A Hierarchical Bayesian Approach for the Evaluation of Surrogate Endpoints in Multiple Randomized Clinical Trials

Ziv Shkedy, Franz Torres, Tomasz Burzykowski and Geert Molenberghs

Limburgs Universitair Centrum, Center for Statistics,Biostatistics, Universitaire Campus, Building D,

B 3590 Diepenbeek, Belgium

Submitted for oral presentation at IWSM18. Presenter: Ziv Shkedy.

Keywords: Hierarchical Bayesian models; Individual level surrogacy ; Surrogate endpoints; Trial level surrogacy;

In Randomized clinical trials, the main interest is to assess the treatment (Z) effect on the primary endpoint (T). However, in some cases, the time needed to observe the endpoint of interest could be long (for example, if the primary endpoint is time to event). In other cases the observation of the primary endpoint could be very expensive. In these cases, one may benefit in using a surrogate endpoint(S) to determine the treatment effect quicker (or less expensive).
In his landmark paper, Prentice (1989) proposed a definition for true and surrogate endpoints and suggested criteria for validation. According to the definition of Prentice (1989) a surrogate endpoint is a variable for which a test of the null hypothesis of no relationship to the treatment group under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint. Freedman, Graubard, and Schatzkin (1992) focused on a single trial and single surrogate setting and suggested to use the proportion of treatment effect explained by the surrogate endpoint in order to evaluate a potential surrogate. However, De Gruttola *et al.* (1997) and Buyse and Molenberghs (1998) have shown that the proportion of treatment effect explained by the surrogate does not restrict to the $[0, 1]$ interval. Buyse and Molenberghs (1998) proposed two measures for evaluation, the relative effect and the adjusted association. The first can be seen as a measure for surrogacy at the trial level while the second measure for surrogacy at an individual level.
In this paper we focus on the case that a potential surrogate is evaluated using data from multiple trials in a meta-analytic fashion. We further assume that for two endpoints from the exponential family the true treatment effects are given by

$$\begin{aligned} g\{E(S|Z=1)\} - g\{E(S|Z=0)\} &= \alpha \\ g\{E(T|Z=1)\} - g\{E(T|Z=0)\} &= \beta \end{aligned} \tag{1}$$

Within the meta-analytic approach the first goal is to establish the association between β and α to asses the quality of trial level surrogacy. This can be done by formulating a model for the joint distribution of treatment effects $[\alpha, \beta]$, or from a model of the conditional distribution $[\beta|\alpha]$. Note that a joint model $[\alpha, \beta]$ imposes a conditional model for $[\beta|\alpha]$ but one can specifiy a model for $[\beta|\alpha]$ without specifying the joint model. The second goal is to assess the quality of individual level surrogacy which can be evaluated from the joint distribution of S and T, $[T, S, |Z]$.
The evaluation of surrogate endpoint within the meta-analytic setting is discussed by Daniels and Hughes (1997) and Buyse *et al.* (2000). Both authors considered a multiple trial setting with a single surrogate endpoint and proposed a two-stage model for the evaluation of a potential surrogate endpoint. Daniels and Hughes (1997)

proposed a hierarchical Bayesian model for the estimated treatment effects in which the joint distribution of the estimated treatment $[\hat{\alpha}, \hat{\beta}]$, is specified at the first stage and the conditional distribution of $[\beta|\alpha]$ is specified in the second stage. Buyse *et al.* (2000) formulated a two-stage model in which the joint distribution $[T, S|Z]$ is specified in the first stage and the joint distribution of the treatment effects $[\beta, \alpha]$, is specified at the second stage. The advantage to use Daniels and Hughes (1997) model is that one does not need to specify the joint distribution of T and S. However, it came with the price that quality of individual level surrogacy cannot be assessed. In this paper we consider the case that individula data are avilable and T, S are notmal distributed.

We consider the following linear predictors for T and S

$$\begin{cases} g\{E(S_{ij}|Z_{ij})\} = \mu_{Si} + \alpha_i Z_{ij}, \\ g\{E(T_{ij}|Z_{ij})\} = \mu_{Ti} + \beta_i Z_{ij}. \end{cases} \tag{2}$$

Here α_i and β_i are trial-specific treatment effects, μ_{S_i} and μ_{T_i} are trial-specific intercepts and S_{ij} and T_{ij} are the surrogate and the true endpoints respectively of subjects j, $j = 1, 2, \ldots, n_i$, in trial i, $i = i, 2, \ldots, N$. We further assume that the two endpoints are normal distributed. Thus, at the first stage of the hierarchical model we specify the joint distribution of T and S,

$$\begin{pmatrix} S_{ij} \\ T_{ij} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_S + m_{Si} + (\alpha + a_i) Z_{ij} \\ \mu_T + m_{Ti} + (\beta + b_i) Z_{ij} \end{pmatrix}, \Sigma \right), \tag{3}$$

where Σ is given by

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

At the second stage of the model the prior models for the 'fixed' effects are specified

$$\begin{aligned} &\mu_S \sim \mathrm{N}(0, \theta^2_{\mu_S}), \\ &\mu_T \sim \mathrm{N}(0, \theta^2_{\mu_T}), \\ &\alpha \sim \mathrm{N}(0, \tau^2_{\alpha}), \\ &\beta \sim \mathrm{N}(0, \tau^2_{\alpha}). \end{aligned} \tag{5}$$

Gamma distributions were specified as (flat) hyperprior models for the precision parameters in (5), i.e., $\theta^{-1}_{\mu_S} \sim$ gamma$(0.0001, 0.0001)$ etc. Similar to the model proposed by Daniels and Hughes (1997) we need to specify a prior distribution to model the association between the treatment effects of the two endpoints. However, while Daniels and Hughes (1997) based their model on $[\boldsymbol{\beta}|\boldsymbol{\alpha}]$, Buyse *et al.* (2000) use the joint distribution of the random effects in order to evaluate trial level surrogacy. In the current model we specify the prior model for joint distribution of the random effects $(\mathbf{m}_S, \mathbf{m}_T, \mathbf{a}, \mathbf{b})$ to be

$$\begin{pmatrix} m_{Si} \\ m_{Ti} \\ a_i \\ b_i \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, D \right), \quad i = 1, \ldots, N. \tag{6}$$

The covariance matrices are assumed to follow a Wishart distribution,

$$\begin{aligned} &D_{4\times 4} \sim \text{Wishart}(R_D), \\ &\Sigma_{2\times 2} \sim \text{Wishart}(R_\Sigma). \end{aligned} \tag{7}$$

In order to assess surrogacy at the trial level, Buyse *et al.* (2000) proposed to use the coefficient of determination

$$R^2_{\text{trial}} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}}. \tag{8}$$

As a measure for individual level surrogacy Buyse *et al.* (2000) proposed to use R^2_{ind} given by

$$R^2_{\text{indiv}} = \frac{\sigma^2_{ST}}{\sigma_{SS}\sigma_{TT}}. \tag{9}$$

Indeed, $R^2_{\text{trial}} = 1$ and $R^2_{\text{indiv}} = 1$ indicate on perfect surrogacy in trial and individual level respectively.

In order to avoid computations problems Buyse *et al.* (2000) proposed a reduced model in which the linear predictors of S and T do not include trial specific intercepts. In the hierarchical model, the likelihood at the first stage of the model can be specified by omitting the trial specific random intercepts from (3).

$$\begin{aligned} &(T_{ij}, S_{ij})^T \sim \text{N}\left((\eta_{T_{ij}}, \eta_{S_{ij}})^T, \Sigma\right), \\ &\eta_{T_i j} = \mu_T + (\alpha + a_i) Z_{ij}, \\ &\eta_{S_i j} = \mu_S + (\alpha + a_i) Z_{ij}. \end{aligned} \tag{10}$$

At the second stage of the model, the prior model for the random effects is given by

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D\right). \tag{11}$$

Note that the covariance matrix in (11) is a 2×2 matrix given by

$$D = \begin{pmatrix} d_{aa} & d_{ab} \\ d_{ab} & d_{bb} \end{pmatrix}, \tag{12}$$

which is assumed to follow a Wishart prior model, $D_{2\times 2} \sim \text{Wishart}(R_D)$. Other prior and hyperprior models remain the same as in the full model. For the reduced model trial level surrogacy is measured with the coefficient of determination which reduces to

$$R^2_{\text{trial}} = \frac{d^2_{ab}}{d_{aa} d_{bb}}. \tag{13}$$

For illustration we consider data from two randomized multicenter trials in colorectal cancer. All data were collected and checked by the Meta-Analysis Group in Cancer between 1990 and 1996 (Corfu-A Group, 1995; Greco *et al.* 1996) in order to confirm the benefits of experimental fluoropyrimidine treatments with 5-fluorouracil (5FU) in advanced colorectal cancer. The treatment comparison made were 5FU plus interferon versus 5FU plus folinic acid in one trial (Greco *et al.* 1995) and 5FU plus interferon versus 5FU alone in the other trial (Greco *et al.* 1996). The final endpoint T_{ij} is the log(survival time), in years, and the potential surrogate endpoint S_{ij} is the log(progression-free survival time), that is the years between the randomization to clinical progression of the disease or death. The data were analyzed previously by Byuse *et al.* (2000).
We fitted the models using MCMC simulation with 9000 iteration followed a burn-in period of 1000 iterations. Table 1 presents the maximum likelihood estimates for both R^2_{trial} and R^2_{indiv} (obtained from the fixed effects model and reported in Buyse *et al.* (2000)) and the posterior means (obtained from the full Bayesian model). The results are comparable, although the standard errors of R^2_{trial} are greater for the Hierarchical Bayesian model. This result in credible intervals for the posterior means which are wider than the confidence intervals for the ML estimates (see Figure 1, panel *c*). Kernel estimates for the posterior density of R^2_{trial} and R^2_{indiv} are shown in panels *a* and *b*, respectively.

Table 1: R^2_{trial} and R^2_{indiv}. The full fixed effects model coresponds to the model in Eq. () and the rediced fixed effects model coresponds to the model in Eq. (). Both model presented in Byuse et al (2000).

MODEL	Trial level R^2_{trial}	Individual level R^2_{indiv}
Reduced (Fixed)	0.928 (0.020)	0.888 (0.0006)
Full (Fixed)	0.94 (0.017)	0.886 (0.0006)
Reduced (Bayesian)	0.925 (0.048)	0.885 (0.0006)
Full(Bayesian)	0.938 (0.038)	0.885 (0.0006)

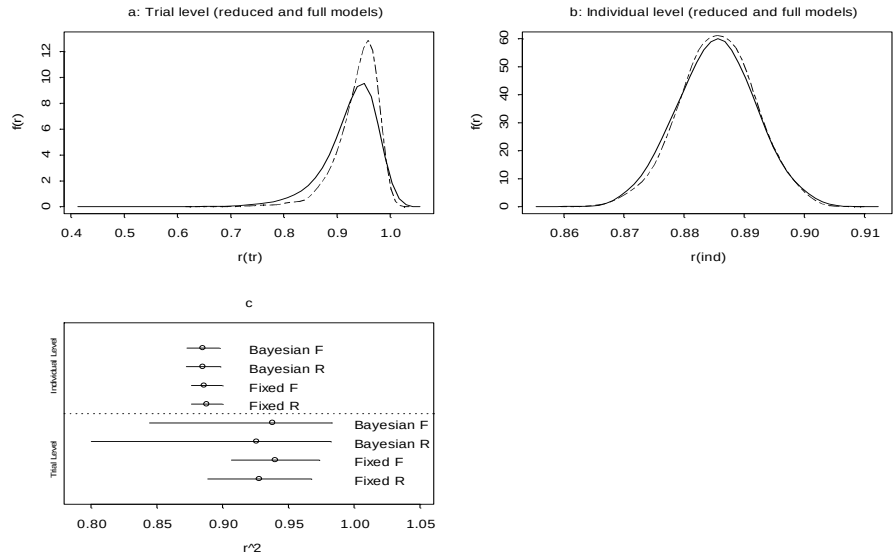

Figure 1: *Ovarian cancer data. Panel a: posterior distribution of R^2_{tr} (reduced model-soild line, full model - long dashed line). Panel b: posterior distribution of R^2_{idv}. Panel c:* 95% *crediable intervals (for the bayesian models) and* 95% *confidence intervals (for the fixed effects model) for R^2_{idv} and R^2_{tr}.*

References

Buyse, M. and Molenberghs, G. (1998). The validation of surrogate endpoints in randomized experiments. *Biometrics* **54**, 186–201.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1**, 000–000.

Daniels, M.J. and Hughes, M.D.(1997), Meta-analysis for the evaluation of potential surrogate markers, *Statistics in Medicine* **16**, 1965-1982.

De Gruttola, V., Fleming, T.R., Lin, D.Y. and Coombs, R. (1997). Validating surrogate markers - are we being naive ? *Journal of Infectious Diseases* **175**, 237–246.

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

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