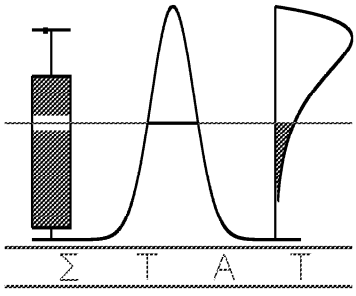
T E C H N I C A L
R E P O R T

0226

Surrogacy Evaluation of Immunological Parameters in Pilot Cancer Clinical Trials

F. Torres Barbosa, F. Tibaldi, J. Cortiñas Abrahantes, H. Geys, G. González, G. Molenberghs



I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Surrogacy Evaluation of Immunological Parameters in Pilot Cancer Clinical Trials

Franz Torres Barbosa[1], Fabián S. Tibaldi[2], José Cortiñas Abrahantes[2], Helena Geys[2], Gisela González[1], Geert Molenberghs[2]

[1]Center of Molecular Immunology, Havana, Cuba

[2]Center for Statistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

Summary

With the new wave of cancer vaccines in clinical trials new endpoints need to be evaluated to show their efficacy. Immunological parameters become very important to show the activity of the vaccine and its relationship with the actual clinical benefit. An important tool to this end is the theory of surrogate marker evaluation. Prentice (1989) and Freedman *et al.* (1992) laid the foundations for the evaluation of surrogate endpoints in randomized clinical studies. Buyse and Molenberghs (1998) supplemented the original proposals with two new measures of validation and Buyse *et al.* (2000a) have formulated a meta-analytic approach, based upon combining the surrogate and true endpoints in a multi-trial setting. Tibaldi *et al.* (2001) studied several simplified modeling approaches in case both the surrogate as well as the true endpoint are continuous. In this paper, we study surrogacy, on survival, of the sera titer achieved by patients immunized with a cancer vaccine using three surrogate candidates and based on two different approaches. The impact of a number of simplifying modeling assumptions as well as the impact of the number of trials, centers and patients used for the validation is studied.

Some Keywords: Immunological Surrogates; Meta-analytic Approach; Plackett-Dale Model; Simplified Models; Surrogate Endpoint Validation.

1 Introduction

Surrogate endpoints are used to substitute the true endpoints (Ellenberg and Hamilton 1989), in the evaluation of response in a clinical trial. The benefit of using a surrogate lies, most often, in achieving either a faster response or a more feasible validation. But a formal validation for shifting to a surrogate endpoint should be undertaken regarding the relationship of the surrogate and true endpoints.

In cancer clinical trials, survival time is considered the more objective endpoint to assess response to treatment, but it requires a long-term trial which then can be influenced by different factors. In

the cancer setting, tumor response (e.g., given by the percentage of tumor shrinkage) has been used as a surrogate endpoint for cytotoxic drugs although it has been questioned at several occasions (Anderson, Cain and Gelber 1983; Ellenberg and Hamilton 1989; Buyse and Piedbois 1996). With the development of biotechnology, new drugs emerge in which no longer a strong cytotoxic effect is suspected but rather a cytostatic one. In this context, other surrogate endpoints like tumor markers and immunological parameters need to be validated. In this study, we address the validity of the immune response to a cancer vaccine as a surrogate for survival time in treated patients with advanced non small cell lung cancer.

The validation of surrogate endpoints in clinical trials has first been formalized by Prentice (1989), who proposed a definition as well as a set of operational criteria. Freedman, Graubard, and Schatzkin (1992) supplemented these criteria with the so-called *proportion explained* (PE). Buyse and Molenberghs (1998) proposed to use the *relative effect* (RE), linking the effect of treatment on both endpoints and an individual-level measure of agreement between both endpoints, after adjusting for the effect of treatment (*adjusted association*), instead of the PE.

Buyse *et al.* (2000a) suggested a new method for the validation of surrogate endpoints in a hierarchical data setting, such as a multicentric trial or a meta-analysis. This type of setting allows for the study of surrogacy at different levels (e.g., trial-level surrogacy and individual-patient level surrogacy). A meta-analytic approach naturally leads to the prediction of the effect of treatment upon the true endpoint, given its observed effect upon the surrogate endpoint. The linear mixed models (Verbeke and Molenberghs 2000) has been used in the case of continuous outcomes, when it is acceptable to assume a normal distribution.

Proper survival-type models were developed in Burzykowski *et al.* (2001) and Burzykowski, Molenberghs, and Buyse (2001). Molenberghs, Geys, and Buyse (2001) extended the theory to handle mixed binary-continuous endpoints with two modeling strategies based on a probit-linear model and the Plackett-Dale model. Longitudinal endpoints are studied by Renard *et al.* (2001).

Even though linear mixed models are well understood and extensively used, Buyse *et al.* (2000a) showed that achieving convergence in this setting is not a trivial task. Tibaldi *et al.* (2001) studied different strategies to deal with the computational problems in validation of surrogate markers,

based on hierarchical models. The trial, endpoint and measurement error dimensions were the three different levels of simplification proposed in their work.

Throughout this paper, we will focus on the behavior in a real application of the simplifying procedures proposed by Tibaldi *et al.* (2001) in terms of *trial-level surrogacy*. To study the numerical issues in such a hierarchical modeling strategy, we analyzed two pilot clinical trials in patients with advanced non small cell lung cancer treated with a cancer vaccine (Section 2). We address the cases of a continuous-scale and binary surrogate for a continuous variable as true endpoint. While it would be appropriate, in principle, to use models for time-to-event endpoints, such a choice is not made here for a number of reasons. First, given the small sample sizes, reaching convergence of continuous-type models is difficult, but reaching convergence survival-type models has not been possible. Second, no fully hierarchical model has been developed for the survival case. Third, since for continuous endpoints a number of simplifying approaches has been proposed, it is possible to study the impact thereof. Fourth, given the advanced state of the disease, no censoring is observed. In Section 3, the statistical models used for the surrogate assessment of the normal-normal and binary-normal cases are reviewed. The different candidate surrogates are evaluated in Section 4 using the data generated in two pilot clinical trials.

2 Non Small Cell Lung Cancer

Two pilot clinical trials were performed, with the aim of testing safety, immunogenicity, and survival of a therapeutic vaccine based on the EGF molecule in patients with advanced Non Small Cell Lung Cancer patients (González *et al.* 1998, 2001).

A first pilot study tested the vaccine in 20 patients, with NSCLC, randomized to the EGF vaccine with two different adjuvants Alum and ISA-51 Montanide. The vaccine is administered in a 5 doses scheme for 51 days. Immunogenicity data is collected weekly during treatment period and monthly during follow-up. The second pilot trial studied the same vaccines of the previous study in 20 more patients, but with a common 3 days pre-treatment with cyclosfosfamide.

In both trials, the scope of patients is reduced to very advanced cancer patients at stages IIIb or IV

Table 1: CT1: Median Survival according to immune response.

Groups	n	**Events**	**Median (months)**
High Responders	12	11	10.43
Low Responders	8	8	4.00
Historical Controls	29	29	5.67

Table 2: Immunological candidate surrogates for survival time.

Approach 1	**Approach 2**	**True Endpoint**
lg Max Titer	lg End Titer	ln(Survival)
Greater 1:2000	Greater 1:2000	ln(Survival)
Greater 1:2000 and 8×	Greater 1:2000 and 8×	ln(Survival)

without any other alternative of oncospecific treatment, with a WHO status less than 3. Survival time was considered from the day of random treatment assignment to the day of death regardless of the cause of death. There were three hospitals including patients: 2 in the first study and 3 in the second.

In a previous analysis, a relationship of the immunological response and the survival time was detected (Torres *et al.* 2001). For one of the trials, there was a clear advantage on survival for the high immunological responders (see Table 1).

For the surrogacy, two approaches were considered, each with 3 alternatives, giving a range of 6 possibles candidate surrogates for survival time. In the first approach, we considered the log of the maximum titer achieved during the immune evaluation for each patient. Considering then as candidates the continuous outcome given by its actual value, a binary outcome being 1 if the titer is greater than 1:2000 (3.3 on the log scale) and a second binary outcome being 1 if the titer is greater than 1:2000 and 8× the original titer (see Table 2).

In the second approach, we considered the log titer achieved by the patient at the end of the treatment period (at 60 days) and the same three alternatives as in the first approach were used.

3 Statistical Models

As indicated earlier, we opt for continuous and/or binary type response models. This implies the (log of) survival time is treated as continuous. This is justified in part by the absence of censoring on the one hand, and the flexibility of the continuous endpoint models, together with the existence of a number of simplifying strategies.

3.1 Full Random Effect Model

We will consider normally distributed endpoints. Let T_{ij} and S_{ij} be random variables denoting the true and the surrogate endpoints for subject $j = 1, \ldots n_i$ in trial $i = 1, \ldots N$. Further, let Z_{ij} denote a binary treatment indicator.

The full random-effects model, as introduced by Buyse *et al.* (2000a) is

$$S_{ij} = \mu_S + m_{S_i} + \alpha Z_{ij} + a_i Z_{ij} + \varepsilon_{S_{ij}}, \tag{1}$$

$$T_{ij} = \mu_T + m_{T_i} + \beta Z_{ij} + b_i Z_{ij} + \varepsilon_{T_{ij}}, \tag{2}$$

where μ_S and μ_T are fixed intercepts, m_{S_i} and m_{T_i} are random intercepts for trial i, α and β are fixed treatment effects and a_i and b_i are random treatment effects. The individual-specific error terms are $\varepsilon_{S_{ij}}$ and $\varepsilon_{T_{ij}}$.

The vector of random effects, $(m_{S_i}, m_{T_i}, a_i, b_i)'$, is assumed to be zero-mean normally distributed with a given covariance matrix D. The individual-level error terms $(\varepsilon_{S_{ij}}, \varepsilon_{T_{ij}})'$ are also zero-mean normally distributed with covariance matrix Σ.

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

$$R^2_{\text{trial (f)}} = R^2_{b_i|m_{Si},a_i} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}}. \tag{3}$$

A good surrogate, *at the trial level*, would have (3) close to 1. Intuition can be gained by considering the simplified case where the prediction of b_0 is done independently of the random intercept m_{S0}.

The coefficient (3) then reduces to

$$R^2_{\text{trial (r)}} = R^2_{b_i|a_i} = \frac{d^2_{ab}}{d_{aa}d_{bb}}. \tag{4}$$

Parameter estimation can be based on, for example, maximum likelihood or restricted maximum likelihood (Verbeke and Molenberghs 2000).

3.2 Two Stage Model

As Buyse *et al.* (2000a) showed the fitting of the full random effects model can be a difficult task. Tibaldi *et al.* (2001) considered several alternatives for trial simplification according to the trial dimension, the measurement error dimension and the endpoint dimension. In particular a two stage modeling approach is considered, as a useful alternative whenever convergence of the full hierarchical model is beyond reach.

In this case, at the first stage the trial-level parameters are treated as fixed, exactly as Buyse *et al.* (2000a), the model can be rewritten as

$$\begin{aligned} S_{ij} &= \mu_{S_i} + \alpha_i Z_{ij} + \varepsilon_{S_{ij}}, \\ T_{ij} &= \mu_{T_i} + \beta_i Z_{ij} + \varepsilon_{T_{ij}}, \end{aligned}$$

where μ_{S_i}, μ_{T_i}, α_i, and β_i are trial-specific intercepts and treatment effects.

At the second stage, a regression model is fitted to the treatment effects, estimated at the first stage:

$$\widehat{\beta}_i = \lambda_0 + \lambda_1 \widehat{\mu}_{S_i} + \lambda_2 \widehat{\alpha}_i + \varepsilon_i. \tag{5}$$

This model can then be used to assess trial-level surrogacy, using the R^2 associated with this regression.

Tibaldi *et al.* (2001) considered the unweighted, weighted and the measurement-error corrected alternatives to account for the measurement error or heterogeneity in information content between trial-specific contributions.

To this end, we introduce models for the estimated trial-specific treatment effects $(\widehat{\mu}_{S_i}, \widehat{\alpha}_i, \widehat{\beta}_i)'$,

given the true trial-specific treatment effects $(\mu_{S_i}, \alpha_i, \beta_i)'$:

$$\begin{pmatrix} \widehat{\mu}_{S_i} \\ \widehat{\alpha}_i \\ \widehat{\beta}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_i} \\ \alpha_i \\ \beta_i \end{pmatrix}, C_i\right). \tag{6}$$

Here, C_i is the variance-covariance matrix of the estimated treatment effects. In case we assume both treatment-effect estimates to be independent (which would result from a univariate choice on the *endpoint dimension*), C_i would assumed to be diagonal, even though this may be unrealistic.

Further, we assume a normal model for the true trial-specific treatment effects around the true overall treatment effects:

$$\begin{pmatrix} \mu_{S_i} \\ \alpha_i \\ \beta_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_i} \\ \alpha \\ \beta \end{pmatrix}, \Sigma\right). \tag{7}$$

The resulting marginal model, combining (6) and (7), is:

$$\begin{pmatrix} \widehat{\mu}_{S_i} \\ \widehat{\alpha}_i \\ \widehat{\beta}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_i} \\ \alpha \\ \beta \end{pmatrix}, \Sigma + C_i\right). \tag{8}$$

Maximum likelihood estimation for this model can be quite easily carried out by using mixed model software, provided the values for C_i can be input and held fixed, as is the case, for example, in the SAS procedure MIXED.

3.3 A Plackett-Dale Formulation

In the case of an ordinal surrogate, the following model was proposed by Molenberghs, Geys, and Buyse (2001). Assume that the cumulative distributions of S_i and T_i are given by F_{S_i} and F_{T_i}. The joint cumulative distribution of both these quantities has been studied by Plackett (1965),

$$F_{T_i,S_i} = \begin{cases} \dfrac{1 + (F_{T_i} + F_{S_i})(\psi_i - 1) - C(F_{T_i}, F_{S_i}, \psi_i)}{2(\psi_i - 1)} & \text{if } \psi_i \neq 1, \\ F_{T_i} F_{S_i} & \text{if } \psi_i = 1. \end{cases}$$

Based upon this distribution function, we can derive a bivariate Plackett "density" function $G_i(t, s)$ for mixed continuous-binary outcomes. Suppose the success probability for S_i is denoted by π_i, then we can define $G_i(\iota, s)$ by specifying $G_i(\iota, 0)$ and $G_i(\iota, 1)$ such that they sum to $f_{T_i}(\iota)$. If we

define $G_i(t,0) = \partial F_{T_i,S_i}(t,0)/\partial t$, this leads to specifying G_i by:

$$G_i(t,0) = \begin{cases} \frac{f_{T_i}(t)}{2}\left(1 - \frac{1+F_{T_i}(t)(\psi_i-1)-F_{S_i}(s)(\psi_i+1)}{C\left(F_{T_i},1-\pi_i,\psi_i\right)}\right) & \text{if } \psi_i \neq 1, \\ f_{T_i}(t)(1-\pi_i) & \text{if } \psi_i = 1. \end{cases}$$

and

$$G_i(t,1) = f_{T_i}(t) - G_i(t,0).$$

In this formulation, we assume $T_i \sim N(\mu_i, \sigma^2)$, with $\mu_i = \mu_T + \beta Z_i$ and $\text{logit}(\pi_i) = \mu_S + \alpha Z_i$ with similar notation as in the probit case. The global odds ratio is assumed to be constant. If we denote

$$\boldsymbol{\theta}_i = \begin{pmatrix} \mu_i \\ \sigma^2 \\ \pi_i \\ \psi \end{pmatrix} \text{ and } \boldsymbol{\eta}_i = \begin{pmatrix} \mu_i \\ \ln(\sigma^2) \\ \text{logit}(\pi_i) \\ \ln(\psi) \end{pmatrix},$$

estimates of the regression parameters $\boldsymbol{\nu} = (\boldsymbol{\mu}, \beta, \alpha, \ln\sigma^2, \ln\psi)$ are easily obtained by solving the estimating equations $\boldsymbol{U}(\boldsymbol{\nu}) = 0$, using a Newton-Raphson iteration scheme, where $\boldsymbol{U}(\boldsymbol{\nu})$ is given by:

$$\sum_{i=1}^{n} \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\nu}}\right)' \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\theta}_i}\right)^{-1'} \left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \ln G_i(t_i, s_i)\right).$$

Note that the adjusted association γ_Z is given by ψ in this case and the relative effect $RE = \beta/\alpha$ can be readily determined.

4 Validation of Candidate Surrogates

To evaluate the surrogacy of the continuous variables *maximal titer* and *end titer*, the full and reduced models proposed by Buyse and Molenberghs (1998) and the simplifications by Tibaldi *et al.* (2001) were used. For these models the hospital was considered as the unit of variability, with 5 units in total. For the reduced case, no random intercepts were included. As it can be seen in Table 3, the R^2_{center} were calculated for each of the alternatives.

Table 3 shows the results of the surrogacy evaluation of the actual *maximal titer* and *end titer*. With the full model the R^2_{center} was higher with the weighted approach with $R^2_{\text{center}} = 0.334$ and 0.440 for the *maximal titer* and the *end titer*, respectively. With the reduced model, R^2_{center} were

Table 3: Results of the trial-level surrogacy analysis for the *maximal titer* and *end titer* with full and reduced models. R^2_{center} are reported.

	Full Model			
	Max Titer		End Titer	
	Univariate	Bivariate	Univariate	Bivariate
Unweighted	0.117	0.117	0.370	0.370
Weighted	0.334	0.334	0.440	0.440
Error Corrected	-	-	-	-
	Reduced Model			
	Max Titer		End Titer	
	Univariate	Bivariate	Univariate	Bivariate
Unweighted	0.111	0.111	0.069	0.069
Weighted	0.234	0.234	0.005	0.005
Error Corrected	0.146	0.146	0.02	0.02

higher for the weighted approach in the case of the *maximal titer* surrogate, for the *end titer* all the R^2_{center} were very low. In general, none of the obtained R^2_{center} are high enough to suggest these parameters as a good surrogate. In the full model there was no convergency of the models with the measurement error correction. Also, there was a clear loss in the R^2_{center} estimates for the reduced model compared to the full model due to lack of correction for random intercepts.

We certainly conclude that, while many of the R^2 values are small, some promise of the surrogate *end titer* is seen in case a weighted version of the full model is used. Evidently, more elaborate sample sizes are needed to confirm or contradict this initial finding.

To evaluate the surrogacy of the binary variables based on the *maximal titer* and *end titer*, the probit and Plackett-Dale models were used. The data from both trials when combined for the surrogacy evaluation. Also in the case of binary data, a very reasonable choice is the use of a meta-analytic framework (Renard *et al.*, 2001). However, in our case, this approach was not feasible because, both, the sample size per center as well as the number of centers were too small. Combined with the limited amount of information coming from binary data, this prohibited the use of the meta-analytic paradigm. Therefore, we believe the concepts from Buyse and Molenberghs (1998) still are useful summaries in this case.

Table 4: Parameter estimates (standard errors) for the probit model.

Parameter	**Binary$_{12}$**	**Binary$_{13}$**	**Binary$_{22}$**	**Binary$_{23}$**
μ_T	-1.969(0.358)	-1.969(0.319)	-2.062(0.339)	-2.062(0.328)
μ_S	0.246(0.843)	1.850(1.002)	1.079(0.673)	1.942(0.872)
β	-0.016(0.227)	-0.016(0.200)	0.030(0.213)	0.030(0.205)
α	-0.758(0.564)	-1.387(0.637)	-0.732(0.421)	-1.095(0.517)
σ	0.673(0.075)	0.673(0.075)	0.666(0.075)	0.666(0.075)
ρ	0.593(0.153)	0.741(0.108)	0.220(0.194)	0.545(0.158)
RE	0.0217(-0.553;0.597)	0.011(-0.266;0.290)	-0.041(-0.605;0.523)	-0.027(-0.386;0.331)

Table 5: Parameter estimates (standard errors) for the Placket-Dale model.

Parameter	**Binary 1**	**Binary 2**	**Binary 3**	**Binary 4**
μ_T	-2.042(0.35)	-1.988(0.29)	-2.122(0.347)	-2.141(0.335)
μ_S	0.342(1.175)	2.143(0.993)	1.741(1.065)	2.737(1.153)
β	0.026(0.216)	-0.003(0.181)	0.068(0.218)	0.091(0.213)
α	-0.994(0.794)	-1.581(0.640)	-1.179(0.668)	-1.566(0.694)
σ	0.679(0.075)	0.679(0.075)	0.670(0.075)	0.671(0.075)
$ln(\psi)$	1.591(0.611)	2.663(0.745)	0.581(0.623)	1.874(0.670)
RE	-0.026(-0.440;0.387)	0.00216(-0.221;0.225)	-0.0585(-0.417;0.300)	-0.0581(-0.307;0.191)

The results of the probit and Plackett-Dale model are shown in Tables 4 and 5. It is clear from these results that all RE's are estimated very imprecisely. In particular, all 95% confidence intervals contain zero. This confirms that even the overall sample size is too small for useful surrogate marker validation. The conclusion from this is two-fold. First, one should aim for the largest possible sample sizes. Second, there is a clear need for alternative frameworks. One such framework is provided by the Bayesian paradigm; this is a topic of ongoing research.

Practically, we cannot make any definitive statement about the surrogacy quality of the postulated candidates. As it was said earlier, future investigations, based on more available data, will be needed.

5 Concluding Remarks

The assessment of surrogate biological markers has been addressed in the work of Choi *et al.* (1993) and Lin *et al.* (1997), evaluating CD4 count in AIDS patients. With the new wave of cancer vaccines and biological response modifiers, the search for surrogate endpoints has become an important issue in cancer research as well.

In this paper, we have contributed towards assessment of immunogenicity as a surrogate of survival, given by the titer achieved by patients after vaccination, either at the end of treatment or through the maximal level achieved. Our evidences of a possible relationship of the immune response to the EGF vaccine with survival came from two pilot trials in NSCLC (Torres *et al.*, 2001). The results obtained in the present work do not support the anti-EGF immune response as a good surrogate of survival time. In this work, two trials comprising 40 patients in total, were analyzed for surrogate validation but this size was not sufficient to obtain precise results. Although some convergence problems were encountered, these could be solved with the simplified methods proposed by Tibaldi *et al.* (2001). To this end, we had to result to continuous-outcome methods rather than survival-type ones. This was justified due to the advanced state of the disease, for which reason there were no problems of censoring.

Also the small size had an influence in the wide confidence intervals obtained. The minimum size for the evaluation of surrogate variables has not been deeply studied. A logical reasoning would require as much evidence as possible, but it might be the case of the existence of a very strong surrogate variable for which it is not necessary a bigger size. We believe this particular issue should be address in further studies.

In this work, several candidate surrogates were proposed for the assessment of surrogacy, as suggested by Molenberghs *et al.* (2000), to get a stronger and less subjective analysis. The discretization of the variables did not have an impact on the final results; the results were consistent either considering the actual titer value or its dichotomized version.

Although there were not really enough trials nor patients, the evaluation of surrogacy is helpful to assess how strong the immune relationship was with survival, based on the evidence accumulated

so far. Other trials with the vaccines are ongoing, and more evidence regarding the relationship of the immune response and survival time will be gathered. Extrapolation of results to other cancer vaccines can be difficult since each vaccine has its own mechanism of action or activation site. Probably, vaccines with a similar receptor family as the one approached by the EGF vaccine will have a similar behavior.

Molenberghs *et al.* (2000) identified practical difficulties of surrogate evaluation and the need for several trials and larger sizes. One possible way to overcome this is by the implementation of full Bayesian models. The latter is a topic of further research.

Acknowledgments

The first author wish to thank "International Union Against Cancer" for the ICRETT grant. The second and sixth authors wish to thank "Bijzonder Onderzoeksfonds" of the Limburgs Universitair Centrum. Research supported by a PAI program P5/24 of the Belgian Federal Government (Federal Office for Scientific, Technical, and Cultural Affairs).

References

Anderson, J.R., Cain, K.C., and Gelber, R.D. (1983) Analysis of Survival by Tumor Response. *Journal of Clinical Oncology*, **1**, 710–719.

Burzykowski, T., Molenberghs, G, and Buyse, M. (2001) The validation of surrogate endpoints using data from randomized clinical trials: A case study in advanced colorectal cancer. *Submitted for publication.*

Burzykowski, T., Molenberghs, G., Buyse, B., Geys, H., and Renard, D. (2001) Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. *Applied Statistics/*, **50**, 405–422.

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. (2000a) The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, **1**, 1–19.

Buyse, M. and Piedbois, P. (1996) On the relationship between response to treatment and survival. *Statistics in Medicine*, **15**, 2797–2812.

Buyse, M., Thirion, P., Carlson, R.W., Burzykowski, T., Molenberghs, G., and Piedbois, P. (2000b) Tumour response to first line chemotherapy improves the survival of patients with advanced colorectal cancer. *Lancet*, **356**, 373–378.

Choi ,S., Lagakos, S. W., Schooley, R. T. and Volberding, P. A (1993) CD4+ lymphocytes are an incomplete surrogate marker for clinical progression in persons with asymptomatic HIV infection taking zidovudine. *Annals of internal Medicine*, **118**, 674–680.

Ellenberg, S.S. and Hamilton, J.M. (1989) Surrogate endpoints in clinical trials: cancer. *Statistics in Medicine*, **8**, 405–413.

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

González, G., Crombet, T., Catalá, M., Mirabal, V., Hernandez, J., González, Y., Marinello, P., Guillén, G., and Lage, A. (1998) A novel cancer vaccine composed of human-recombinant epidermal growth factor linked to a carrier protein: Report of a pilot clinical trial. *Annals of Oncology*, **9**, 431–435.

Gonźalez, G., Crombet, T., Catalá, M., Torres, F., Lage, A. (2001) Epidermal growth factor cancer vaccine for non small cell lung cancer therapy: Results of two phase I clinical trials. *In: Proceedings ASCO meeting 2001*.

Lin, D.Y., Fleming T.R., and De Gruttola, V. (1997) Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine*, **16**, 1515–1527.

Molenberghs, G., Buyse, M., Geys, H., Renard, D., and Burzykowski, T. (2001) Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Submitted for publication.*

Molenberghs, G., Geys, H., Buyse, M. (2001) Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine*, **20**, 3023–3038.

Plackett, R.L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**,516–522.

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

Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., Buyse, M., Vangeneugden, T., and Bijnens, L. (2001) Validation of a longitudinally measured surrogate marker for a time-to-event endpoint.*Journal of Applied Statistics*, **00**, 000–000.

Tibaldi, F., Cortiñas, J., Molenberghs, G., Renard, D., Burzykowski, T., Buyse, M., Parmar, M., Stijnen, T., and Wolfinger, R. (2001) Simplified hierarchical linear models for the evaluation of surrogate endpoints. *Submitted for publication.*

Torres, F., Gonzàlez, G., Crombet, T., and Lage, A. (2001) Use of the Bayesian approach in the analysis of clinical trials in patients with advanced lung cancer. *Submitted for publication.*

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Model for Longitudinal Data.* New York: Springer-Verlag.

Franz Torres Barbosa

Gisela González

Center of Molecular Immunology

La Habana

Cuba

Fabián S. Tibaldi

José Cortiñas Abrahantes

Helena Geys

Geert Molenberghs

Center for Statistics

Limburgs Universitair Centrum

Universitaire Campus, Building D

B3590 Diepenbeek

Belgium