF}}$ of 0.8 would suggest that a significant portion of the treatment effect of the clinical endpoint is being explained by other endpoints, but it does not convey the information as to how much and with what probability. The

measure P_c addresses this later question conditionally given that the result in the trial for the main endpoint T is statistically significant.

In the following, we describe the calculations of this LRF. Let the errors for the marginal model (2.2) and the conditional model (2.3) be normally distributed as $N(0, \sigma_T^2)$ and $N(0, \sigma_{T|S}^2)$, respectively, for $j = 1, ..., n_i$ and $i = 1, 2$. Then the likelihood ratio statistics Λ for evaluating the added information by the model (2.3) is given by

$$\Lambda = \left[\frac{\hat{\sigma}_{T|S}^2}{\hat{\sigma}_T^2} \right]^{n/2}, \quad \text{where } n = n_1 + n_2, \tag{6.1}$$

$$\hat{\sigma}_T^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \{T_{ij} - (\hat{\mu}_T + b_T I_{ij})\}^2 / n$$

and

$$\hat{\sigma}_{T|S}^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \{T_{ij} - (b_{20} + b_{21} I_{ij} + b_{22} S_{ij})\}^2 / n.$$

Estimators $\hat{\mu}_T$, b_T, b_{20}, b_{21} and b_{22} are maximum likelihood estimators. They are the same as given in Section 3. Therefore, if G^2 is the log-likelihood ratio statistics, the LRF, as defined by Alonso et al. [1], is given by

$$\text{LRF} = 1 - \exp(-G^2/n) = \hat{\sigma}_{T|S} / \hat{\sigma}_T. \tag{6.2}$$

7. A Numerical Example of an Acne Trial

Acne clinical trials, for evidence of efficacy of a treatment, typically analyze three endpoints after 11 weeks of treatment. These endpoints are: (1) IGA (investigator global assessment) on an improvement scoring scale, (2) inflammatory lesion counts, and (3) non-inflammatory lesion counts. Clinicians consider IGA an important clinical endpoint. The IGA endpoint is expected to consider a wider aspect of the disease as it considers facial lesions and other clinical aspects, such as, the extent of facial itching, burning, etc. Such acne trials, for a clinically meaningful evidence of treatment benefit, generally require that, the IGA endpoint shows convincing evidence of treatment efficacy, and the two lesion count endpoints also show some evidence of treatment efficacy.

It is of interest to know as to how much of the treatment efficacy in the IGA endpoint is explained jointly by the treatment effects of the two lesion count endpoints. In this regard, we first briefly describe the results a specific vehicle-controlled acne clinical trial without mentioning the name of the trial and the name of the product studied. We then illustrate the determination of P_c and PTE for answering, for this specific trial, as to how much of the IGA treatment effect is being explained by the joint treatment effects of the other two endpoints by the linear modeling technique we have presented in this article. Our results show that for this trial treatment effects for the two lesion count endpoints do explain a certain percentage of the treatment effect of the IGA endpoint but this percentage is not large enough to ignore requiring a convincing result of efficacy for this endpoint for a clinically meaningful benefit of the study treatment.

This specific acne trial was a double-blind randomized trial in which the treatment arm was compared to a vehicle arm after 11 weeks of treatment. After excluding a very few dropouts, there were $n_1 = 93$ patients in the treated arm and $n_2 = 56$ patients in the vehicle arm. The randomization scheme for the trial was $2:1$, i.e., for every 2 patients assigned to a treated arm, one patient was assigned to the vehicle arm. In this trial, each of the three endpoints, i.e., the IGA and the two lesion count endpoints, showed evidence of treatment efficacy at the significance level of $\alpha = 0.01$, and there was evidence of associations among the three endpoints (see box plots and scatter plots in Figure 1). In this example, T stands for the IGA and $S^{(1)}$ and $S^{(2)}$ for the non-inflammatory and inflammatory lesion counts, respectively. The Q-Q normal plot (Figure 2) showed that the residuals for the conditional model in (5.1) followed normal density except for a few points in the tail.

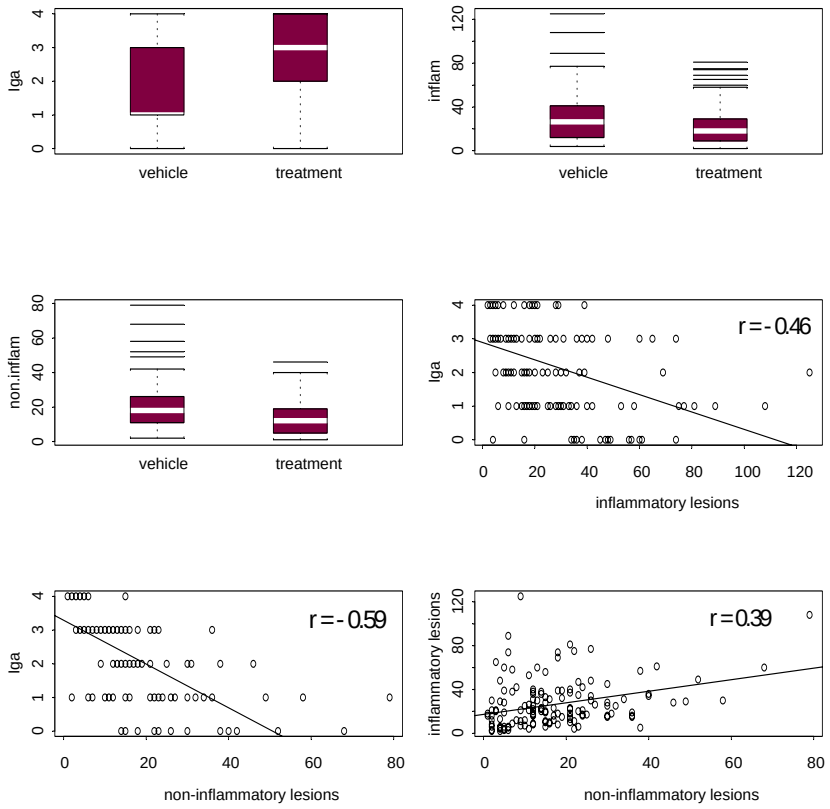

Figure 1 (Acne trial)**.** Box plots and scatter plots for the three endpoints, investigator global assessment (Iga), inflammatory lesion counts (inflam) and non-inflammatory lesion counts (non.inflam).

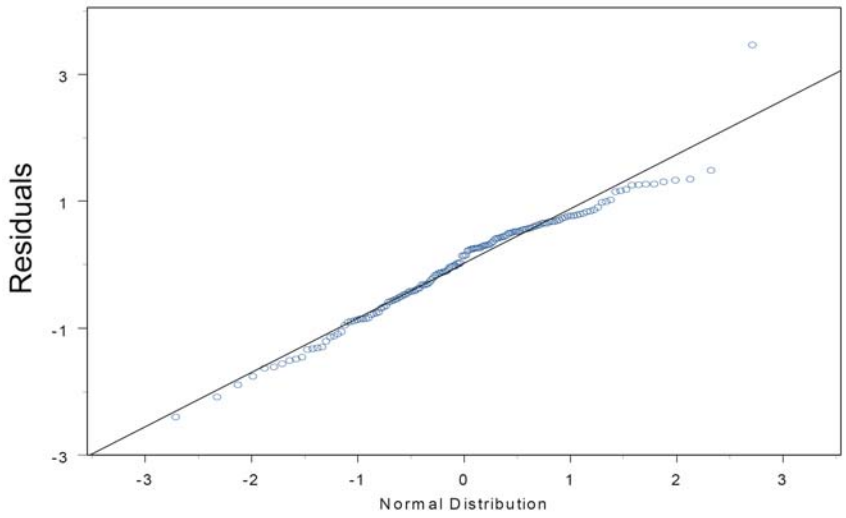

Figure 2. *Q-Q* normal plot for the residuals of the conditional model (5.1) for the acne trial example shows that, except for a few points in the right tail, residuals pretty much follow normal distribution.

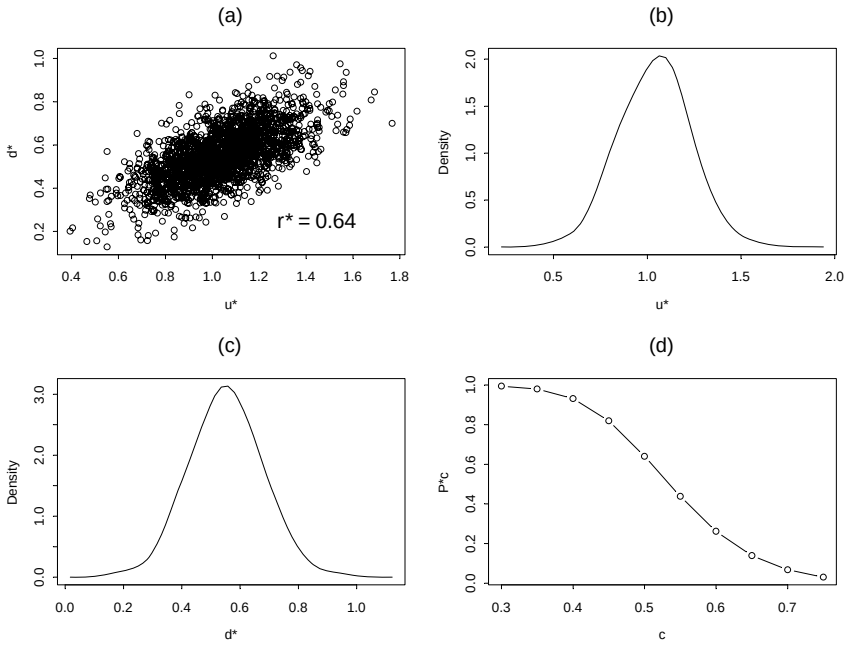

Figure 3 (Acne trial)**.**

- (a) Scatter plot (u^*, d^*) is based on 1,997 bootstraps from a total of 2,000 bootstraps drawn with replacement from the original acne trial data. In a pair (u^*, d^*), u^* is a treatment effect for the endpoint T (= Investigator global assessment), and d^* is the extent of it explained jointly by the endpoint $S^{(1)}$ (non-inflammatory lesion counts) and $S^{(2)}$ (inflammatory lesion counts). Correlation $(u^*, d^*) = r^* = 0.64$.
- (b)-(c) Density plots of u^* and d^* based on the 1,997 bootstraps in (a).
- (d) Plot of P_c^* versus c. In this plot, points (c, P_c^*) were obtained by the Bootstrapping Algorithm 1 (Appendix C).

Figures 3(a) to 3(c) display some properties of a data set on (U, D) generated by the bootstrapping of the original data by using equation (5.3) for D and equation (3.1) for U. A scatter plot of this data set indicates association between the two statistics U, D with $r^* = 0.64$; U is a treatment effect measure for T and D is the part of the treatment effect of T that is explained jointly by $S^{(1)}$ and $S^{(2)}$. Density plots for D and U show that they are about normally distributed. Values of P_c^* for different c were from the bootstrapping algorithm of Appendix C (Table C1). Figure 3(d) shows a plot of P_c^* versus c. In Figure 3(a), the scatter plot (u^*, d^*) is based on 1,997 bootstraps from a total of 2,000 bootstraps drawn with replacement from the original acne trial data. In three bootstraps, u^* did not meet statistical significance at the 0.025 level by 1-sided test, and in one case d^* exceeded u^*. In a pair (u^*, d^*), u^* is a treatment effect for the endpoint T (IGA) and d^* the extent of it that is explained jointly by the endpoint $S^{(1)}$ (non-inflammatory lesion counts) and $S^{(2)}$ (inflammatory lesion counts). The value of r^* without the treatment effect restriction based on 10,000 bootstraps was $r^* = 0.625$ and the value of $\sqrt{\text{LRF}}$ was 0.518 with 95% confidence interval of (.436, 618). This confidence interval was the bootstrapping method.

Values of P_c^* at $c = 0.40$ and 0.45 are 93 and 82 percent, respectively. The value of P_c^* at $c = 1/2$, although not in the 80 to 90 percent range, is 64 percent

which is greater that 50 percent. However when the value of c is 0.70 the value P_c^* is relatively very small. Therefore, it is less likely that the true value of PTE is greater than 70 percent but there is a high probability that it is at least 45 percent. For the endpoints S_1 and S_2 to be joint surrogates for T, one would expect a high value of P_c^* at $c = 1/2$ (e.g., 90 percent) and value P_c^* of greater than 50 percent at $c = 3/4$. These results for this trial indicate that there is some overlap between treatment effect of T and that contributed jointly by the treatment effects of S_1 and S_2. However, this overlap is not small and also not large enough for replacing the clinical endpoint by these two objective endpoints. On the other hand, the extent of this overlap suggests that positive findings in objective endpoints strengthen the positive finding of the clinical endpoint. In addition, the IGA endpoint justifiably provides some extra information regarding the treatment effect than those provided by the two lesion count endpoints.

8. Some Concluding Remarks

The results of this article can be useful for Phase III confirmatory clinical trials and also for early phase exploratory trials. Early phase trials explore the effects of a test compound on many different types of endpoints. Some are for capturing pharmacologic or biologic effects of the test treatment. These endpoints are different from the clinical endpoints but early trials are easier to do with these endpoints. There is generally an interest in learning if the treatment effects for some of these non-clinical endpoints can be used to explain the treatment effect for certain clinical endpoints. The proposed method can be used for this purpose if appropriate data are available for the test compound or for compounds of similar mechanism of actions.

Late phase confirmatory trials generally include primary clinical endpoints for addressing primary objectives of a trial and a few key secondary clinical endpoints for either strengthening the results of the primary endpoints or for additional treatment benefit. In this regard, it is of interest to know how the treatment effects for these endpoints overlap in explaining each other. The proposed methods can be used for this purpose. The results obtained can be useful in claiming an extra benefit for a secondary endpoint if this secondary endpoint had a beneficial treatment effect for the experimental treatment with little or no explaining ability of the beneficial treatment effects already found for the primary endpoints. On the other hand, if this secondary endpoint treatment effect is able to explain most of the treatment effects

of the primary endpoint, then this secondary endpoint result would strengthen the results of the primary endpoints but may not qualify for an extra benefit of the treatment.

Methods proposed can be further improved on including baseline covariates in the marginal and conditional models of this article. This can reduce the error variances of these models, and thus can further refine the conditional probability curve of this article for evaluating PTE and in providing the r^* estimate. These methods, however, are for continuous type endpoints and assume that the use of linear models is justified. The later can be checked by investigating the residuals of the fitted models. These methods may not be appropriate for other types of endpoints such as binary and for other types of modeling.

It is worth mentioning that there may be some other clinical or biological phenomenon for which the two endpoints do not associate directly, but one of them could be capable of capturing an effect of the other. This can happen in a clinical trial when two endpoints for a given drug are not on the same linear causal pathway but link through an indirect network of pathways which may or may not be identifiable depending on the complexity of the mechanism of the action of the drug and the site of the action of the drug. For example, the test drug may breakdown within the human body after some chemical reactions into multiple active metabolizes which may form an indirect complex bio-chemical metabolic network for targeting the two endpoints in inducing treatment effects. The methods proposed in this article are not meant to capture this type of complex dependence between the induced treatment effects of the two or more endpoints. A meta-regression approach of multiple similar studies, however, may be able to better capture such dependence. This is an area of further research.

Disclaimer

The views expressed in this article are of the authors and not necessarily of the U.S. Food and Drug Administration.

Appendix A

On partitioning the matrix $\boldsymbol{X}'\boldsymbol{X}$ as

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{11} = n_1 + n_2$, $A_{12} = [n_1 \quad (n_1\bar{S}_{1.} + n_2\bar{S}_{2.})]$ and A_{22} is the matrix formed on deleting the first row and first column of the matrix $\boldsymbol{X'X}$ in (2.4). Let the matrix $\boldsymbol{C}$ be the inverse of $\boldsymbol{X'X}$ with sub-matrices C_{11}, C_{12}, C_{21} and C_{22} corresponding to sub-matrices A_{11}, A_{12}, A_{21} and A_{22}, respectively. Then

$$C_{11} = A_{11.2}^{-1} = (A_{11} - A_{12} \; A_{22}^{-1} \; A_{21})^{-1}, \qquad C_{12} = -A_{11.2}^{-1} \; A_{12} \; A_{22}^{-1}$$

and

$$C_{22} = A_{22}^{-1} \; A_{21}(A_{11.2}^{-1} \; A_{12} \; A_{22}^{-1}) + A_{22}^{-1}.$$

On evaluating the above C sub-matrices,

$$C_{11} = [1/n_2 + \bar{S}_{2.}^2/S^2], \qquad C_{12} = [-1/n_2 + \bar{S}_{2.}(\bar{S}_{1.} - \bar{S}_{2.})/S^2, \; -\bar{S}_{2.}/S^2]$$

and

$$C_{22} = \begin{bmatrix} (1/n_1 + 1/n_2) + (\bar{S}_{1.} - \bar{S}_{2.})^2/S^2, & -(\bar{S}_{1.} - \bar{S}_{2.})/S^2 \\ -(\bar{S}_{1.} - \bar{S}_{2.})/S^2, & 1/S^2 \end{bmatrix}.$$

Appendix B

Variances and covariances of D and U

$\text{Var}(D)$ follows from the standard formula:

$$\text{Var}(D) = E_S V_{T|S}[b_{22}(\bar{S}_{1.} - \bar{S}_{2.})] + V_S E_{T|S}[b_{22}(\bar{S}_{1.} - \bar{S}_{2.})].$$

As $E_{T|S}[\hat{\beta}_{22}] = \beta_{22}$, the second term reduces to

$$V_S[\beta_{22}(\bar{S}_{1.} - \bar{S}_{2.})] = \beta_{22}^2\sigma_S^2(1/n_1 + 1/n_2).$$

The first term is $\tau^2 E_S[(\bar{S}_{1.} - \bar{S}_{2.})^2/S^2]$ from (2.7). This term, however, reduces to $E_S[(\bar{S}_{1.} - \bar{S}_{2.})^2]$. $E_S[1/S^2]$ under the assumption that observations in each treatment arm are independently and identically distributed as normal random variables. Also, under this normal distribution assumption, $S^2/\sigma_S^2 = \chi^2_{n_1+n_2-2}$, a chi-square random variable with $n_1 + n_2 - 2$ degrees of freedom with $E(1/\chi^2_{n_1+n_2-2})$

$= 1/(n_1 + n_2 - 4)$. Therefore, the first term $E_S V_{T|S}[b_{22}(\bar{S}_{1.} - \bar{S}_{2.})]$ reduces to

$$\tau^2(\beta_S^2/\sigma_S^2)/(n_1 + n_2 - 4) + \tau^2(1/n_1 + 1/n_2)/(n_1 + n_2 - 2).$$

Consequently,

$$\mathrm{Var}(D) = \beta_{22}^2\sigma_S^2(1/n_1 + 1/n_2) + \tau^2 E_S[(\bar{S}_{1.} - \bar{S}_{2.})^2/S^2] \tag{B.1}$$

$$= \beta_{22}^2\sigma_S^2(1/n_1 + 1/n_2) + \tau^2(\beta_S^2/\sigma_S^2)/(n_1 + n_2 - 4)$$

$$+ \tau^2(1/n_1 + 1/n_2)/(n_1 + n_2 - 2). \tag{B.2}$$

The result in (B.1) is general, but the result (B.2) is based on the normal assumption. The quantity $E_S[(\bar{S}_{1.} - \bar{S}_{2.})^2/S^2]$ in (B.1) for non-normal situations can be found, for example, by the bootstrapping technique. For sufficiently large sample sizes, the last term in variance formula (B.2) can be ignored as it is of order $O(1/n^2)$,

$$\mathrm{Var}(b_{21}) = E_S V_{T|S}(b_{21}) + V_S E_{T|S}(b_{21})$$

$$= \tau^2[(1/n_1 + 1/n_2) + E_S[(\bar{S}_{1.} - \bar{S}_{2.})^2/S^2] + 0.$$

However,

$$\mathrm{Var}(b_{21}) = \mathrm{Var}(U) + \mathrm{Var}(D) - 2\,\mathrm{COV}(U, D).$$

Therefore,

$$\mathrm{COV}(U, D) = \{\mathrm{Var}(U) + \mathrm{Var}(D) - \mathrm{Var}(b_{21})\}/2.$$

This simplifies to

$$\mathrm{COV}(U, D) = \beta_{22}^2\sigma_S^2(1/n_1 + 1/n_2) \tag{B.3}$$

$$= \rho^2\sigma_T^2(1/n_1 + 1/n_2). \tag{B.4}$$

The expression (B.4) follows under the normal distribution assumption which gives $\beta_{22}^2 = \rho^2\sigma_T^2/\sigma_s^2$. The variance $\mathrm{Var}(U)$ can also be written as

$$\mathrm{Var}(U) = \sigma_T^2(1/n_1 + 1/n_2) = (\beta_{22}^2\sigma_S^2 + \tau^2)(1/n_1 + 1/n_2). \tag{B.5}$$

Table B1. Justification for the same sign for *U* and *D*

Cases	Est. of ρ	b_{22}	$\bar{S}_{1.} - \bar{S}_{2.}$	$U = \bar{T}_{1.} - \bar{T}_{2.}$	$D = b_{22}(\bar{S}_{1.} - \bar{S}_{2.})$
I	+	+	+	+	+
II	+	+	–	–	–
III	–	–	+	–	–
IV	–	–	–	+	+

Appendix C

Bootstrapping Algorithm 1

Step 1:

(a) Check that both *T* and *S* enjoy evidence of some treatment effect and ρ is different from zero.

Step 2:

(a) Calculate $U = \bar{T}_{1.} - \bar{T}_{2.}$ and $D = b_{22}(\bar{S}_{1.} - \bar{S}_{2.})$ from the original data $\{(T_{ij}, I_{ij}, S_{ij}) : j = 1, ..., n_i \text{ and } i = 1, 2\}$.

(b) Check that both U and D have the same sign. If both U and D are negative numbers then change them to both positive numbers.

Step 3:

(a) Create a new data $\{(T_{ij}^*, I_{ij}^*, S_{ij}^*) : j = 1, ..., n_i \text{ and } i = 1, 2\}$ by bootstrapping the original data by the method of bootstrapping with replacement.

(b) Calculate new statistics U^* and D^* from this new data using the same formulae as for *U* and *D*.

(c) Retain U^* and D^* if both have the same sign otherwise discard these values and repeat 3(a)-3(b).

(d) If both U^* and D^* are negative numbers then change them to both positive numbers.

Step 4:

(a) Repeat Step (3) N_a times giving paired data $\{(U_k^*, D_k^*) : k = 1, ..., N_a\}$.

Step 5:

(a) Calculate $A_1^* = \{\#\text{ of times } D_k^* > cU_k^* \text{ and } U_k^* > u_0\}/N_a$, $A_2^* = \{\#\text{ of times } U_k^* > u_0\}/N_a$.

(b) Calculate $P_c^* = A_1^*/A_2^*$ and calculate r^* value from the bootstrapped data $\{(u_j^*, d_j^*) : j = 1, ..., N_a\}$ by the product moment correlation coefficient formula.

Table C1 (Acne trial)**.** Values of (c, P_c^*) from the bootstrapping distribution of statistics (U, D)

c	P_c^*	c	P_c^*
0.30	0.994	0.35	0.980
0.40	0.931	0.45	0.819
0.50	0.640	0.55	0.438
0.60	0.261	0.65	0.138
0.70	0.067	0.75	0.030

The given trial was re-sampled 2,000 times satisfying certain condition (see Bootstrapping Algorithm 1)

Appendix D

Calculation of P_c by numerical integration method

Calculation of P_c readily follows from the bivariate normal distribution of U and D with means (μ_U, μ_D) and σ_D^2 (given in (B.1)), σ_{DU} (given in (B.4)) and σ_U^2 (given in (B.5)). In which case, the conditional distribution of D given $U = u$ is normal with mean $\mu_D + \rho_{DU}(\sigma_D/\sigma_U)(u - \mu_U)$, and variance $(1 - \rho_{DU}^2)$. It is easy to verify that

$$P_c = 1 - \frac{I(u_0, c)}{1 - \Phi((u_0 - \mu_U)/\sigma_U)}, \tag{D.1}$$

where the integral

$$I(u_0, c) = \int_{(u_0 - \mu_U)/\sigma_U}^{\infty} \varphi(x)\Phi\left(\frac{x(c\sigma_U - \rho_{DU}\sigma_D) + (c\mu_U - \mu_D)}{\sigma_D\sqrt{1 - \rho_{DU}^2}}\right)dx. \tag{D.2}$$

Functions $\varphi(x)$ and $\Phi(x)$ are, respectively, the density and cumulative distribution functions of a standard normal random variable *X*. The integral $I(u_0, c)$ in (D.2) can be readily calculated by a standard numerical method. One can use the SAS/IML function QUAD (1999) for calculating this integral on assuming the unknown parameters.

References

[1] A. Alonso, G. Molenberghs, T. Burzykowski, D. Renard, H. Geys, Z. Shkedy, F. Tibaldi, J. C. Abrahantes and M. Buyse, Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints, Biometrics 60(3) (2004), 724-728.

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

[3] M. Buyse and G. Molenberghs, Criteria for the validation of surrogate endpoints in randomized experiments, Biometrics 54(3) (1998), 1014-1029.

[4] A. Chakravarty, Regulatory aspects in using surrogate markers in clinical trials, The Evaluation of Surrogate Endpoints, T. Burzykowski, G. Molenberghs and M. Buyse, eds., Springer, 2005, pp. 13-51.

[5] L. S. Freedman, B. I. Graubard and A. Schatzkin, Statistical validation of intermediate endpoints for chronic diseases, Stat. Med. 11(2) (1992), 167-178.

[6] SAS Institute Inc., SAS/IML User's Guide, Version 8, SAS Institute Inc., Cary, NC, 1999.

[7] C. J. Weir and R. J. Walley, Statistical evaluation of biomarkers as surrogate endpoints: a literature review, Stat. Med. 25(2) (2006), 183-203.