Journal of Statistical Theory and Application ISSN 1538-7887
Volume 10, Number 4, 2011, pp.633-651

BAYESIAN EVALUATION OF BIOMARKERS AS SURROGATE ENDPOINTS AND ITS APPLICATION IN LONG-TERM EFFECTS OF SKIN LESIONS DUE TO MUSTARD GAS

Shohreh Jalaie[1], Soghrat Faghihzadeh[2*], Farzad Eskandari[3], Tooba Ghazanfari[4], and M. Reza Meshkani[5]

[1]Faculty of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran jalaeish@sina.tums.ac.ir

[2]Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University,Tehran, Iran faghihz@modares.ac.ir

[3]Department of Statistics and Mathematics, Allameh Tabatabai University, Tehran, Iran f-eskandari@cc.sbu.ac.ir

[4]Department of Immunology, Medical Faculty, Shahed University, Tehran, Iran tghazanfari@yahoo.com

[5]Department of Statistics; School of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran mrmeshkani@gmail.com

Abstract

The paper aims at undertaking a Bayesian evaluation of surrogate endpoints which has the potential to decrease the exposure of patients while at the same time being cost-effective. As such, the current research, with an individual-level evaluation, intends to introduce a new Bayesian criterion based on some of the previous research works. Taking into account the surrogate and the true endpoint variables, the statistics that derived for the exponential family of distributions particularly is for a binomial one. In due process, the paper shows the relationship between the Bayesian likelihood reduction factor (LRF_B) and its frequentist counterpart (LRF). Finally, to show some intuition in the nature of LRF_B, it has been applied in a multi-center simulation and a uni-center real example. In real data, LRF_B is carried out to evaluate the immunological factor *IL-18 BPa Serum* in the long-term effect of skin lesions of people who were exposed to mustard gas.

Key words: Clinical Endpoint; Criterion; Individual Level; Surrogate; Bayes Factor

* Correspondence to: Soghrat Faghihzadeh, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. faghihz@modares.ac.ir

1. Introduction

Gradual application of biomarkers as the surrogate endpoints in medical and pharmaceutical research has stemmed out of various reasons, a complete list of which can be found in a review paper by Weir and Walley (2006). In fact, Prentice (1989), in his seminal paper, for the first time highlighted the statistical evaluation of the surrogates. As such, he proposed the first criterion to test the hypothesis. Freedman et al. (1992) extended the idea further to estimate the degree of validity of a surrogate endpoint.

Using a potential outcome frameworks introduced by Frangakis and Rubin (2002), later researchers like Li and Elliott (2010) proposed a Bayesian estimation to evaluate the causal probabilities associated with the cross-classification of potential outcomes of S and T when both are binary. To evaluate the potential outcomes, they preferred a log-linear model.

With respect to meta-analysis, Buyse et al. (2000) followed the Prentice criterion to evaluate the validity of surrogate. Alonso et al. (2004) found a lack of unified approach for applying these two criteria when according to them; neither the biomarker nor the true endpoint is normally distributed. Consequently, they proposed a likelihood reduction factor (LRF) which is not restricted to the normal variables. This measure of individual-level association could be used under any generalized linear model for a single trial or meta-analysis. In other words, with a generalized linear model, the LRF is derived from modeling the effect of treatment on the biomarker and the true endpoints. It can be scaled between 0 and 1 while at the same time does not have the PTE weaknesses i.e. lack of interaction between the surrogate and the true endpoints as alluded above.

Following the Bayesian approach, the current paper uses the LRF to evaluate biomarkers as the surrogate endpoints. Indeed, the Bayesian LRF denoted by LRF_B benefits from the prior knowledge on the situation under study. The proposed study is mainly concerned about the exponential family of distributions for the surrogate and the true endpoint variables.

The rest of the paper is as follows. Section 2 deals about methods applied during the course of study and how the LRF_B criterion is derived for the exponential family. Section 3 presents a simulation study with the aim to measure its performance. Section 4 illustrates the proposed methodology incorporating a real example concerning patients suffering from the mustard gas exposure. Finally, section 5 concludes with a discussion.

2. Methods

In order to lay down the foundation, the authors briefly took into account the prerequisite materials for the proposed study. As mentioned before, Alonso et al. (2004) employed two

generalized linear models to explain the effects of treatment on the surrogate (S) and true (T) endpoints.

2.1 The model

Taking into account the ideas of Alonso et al, at the i_{th} trial, the two following generalized linear models (a reduced and a full model) were considered:

$$g_T\{E[T_{ij}|Z_{ij}]\} = X_{ij}^{(1)}\beta_i^{(1)} \quad (1)$$

$$g_T\{E[T_{ij}|Z_{ij},S_{ij}]\} = X_{ij}^{(2)}\beta_i^{(2)} \quad (2)$$

Where it is assumed that i=1… N trials are available in the i_{th} of which j=1, … n_i subjects are enrolled. S_i denotes the variables of interest; j_{th} is the value of surrogate marker which subjected to i_{th} trial with any kind of distribution. The value of the true endpoint T_j is corresponding to S_j and Z_j, and Z_j=1 is a binary variable denoting the treatment group. The vectors used in the equations (1) and (2) represent the explanatory variables and their coefficients in a way that $X_{ij}^{(1)} = (x_{0ij}, x_{1ij},...,x_{(q-1)ij}) = (1, Z_{ij}, x_{2ij},\ldots, x_{(q-1)ij})$: $1\times q$ which is a corresponding vector of covariates such as $X_{ij}^{(2)} = (X_{ij}^{(1)}, S_{ij})$: $1\times(q+1)$ and $\beta_i^{(1)} = (\beta_{0i}^{(1)},....,\beta_{(q-1)i}^{(1)})^T$, $\beta_i^{(2)} = (\beta_{0i}^{(2)},....,\beta_{(q-1)i}^{(2)},\beta_{qi}^{(2)})^T$.

The link function g(.) is appropriately chosen to relate the explanatory variables the conditional mean of the true endpoint variable.

Based on this formulation, Alonso et al. (2004) presented the likelihood reduction factor (LRF) as:

$$LRF = 1 - \frac{1}{N}\sum_i \exp\left(-\frac{G_i^{\,2}}{n_1}\right) \quad (3)$$

Where N is the number of trials, each involving n_i subjects, and G_i^2 is the likelihood test ratio for comparing models (1) and (2) within trial i. The LRF equals $R^2_{\ indiv}$ in a special case of the normally distributed surrogate and the true endpoints.

It is worth mentioning that the Bayesian approach is beneficial with respect to its simplicity, exactness and coherency. And, the Bayesian inference provides a formal mechanism for incorporating and updating the prior knowledge (Link and Barker, 2010). As such, the Bayesian

approach could be useful while dealing with binary and count data with asymmetric regression coefficients. In other words, in clinical studies, the Bayes factor seems to be more appropriate criterion than the non-Bayesian statistics. Consequently, a new criterion (LRF_B) is proposed to evaluate the validity of surrogate biomarkers at individual level in a single or multiple trials. This criterion is based on the Bayes factor as well as the approximation of Schwartz Criterion (BIC) (Robert 2001).

2.2 The Bayesian approach

Considering the models (1) and (2) wherein the true endpoint variable j_{th} subjected to i_{th} trial, T_j, it is assumed to follow an exponential family distribution. The generic variable y is said to have an exponential family distribution if its probability is expressed as:

$$P(y|\theta) = h(y)\exp\{\sum_{i=1}^{s}\eta_i(\theta)V_i(y) - B(\theta)\} \quad (4)$$

Where, $\eta = (\eta_1, \cdots, \eta_s)^T$ is the natural parameter vector and $V = \{V_1, ..., V_s\}$ is the vector of sufficient statistics. The quantities T_j and β in (1) and (2) are identified by y and η in (4), respectively.

To perform a Bayesian analysis, one needs to specify priors for elements β. The proposed study assumed some independent priors for β_j's. Under the Bayesian paradigm, since the prior distributions for parameters are assumptions, the effects of priors on inference can be evaluated by trying some of them (Link and Barker, 2010). In this study, normal priors have been considered. Thus, one option is $\beta_i \sim N(\mu_i, \sigma_i^2)$. In case of lack of reliable prior knowledge, one may use a non-informative prior. In fact, the Bayes factor as a model criterion has been advocated by previous researchers such as Kass and Raftery (1995), and Berger and Perrichi (1996). For comparing the full and the reduced models, the Bayes factor BFB_{21} was considered which is defined in each center (i) as:

$$BF_{i21} = \frac{P(T_i|M_2)}{P(T_i|M_1)} \quad (5)$$

Where $p(T_i|M_k)$ is the marginal distribution of T_i under model k (k=1, 2) that can be shown as:

$$p(T_i|M_k) = \int p(T_i|\beta, M_k)\pi(\beta|M_k)d\beta \quad (6)$$

Where $\pi(\beta|M_k)$ is the prior function on the β. In the exponential family, a closed form of (6) is available for the conjugate priors. In general, it does not have a closed form, but a good approximation can be provided via Laplace expansion, Tierney and Kadane (1986), Erkanli (1994), and Kadane and Lazar (2004), in a way that for ith center and kth model, the equation (6) via Laplace expansion is approximated as:

$$P(T_i|M_k) \approx (2\pi)^{\frac{q_k}{2}} \left|\hat{\Sigma}_k\right|^{\frac{1}{2}} P(T_i|\beta, M_k)\pi(\beta|M_k) \qquad (7)$$

Where β is substituted by its MLE, q_k is the dimension of β under the model k, and $\hat{\Sigma}_k$ is the inverse of the corresponding negative Hessian matrix. Alternatively, $\hat{\Sigma}_k$ can be the inverse of the observed or expected Fisher information matrix, evaluated at the maximum likelihood estimation of β. According to Kass et al. (1990), this approximation is good enough for sample sizes (n) of greater than $20q_k$ up to $o(n^{-1})$.

Applying (7) in the numerator and denominator of (5), the logarithm of Bayes factor for i_{th} center is approximated as:

$$\log BF_{i12} \approx \log\lambda_{n_i,12} - \frac{1}{2}(q_1 - q_2)\log(n_i) - \frac{1}{2}(q_2 - q_1)\log(2\pi) + \frac{1}{2}\log(\frac{|\Sigma_1|}{|\Sigma_2|}) + \log(\frac{\pi_1(\hat{\beta}^{(1)})}{\pi_2(\hat{\beta}^{(2)})}) \qquad (8)$$

Where, $\lambda_{n_i,12} = P(T_i|\hat{\beta}, M_1) \big/ p(T_i|\hat{\beta}, M_2)$ is the likelihood ratio of model 1 to model 2 for center i with sample size n_i. Thus, there found to be close relations between BF and λ_n. In a special case where $\pi_k(\beta^{(k)})$ is degenerated at $(\beta^{(k)})$ and $P(MB_{1B})=P(MB_{2B})=\frac{1}{2}$, the Bayes factor reduces the likelihood ratio, implying that the frequentist test is a special case of the Bayesian. In (3), G_i^2 is the same as $\lambda B_{ni,12B}$ for the i_{th} center. In fact, it is equivalent to $BFB_{i,12}$.

Schwartz's Bayesian information criterion (BIC) is yet another vastly used criterion for the model choice as suggested by Smith and Spiegelhalter (1980), Nishii (1984), Haughton (1988), Kass and Wasserman (1995). In the present study, it is denoted by 'Sch'. Here, Schwartz's Bayesian information criterion for i_{th} center is:

$$\mathrm{Sch}_i = \log(\lambda_{n_i,12}) - \frac{1}{2}(q_1 - q_2)\log(n_i) \quad (9)$$

Kass and Wasserman (1995) established that for the prior $\pi(\beta) \sim N(0,\Sigma)$ with $|\Sigma| = |I(\beta)|$, the Schwartz criterion can be used as an asymptotic equivalent for log(BF) with the relative error of $o(n^{-1/2})$. Thus, for the prior $N(0,\Sigma)$, the Schwartz criterion can be viewed as a good approximate of log(BF).

The current paper is searching for an exact link between BF and 'Sch' and hence: substituting (9) in (8) in each center, one can have

$$\log \mathrm{BF}_{12} \approx \mathrm{Sch} - \frac{1}{2}(q_2 - q_1)\log(2\pi) + \frac{1}{2}\log(\frac{|\Sigma_1|}{|\Sigma_2|}) + \log(\frac{\pi_1(\hat{\beta}^{(1)})}{\pi_2(\hat{\beta}^{(2)})}) = \mathrm{Sch} + K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) \quad (10)$$

Where $K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$ is remainder terms in (10),

$$K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) = -\frac{1}{2}(q_2 - q_1)\log(2\pi) + \frac{1}{2}\log\frac{|\Sigma_1|}{|\Sigma_2|} + \log(\frac{\pi_1(\hat{\beta}^{(1)})}{\pi_2(\hat{\beta}^{(2)})}) \quad (11)$$

Consequently, one obtains

$$G_i^{\,2} = 2\log \mathrm{BF}_{i,21} - (q_1 - q_2)\log(n_i) - K'(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) \quad (12)$$

Such that $K'(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) = -2K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$. As the equation (12) shows the relations between G^2 and log (BF_{12}), the paper proceeds to use it in evaluating the surrogate endpoints.

Prentice (1989) believed that one requires $p(T|S,Z) = p(T|S)$ while evaluating the validity of the surrogate. And, based on Kent's (1983) suggestion, the LRF measures partial correlation between S and T, accounting for the effect of Z, fitting the models (1) and (2) by equation (3). In the Bayesian approach, this correlation is computed through the comparison of models in each center by BF_i. Since $K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) = o(n^{-1/2})$ as $n \to \infty$, it can be shown that

$$\rho(T,S|Z) = LRF_B = 1 - \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\frac{2\log(BF_{i21})}{n_i} + \frac{\log(n_i)^{q_1-q_2}}{n_i}\right) \quad (13)$$

LRF_B like LRF can be employed when multiple markers are studied as surrogates and when there is a nonlinear relation between S and T.

Close examination of the equation (13) reveals that when the smallest $n_i \to \infty$, the term $\log(n_i)/n_i \to 0$. It can be ignored for moderate to large sample sizes. Also, the interpretation of LRF_B is similar to that of BF_{21}. Indeed, if the model (2) is not better than the model (1), it implies that $BF_{i21} \approx 1$, $\log(BF_{i21}) = 0$ for all i and hence; LRF_B=0. If $\rho(T,S|Z) = 0$, it implies that S is not a good surrogate for T. The proposed study proves the situation Lim $LRF_B = LRF$ by a theorem mentioned in appendix. However, if BF_{i21} is large enough, $LRF_B \to 1$, favoring S as a good surrogate.

2.3. Exact Value of Bayesian Likelihood Reduction Factor in the Exponential Family of Distributions

Based on the equations (3) and (12), the exact form of LRF_B is shown as:

$$LRF_B = 1 - \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\frac{2\log(BF_{i21})}{n_i} + \frac{\log(n_i)^{q_1-q_2}}{n_i} + \frac{K'(\hat{\beta}^{(1)},\hat{\beta}^{(2)})}{n_i}\right) \quad (14)$$

In moderate samples, one needs to compute the exact value of LRF_B relative to their priors $\pi(\beta_i^{(k)}|M_k), k = 1,2$ and i=1,2,…,n This, however, requires the computation of the remainder $K(\hat{\beta}_{1,n_i},\hat{\beta}_{2,n_i})$ in (11). In the present work, three types of priors have been considered each of which reflects the degrees of prior knowledge available to the researchers:

(i) For i=1, 2,…, n, there is independent normal priors with known variance σ^2, i.e., $\pi(\beta_i^{(k)}|\sigma^2) \sim N(0,\sigma^2 I)$

(ii) For i=1,2,…,n, there is independent reference normal prior with covariant matrix Σ_i equal to the inverse of the Fisher information matrix $I(\beta_i^{(k)})$, i.e., $\pi(\beta_i^{(k)}|I(\beta_i^{(k)})) \sim N(0, I(\beta_i^{(k)}))$.

(iii) For i=1,2,…,n, there is independent non-informative prior, i.e., $\pi(\beta_i^{(k)}) \sim 1$

With this approach, if the reference prior (ii) is chosen, it needs the Fisher information matrices $I(\beta_i^{(k)})$ for $\beta^{(1)}$ and $\beta^{(2),}$ in the contexts of models (1) and (2) in each center i. This part shows how to compute $I(\beta_i^{(k)})$ for 3 forms.

1- With respect to model (1) in special case, supposing N=1 where it is assumed to have one center (i=1) and T as a response variable such that T_{1j} are independently binomial (n_1, p_1), for (j=1, …, n_1), and then further $X_{1j}^{(1)} = (x_{01j}, x_{11j}, ..., x_{(q-1)1j})^T = (1, Z_{1j}, x_{21j}, \ldots, x_{(q-1)1j})^T$ is a corresponding vector of covariates where g(.) = log{p_1/(1 –p_1)} and $\beta_1^{(1)} = (\beta_{01}^{(1)}, ..., \beta_{(q-1)1}^{(1)})^T$. The information matrix $I(\beta_1^{(1)})$ is given by $\hat{I}(\beta_1^{(1)}) = X_{1j}^{(1)T} W X_{1j}^{(1)}$ and W is the diagonal matrix with non-zero elements $W_{JJ} = n_1\, p_1(1 - p_1)$ for (j=1, …, n_1) similarly in the model (2) (Robert and Suresh, 1992)

2-We illustrate the above procedure for the binomial distribution. As such, we assume $T_i \sim Bin(n_i, p_i), i = 1,...,N$, then $\eta_i = g_i(.) = \log\{p_i/(1-p_i)\}$. In what follows, it is assumed that the vector of $\beta^{(2)} = (\beta^{(1)}, \beta_S)^T$, and the design matrix $X^{(2)} = (X^{(1)}, S)^T$ are partitioned correspondingly and hence; we shall have dim($\beta^{(1)}$) =q_1< q_2=dim($\beta^{(2)}$).

In $I(\beta^{(k)}) = X^T W X$, W has the diagonal elements $W_{ii} = n_i p_i (1 - p_i)$. The estimate of elements of information matrix $\hat{I}(\beta^{(k)})$ k=1,2 in each cell (ab_{th}) is given as:

$$\hat{I}_{ab}(\beta^{(k)}) = \frac{\sum_{i=1}^{N} x_{ia} x_{ib} n_i \exp(\sum_j \beta_j^{(k)} x_{ij})}{\{1 + \exp[\sum_j \beta_j^{(k)} x_{ij}]\}^2}$$

3- Here, supposing T as a response variable with a distribution belonging to the exponential family or presuming that the trials, in the contexts of models (1) and (2), have been performed in N centers. Consequently, the information matrix $I(\beta^{(k)})$ for $\beta^{(k)}$ of the generalized linear models under the model k is a square matrix with dimension $\beta^{(k)}$ whose entries are defined below. It is clear that $I(\beta^{(k)}) = X^T W X, W = \text{diag}\{w_1, ..., w_N\}$ with $w_i = I_{ii}(\theta)(\frac{\partial \theta_i}{\partial L})^2$. Regarding the equation (4) and the link function $\eta(\theta_i) = L \equiv x_i \beta$, we have:

$I_{ab}(\beta_1, ..., \beta_{q+1}) = \sum_i x_{ia} x_{1b} (B''(\theta_1)) - (\frac{\eta''(\theta_1) B'(\theta_1)}{\eta'(\theta_1)})(\frac{\partial \eta^{-1}}{\partial L})^2$. In other words, when η is a canonical link function, the information matrices are same and the approximation is better.

Table 1 indicates the exact values of $K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$ for the exponential family of distributions, in general, and {within the braces} for the binomial distribution, in particular. Now, the exact value of LRF_B can be computed for moderate sample sizes by substituting $K(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$ as in equation (14). For instance, this value for reference normal prior is

$$LRF_B = 1 - \frac{1}{N}\sum_i \exp\left(-\frac{2\log(BF_{i21})}{n_i} + \frac{\log(n_i)^{q_1 - q_2}}{n_i} - \frac{(\hat{\beta}_1'(\hat{I}(\beta_1))^{-1}\hat{\beta}_1 + \hat{\beta}_2'(\hat{I}(\beta_2))^{-1}\hat{\beta}_2)}{2n_i} \right) \quad (15)$$

3. A Simulation Study

In the present study, a simulation example has been used with the help of R-software to show the behavior of LRF_B against its frequentist counterpart LRF to ascertain the advantage of the proposed criterion in multi-center. The generated data simulate 150 different multicenter trials datasets at 3 positions:

1) 50 datasets with 5 centers in each set,

2) 50 datasets with 10 centers in each set,

3) 50 datasets with 50 centers in each set,

500 unicenter sets of data are generated too.

Each center contains 500 observations, with half of them for treatment and half of them for observing the controlled groups. When S and T lack strong relations (0.5<LRF<0.6), it is assumed that the treatment has a discernible effect on both. The clinical endpoint variable follows a binomial distribution with success probability equals to 0.8 for the treatment and 0.1 for the control. The surrogate endpoint is simulated as a continuous variable for four possible situations: (T=0, Z=0), (T=1, Z=0), (T=0, Z=1), (T=1, Z=1) from normal distributions with different means and variances.

To evaluate the surrogacy value, the current research initially fit the models (1) and (2) into the generated data (T, S, Z). Here, the logistic link function is used. Next, from the equation (5), the Bayes factor BFB_{21B} is computed for 8 different values of the prior variance σ^2.

Finally, LRF_B is computed numerically from the equation (13) by N=50 and ni=500 i=1, …, 50 for each multicenter dataset. The corresponding LRF is computed from equation (3) where G^2 is the likelihood ratio statistic. Comparing models (1) and (2) in all 500 unicenter datasets, their results are computed separately with N=1, n=500. Table 2 gives the means of LRF and LRF_B for various choices of prior variance σ^2.

Table 2 and Figs. 1 & 2 make it clear that LFR is almost larger than LRF_B. The difference between the two becomes zero and results of both the Bayesian and frequentist approaches approximately equal when the variance of prior distribution goes to infinity. In fact, the LFR can be viewed as a special case of LRF_B.

4. Application

The proposed method is illustrated by a real example concerning the validity of immunologic factor "*IL-18 BPa Serum"* as a surrogate for long-term effect of dermatologic lesions caused by the mustard gas exposure during the Iraq-Iran war. The clinical endpoint T is a binary variable denoting the existence of Xerosis, while the surrogate is a continuous variable, measuring the amount of *IL-18 BPa* in the patient's blood. *IL-18 BPa* is precisely measured in the laboratory and recorded for each subject by milligram (*IL-18 BPa*/100). The explanatory variable Z is an indicator variable with Z=1 for the exposed subjects and Z=0, for otherwise.

The data are borrowed from a large scale epidemiologic study in a war zone (Ghazanfari et al. 2009). The study includes 461 people including 339 from the town of Sardasht who were injured during chemical bombardments of the area, and 122 persons from the town of Rabat as the control group. Among those exposed, 159 were hospitalized and 180 were treated as outdoor patients.

The controlled group was similar to the exposed one in terms of their age or ethnic, cultural and religious characteristics, even the amount of stress caused to them while living in the war zone. The only difference between these two population groups was that the Sardasht was subjected to chemical bombardment by Iraqi forces.

Now, we are to evaluate the validity of IL-18 BPa factor as a surrogate for the long-term effect of Xerosis. With logit link functions used in the models (1) and (2), we therefore have:

$$\text{Logit}(\pi_i^{TIZ}) = -1.514 + 1.365 Z_i \qquad \text{(P-value<0.0001)}$$

$$\text{Logit}(\pi_i^{TI,S,Z}) = -1.706 + 1.334 Z_i + 0.045 S_i \qquad \text{(P-value=0.024)}$$

Other explanatory variables did not have a discernable effect on the response, hence; are deleted from the models. It must be noted that we have only one big center, i.e., N=1, n=461, and since there is only one surrogate, $q_2 - q_1 = 1$, therefore,

$$LRF_B = 1 - \exp\left(-\frac{2\log(BF_{21})}{n} - \frac{\log(n)}{n} \right)$$

To compute the Bayes factor applying R software for the above models, normal priors, $\beta_j \sim N(0, \sigma_j^2)$ were used. The numerical results were stable after 1000 iterations. The value of

LRF from equation (3) is 1.05%, and the value of LRF_B depended on the prior variance σ^2. As Fig. 3 shows, the values ranged from 1.03% up to 4.84% corresponding to the values of $\sigma_j^2 = 90$ up to $\sigma_j^2 = 0$. Both approaches show a low validity for this surrogate hence; no significant relations could be found between the laboratory measurements of *IL-18 BPa* and Xerosis existence, after adjusting the mustard gas exposure.

5. Discussion

In this study, we have tried to explore a Bayesian approach with the aim to evaluate the validity of a surrogate at the individual level. Prior knowledge is inevitably used for all inferences, no matter Bayesian or otherwise, if only to provide a context in which "data" is more than a list of numbers or symbols. The legitimate inference always acknowledges and ponders its assumptions, thus reducing the chance of self-deceit as well as the appearance of biased advocacy. The Bayesian inference is not merely data analysis, but analysis of data and priors. This paper is based on the framework laid down by Prentice (1989) which was later modified by Buyse and Molenberghs (2000) and then extended further to meta-analysis for multi-center trials by Alonso et al. (2004). These efforts led to the likelihood reduction factor, LRF, which is employed to evaluate the validity of surrogate at the individual level. A viable alternative is the Bayesian approach which we have followed here and compute the LRF via the Bayesian approach, namely LRF_B. . We used an approximation to BIC criterion in model choice. Preference given to BIC is based on the argument that the likelihood ratio statistic G^2 should not be considered alone without its degrees of freedom. While BIC itself takes care of the degrees of freedom, it can be used in moderate samples too. Furthermore, Bayes factor makes sense that there should be a simple mechanism for describing the accumulation of evidence in favor of one model over another (Link and Barker, 2010).

Due to sensitivity of the test results to the employed prior in a Bayesian analysis, one must be careful in choosing a prior. To illustrate this point, we have simulated different trials with different priors in the logistic regression models. The results show that the LRF can be viewed as a special case of LRF_B relative to a certain prior. Hence, the importance of prior knowledge in the Bayesian analysis is shown.

The proposed procedure has been illustrated in a real example concerning the skin lesions of people who exposed to the chemical warfare, i.e. mustard gas used by Iraqi forces during the Iraq-Iran war. The results concerning the evaluation of *"IL-18 BPa Serum"* as a surrogate for long-term effects of skin lesions in injured persons have been obtained.

References

[1] Alonso A., Molenberghs G., Burzykowski T., et al. (2004) Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. Biometrics; 60: 724 –728.

[2] Berger JO., Perrichi RL. (1996) The intrinsic Bayes factor for model selection and prediction. Journal of the American Statistical Association; 91: 109-22.

[3] Buyse M., Molenberghs G., Burzykowski T., et al. (2000) The validation of surrogate endpoints in meta analyses of randomized experiments. Biostatistics; 1: 49–67.

[4] Erkanli A. (1994) Laplace approximations for posterior expectations when the mode occurs on the boundary of the parameter space. Journal of the American Statistical Association; 89:250-258.

[5] Frangakis CE., Rubin DB.(2002) Principal stratification in causal inference. Biometrics; 58:21-29.

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

[7] Ghazanfari T., Faghihzadeh S., Aragizadeh H., et al. (2009)Sardasht-Iran cohort study of chemical warfare victims, Design and Methods. Archives of Iranian Medicine;12(1):5-14.

[8] Haughton DMA. (1988) On the choice of a model to fit data from an exponential family. Ann. Statist.; 16: 342-55.

[9] Kadane J., Lazar N. (2004) Methods and criteria for model selection. Journal of the American Statistical Association; 99: 279-290.

[10] Kass RE., Raftery AE. (1995) Bayes factors. Journal of the American Statistical Association; 90: 773-95.

[11] Kass RE., Tierney L., Kadane JB. (1990) The validity of posterior expansions based on Laplace's method, Essays in Honor of George Bernard, eds. S. Geisser, J.S. Hodges, S.J. Press, and A. Zellner, Amsterdam: North Holland,; 473-488.

[12] Kass RE., Wasserman L. (1995) A reference Bayesian test for nested hypotheses with large samples. Journal of the American Statistical Association; 90: 928-34.

[13] Kent J.(1983) Information gain and a general measure of correlation. Biometrika; 70:163 –173.

[14] Link WA., Barker RJ. (2010) Bayesian inference with ecological applications. Elsevier, ISBN: 978-0-12-374854-6. First edition; p: 7-8.

[15] Li Y., Taylor J. M.G., Elliott M. R. (2010) A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. Biometrics; 66 :523-531.

[16] Nishii R. (1984) Asymptotic properties of criteria for selection of variables in multiple regression. Ann. Statist.; 12: 758–765.

[17] Prentice RL. (1989) Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in Medicine; 8 : 431– 440.

[18] Robert CP. (2001) The Bayesian Choice T, Second edition. Springer- Verlag.

[19] Robert E Kasst, Suresh K Vaidyanathan. (1992) Approximate Bayes factors and Orthogonal parameters, with application to testing equality of two Binomial Proportions. Journal of the Royal Statistical Society, Series B (Methodological), Vol. 54, No. 1, pp.129-144.

[20] Smith AFM, Spiegelhalter DJ. (1980) Bayes factors and choice criteria for linear models. J. R. Statist. Soc. B; 42 :213-20.

[21] Tiernny L, Kadane JB. (1986) Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association; 81:82-6.

[22] Weir CJ, Walley RJ. (2006) Statistical evaluation of biomarkers as surrogate endpoints: a literature review, Statistics in Medicine; 25: 183–203.

Appendix

Theorem: For large sample size, when we have $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(2)}\right| \le \delta$ (small G^2)

$$\text{LimLRF}_B = \text{LRF}$$

Proof: We have from equation (10)

$$\log \text{BF}_{12} = \text{Sch} + \text{K}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$$

In which

$$\text{K}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}) = -\frac{1}{2}(q_2 - q_1)\log(2\pi) + \frac{1}{2}\log\frac{|\Sigma_1|}{|\Sigma_2|} + \log(\frac{\pi_1(\hat{\beta}^{(1)})}{\pi_2(\hat{\beta}^{(2)})})$$

If, $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(2)}\right| \le \delta$, it means $\pi(\hat{\beta}^{(1)}) \approx \pi(\hat{\beta}^{(2)})$ that we have

$$\pi(\hat{\beta}^{(1)}) = \pi(\hat{\beta}^{(2)}) + \varepsilon_1(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$$

$$\text{Log}\frac{\pi(\hat{\beta}^{(1)})}{\pi(\hat{\beta}^{(2)})} = \text{Log}(1 + \alpha_1(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}))$$

$$\left|\alpha_1(\hat{\beta}^{(1)},\hat{\beta}^{(2)})\right| = \left|\frac{\varepsilon_1(\hat{\beta}^{(1)},\hat{\beta}^{(2)})}{\pi(\hat{\beta}^{(2)})}\right| \le 1$$

When n goes to infinite, so we have:

$$\mathrm{Log}(1+\alpha_1(\hat{\beta}^{(1)},\hat{\beta}^{(2)})) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{(\alpha_1(\hat{\beta}^{(1)},\hat{\beta}^{(2)}))^n}{n} \to 0$$

Also we have

$$\left|\Sigma_1(\hat{\beta}^{(1)})\right| = \left|\Sigma_2(\hat{\beta}^{(2)})\right| + \varepsilon_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)})$$

$$\mathrm{Log}\frac{\left|\Sigma_1(\hat{\beta}^{(1)})\right|}{\left|\Sigma_2(\hat{\beta}^{(2)})\right|} = \mathrm{Log}(1+\alpha_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)}))$$

Such that

$$\left|\alpha_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)})\right| = \left|\frac{\varepsilon_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)})}{\pi(\hat{\beta}^{(2)})}\right| \le 1$$

So

$$\mathrm{Log}(1+\alpha_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)})) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{(\alpha_2(\hat{\beta}^{(1)},\hat{\beta}^{(2)}))^n}{n} \to 0$$

Finally, because of $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(2)}\right| \le \delta$

$$\frac{1}{2}(q_2 - q_1)\mathrm{Log}2\pi \to 0$$

And proof is complete.

Table 1.The exact amounts of $K'(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$ based on different priors for exponential family true endpoints. (The results for the special case of *Binomial distribution* are given within braces)

Prior Distribution $\pi_i(\beta_i \mid \sigma^2)$	Remainder term $K'(\hat{\beta}^{(1)}, \hat{\beta}^{(2)})$ when true endpoints are from exponential family distribution {*Binomial distribution*}
$N(0, \sigma^2 I)$	$-\frac{1}{2}[(m_2 - m_1)\log(\sigma^2) + \frac{(\hat{\beta}^{*\prime}\hat{\beta}^{*} - 2\hat{\beta}_{11}'\hat{\beta}_2)}{\sigma^2} - \log\frac{\prod_{j'}(X_{j'}^{*})'W^{*}}{\prod_{j}(X_{j}^{*})'W^{*}}]$ $\{-\frac{1}{2}[(m_2 - m_1)\log(\sigma^2) + \frac{(\hat{\beta}^{*\prime}\hat{\beta}^{*} - 2\hat{\beta}_{11}'\hat{\beta}_2)}{\sigma^2} - \log\frac{\prod_{j'}(X_{j'}^{*})'P^{*}}{\prod_{j}(X_{j}^{*})'P^{*}}]\}$
$N(0, I(\beta_i))$	$-\frac{1}{2}(\hat{\beta}_1'(I(\beta_1))^{-1}\hat{\beta}_1 + \hat{\beta}_2'(I(\beta_2))^{-1}\hat{\beta}_2)$ $\{-\frac{1}{2}(\hat{\beta}_1'(I^{logit}(\beta_1))^{-1}\hat{\beta}_1 + \hat{\beta}_2'(I^{logit}(\beta_2))^{-1}\hat{\beta}_2)\}$
1	$-\frac{1}{2}[(m_1 - m_2)\log(2\pi) - \log\frac{\prod_{j'}(X_{j'}^{*})'W^{*}}{\prod_{j}(X_{j}^{*})'W^{*}}]$ $\{-\frac{1}{2}[(m_1 - m_2)\log(2\pi) - \log\frac{\prod_{j'}(X_{j'}^{*})'P^{*}}{\prod_{j}(X_{j}^{*})'P^{*}}]\}$

$$W^{*} = ((B''(\hat{\theta}_1) - \frac{\eta''(\hat{\theta}_1)B'(\hat{\theta}_1)}{\eta'(\hat{\theta}_1)})(\frac{\partial\eta^{-1}(\hat{\theta}_1)}{\partial L})^2 \quad \dots \quad (B''(\hat{\theta}_i) - \frac{\eta''(\hat{\theta}_i)B'(\hat{\theta}_i)}{\eta'(\hat{\theta}_i)})(\frac{\partial\eta^{-1}(\hat{\theta}_i)}{\partial L})^2)_{1.i}^{T}$$

$$P^{*} = W^{logit*} = (\frac{n_1\, exp(x_1\beta)}{1 + exp(x_1\beta)} \quad \dots \quad \frac{n_i\, exp(x_i\beta)}{1 + exp(x_i\beta)})_{1.i}^{T}$$

$$X_j^{*} = (x_{1j}^{2} \quad \dots \quad x_{ij}^{2})_{1.i}^{T}$$

Table 2.Mean and standard error of LRF and LRF_B based on 8 different values of prior variance σ^2 in each dataset for 150 multicenter and 500 unicenter simulated dataset

Number of Simulated dataset		500	50	50	50
Number of Center in each dataset(N)		1	50	10	5
Number of Subjects in each center		n=500	ni=500 i=1,…,50	ni=500 i=1,…,10	ni=500 i=1,…,5
LRF		0.547774	0.548610	0.547933	0.548155
LRF_B	Var = 1000	0.545626	0.546479	0.545830	0.546028
	Var = 100	0.543605	0.544438	0.543709	0.543980
	Var = 10	0.519660	0.520482	0.519572	0.519942
	Var = 5	0.501217	0.502049	0.500774	0.501507
	Var = 1	0.422984	0.423868	0.422184	0.423415
	Var = 0.1	0.223600	0.224389	0.222757	0.223791
	Var = 0.01	0.096893	0.097471	0.096758	0.096751
	Var =0.001	0.083666	0.084310	0.083837	0.083364

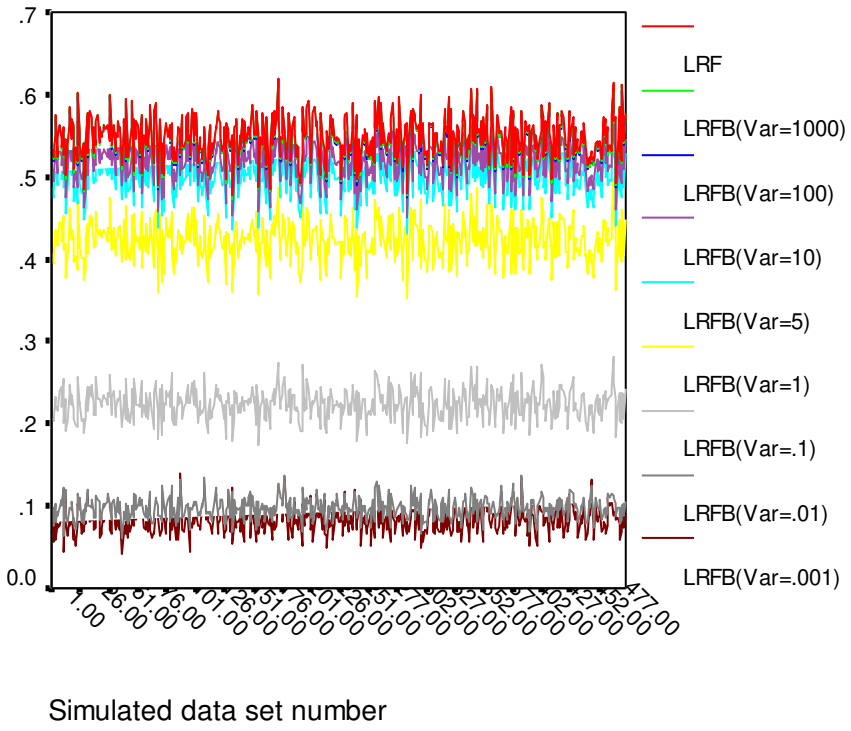

Figure 1: Plots of individual level criterion of surrogacy for LRF and LRF_B relative to various values of the prior variance in 500 different unicenter datasets.

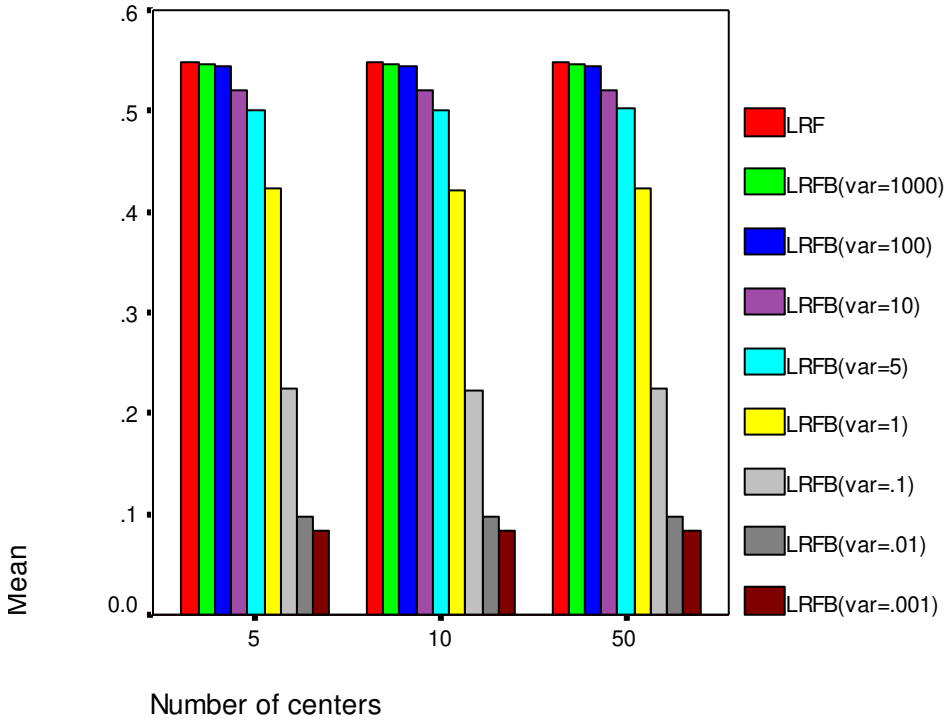

Figure 2: Plots of mean of individual level criterion of surrogacy for LRF and LRF_B relative to various values of the prior variance in 150 multicenter different datasets

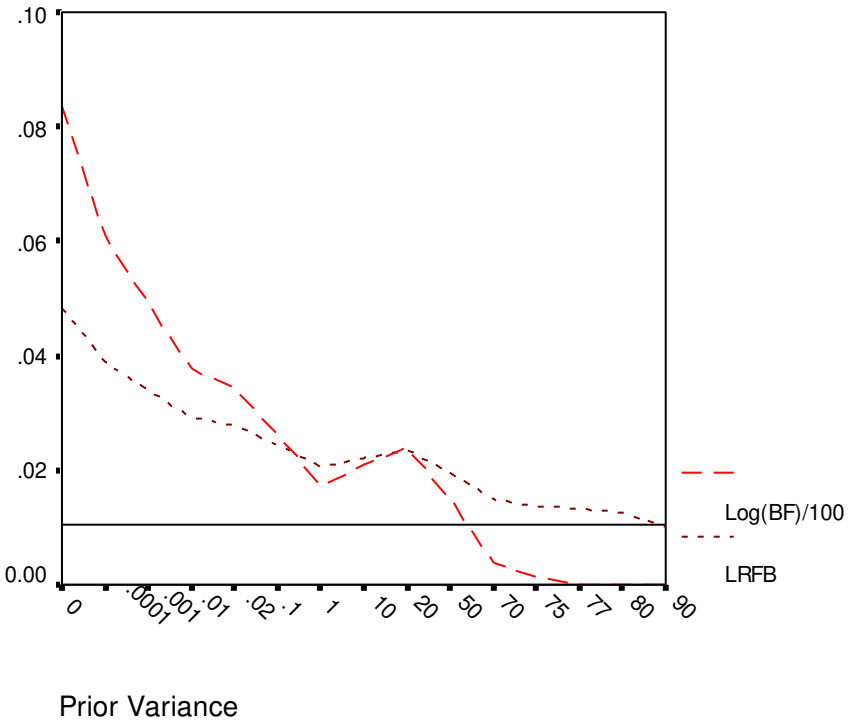

Figure 3: Plot of LRF_B against the different values of the prior variance used for evaluation of IL-18 BPa as a surrogate for Xerosis due to long term effects of mustard gas. (Reference line indicates LRF)