Article

# IDENTIFICATION OF INTERACTION TESTS OF PREDICTIVE BIOMARKERS IN CANCER CLINICAL TRIALS

V. Ndahindwa[1*]

[1]*Department of Biostatistics and Epidemiology, School of Public Health, National University of Rwanda*

## ABSTRACT

Predictive biomarkers are covariates that interact with treatment in relation to the outcome and thus, predictive biomarkers are characterized by interactions between the treatment and covariates. Many questions remain unanswered in this topic: What is the best design for detecting and validating a predictive biomarker? What can be the sample size required? What could be the statistical methods used to identify those interactions? The major problem of interaction tests is that they lack power, so that a very large trial would be required for the test to reach significance. The identification of a predictive factor becomes difficult if interactions of higher orders have to be investigated. We discussed the use of Martingale residuals combined with the classification and regression trees (CART) to identify which could be the optimal cut point in a continuous marker through data simulation. Our findings using these methods were very close to the expected results given the parameters of the simulation. Our conclusion is that the CART applied to Martingale residuals can be the good alternative of identifying predictive biomarkers. In practice we may need a cut point for a predictive biomarker so that we can know who among patients can benefit from the treatment and those who may be harmed by the treatment, especially when drugs are highly toxic.

**Key words**: Predictive biomarker - Interaction - Martingale residuals - Classification tree - Clinical trials.

## RESUME

Les biomarqueurs prédictifs sont des covariables qui interagissent avec le traitement vis à vis du résultat et donc, les biomarqueurs prédictifs sont caractérisés par des interactions entre le traitement et les autres covariables. Plusieurs questions restent sans réponses sur ce sujet: Quelle est la bonne méthode pour détecter et valider un biomarqueur prédictif? Quelle serait la taille d'échantillon requise ? Quelles seraient les méthodes statistiques à utiliser pour identifier ces interactions? Le problème majeur des tests d'interaction est qu'ils ont une faible puissance statistique, et qu'une grande taille d'échantillon serait requise pour que le test soit significatif. L'identification d'un facteur prédictif devient plus difficile lorsqu'on étudie les interactions d'ordre supérieur. Nous avons discuté l'utilisation des résidus de Martingale combinée à l'arbre de régression et de classification (CART) pour identifier, à travers des simulations des données, le point de coupe optimal dans le cas d'un facteur continu. Nos résultats, utilisant ces méthodes, ont été très proches des résultats escomptés connaissant les paramètres de simulation. Notre conclusion est que le CART appliqué aux résidus de Martingale peut être une alternative dans l'identification des biomarqueurs prédictifs. En pratique, nous pouvons avoir besoin de déterminer un point de coupe pour un facteur prédictif pour que nous sachions qui parmi les patients peut bénéficier du traitement et ceux dont ce traitement peut être nocif, plus spécialement lorsque le médicament est hautement toxique.

**Mots clés** : Biomarqueur Prédictif - Interaction - Résidus de Martingale - Arbre de Classification - Essais cliniques.

## BACKGROUND

Biomarkers play an increasing role in the development of new treatments, mainly for cancer treatments. A biomarker can be formally defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [1, 2].

Biomarkers can be broadly classified as prognostic markers (associated with disease outcome) or predictive markers (associated with drug response) [3].

Predictive biomarkers are characterized by an interaction between the treatment group and covariates, where the treatment can be beneficial in one group of patients and be harmful in another group of patients with respect to their characteristics [4, 5]. We need therefore to test interactions between the effect of treatment and the biomarker status. However, the statistical significance is always the main challenge in the interaction identification.

In our case, we will consider a prospective way of identifying predictive biomarkers trough a randomized clinical trial.

We shall focus our attention on predictive biomarkers and discuss these challenges on interaction identification and try to bring some answers on what could be the statistical methods used to identify those interactions ? Our work will be based on clinical trials, mainly where survival analysis is used and the CART [6, 7] applied to Martingale residuals was used on simulated data.

## METHODS

As predictive biomarkers are widely used in cancer clinical trials, we assumed that our data concerned cancer patients, receiving 2 types of treatment, an experimental treatment compared to a standard treatment. We also assumed that some patients have a kind of mutation which was expressed by a high level of biomarker concentration.

*Correspondence to:
*Védaste Ndahindwa*
*School of Public Health*
*National University of Rwanda*
*Tel: +250 788 45 46 13*
*E-mail address: ndahindwa@nursph.org*

We have generated survival data from simulations. The aim of these simulations was to generate datasets that can be used to identify interactions between the treatment and covariates. Once interactions are specified by a Cox model, we need to identify a possible cut point justifying those interactions. We will introduce an interaction which is highly significant in a normally distributed biomarker with a predefined cut point, and compare the results of CART splits with the expected results.
The other covariates will be: age of the patient, the sex, the weight and the biomarker level which can be considered as the number of mutations as it is the case in cancer clinical trials. We will assume that patients were randomly assigned to two treatments (an Experimental treatment and a Standard treatment). The censoring variable will be death and the time will be assigned so that the survival curves follow an exponential distribution.

*Simulation software and Sample size*

We used the R software for simulation. R is a statistical computer program made available through the Internet under the General Public License (GPL). R provides an environment in which you can perform statistical analysis and produce graphics. As using function runif generates almost different sets of random numbers, we set the random number seed (set.seed(i)) before generating the number, in order to get the same results across data simulations, as we want the results to be reproducible.
For our scenario, the sample size was 500 patients (400 without mutation and 100 with mutation) in both treatment groups with a well balanced ratio 1:1.
Our variable of interest is "time" which is the time-to-failure in days. The dataset will help us to analyze time of death in days among the 2 treatment groups (Experimental treatment and Standard treatment). The censoring variable is death with 1 as dead or 0 as censored.

*Analysis*

We analyzed our data in the intention-to-treat population, and all data have been considered. The primary end point was considered as the progression-free survival as commonly used in cancer trials. The aim was to determine the interaction test using the Cox model. The Kaplan Meier estimations were performed to estimate survival curves in the two treatment groups and the log-rank test used to compare them.
We realized our analysis on the basis of Martingale residuals with the idea of identifying the optimal cut point for a biomarker. A Cox model was performed on patients under standard treatment, and this model was applied to the experimental treatment dataset. Then we get Martingale residuals which were divided into 2 groups and create a new variable named "target": positive martingale residuals were coded 1 and those which are negative or equal to zero were coded 0. We therefore performed a CART with target as our variable of interest. This method gave us the cut point for the covariates which are in interaction with the treatment. Each time we got a cut point, we tried to estimate Kaplan Meier curves within the two datasets created and compare differences
We used R software to analyze our simulated data. The survival analysis and Cox model were performed using the package called survival described by Terry Therneau [8] and ported to R software by Thomas Lumley [9], whereas CART were performed with the package named party.
Kaplan-Meier estimator for survival function was done with a function called survfit. Computing the log-rank test was done by the function survdiff of R software, whereas computing Cox model was performed by the function coxph.

## RESULTS

We would like to set one value to be considered as a cut point. We chose the value 225.65, which divides our data set into 2 sub data sets in the proportion of 0.80 and 0.20 respectively.

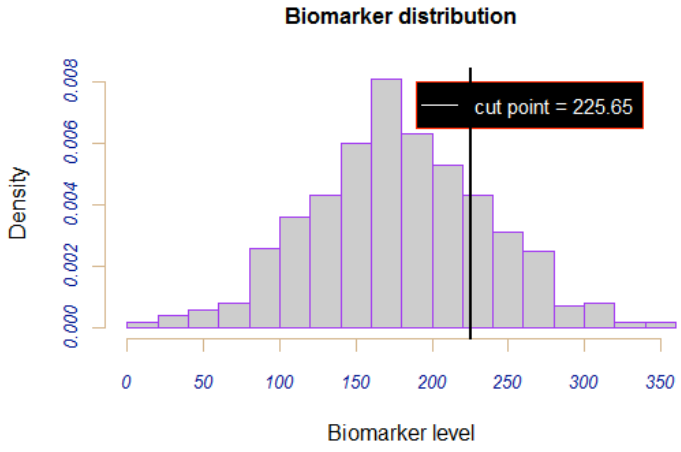


**Figure 1**: Biomarker distribution

The marker variable has the mean of 178.2 and standard deviation of 58.33 and is normally distributed. The minimum was 0.41 while the maximum was 357.95. The cut point was set to 225.65 so that the 20% upper values will be considered as mutation group.
The Figure 2(a) shows that the medians of the biomarker value among patients with and without mutation were different and were high in the mutation group as predefined (biomarker level above 225.65). However, patients were randomized in the two treatment groups with a well balanced distribution (Figure 2(b)).

Figure 3 shows the Kaplan Meier estimation of progression-free survival curves for the two treatment groups. The progression-free survival for Standard treatment group is less than the one for Experimental group. The log-rank test

(p = 0.0346) was significant to reject the null hypothesis of the equality over strata.
Hazard ratio (Standard versus Experimental) was 1.3025 and 95% CI = [1.018; 1.666], thus there is a significant difference between the two hazard rates at the significance level $\alpha = 0.05$ (p = 0.0965).

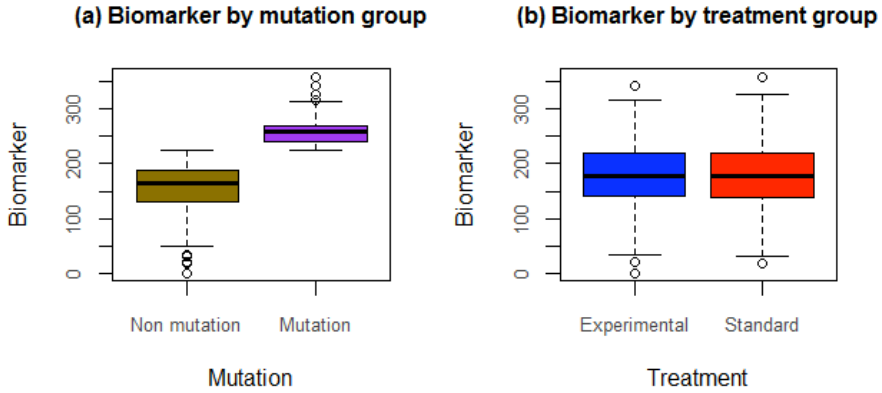


**Figure 2**: Box plots for the variable "biomarker" by mutation and treatment groups respectively

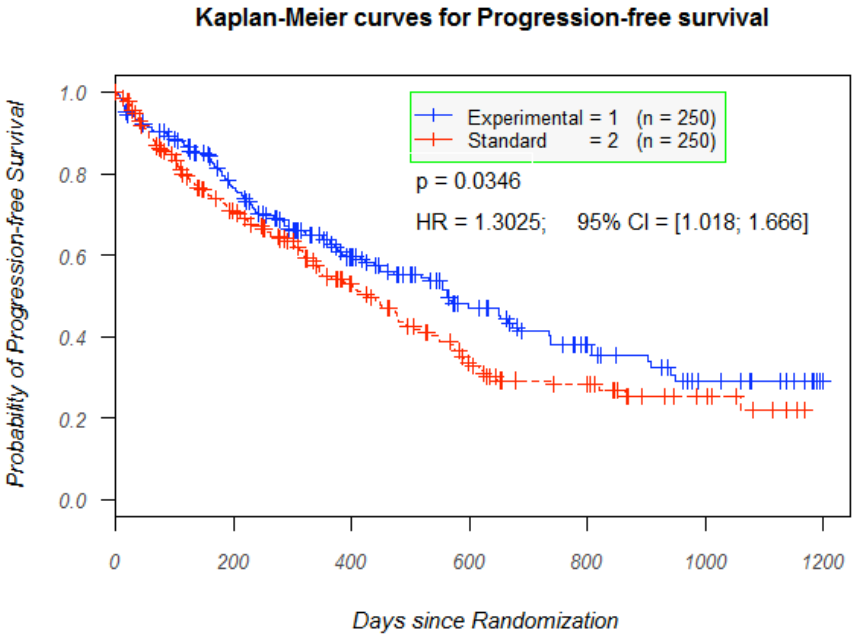


**Figure 3**: Kaplan-Meier curves for Progression-free survival

**Table 1**: Cox model fitting for all covariates

| Parameter | Parameter Estimate | Standard Error | Pr(>\|z\|) | Hazard Ratio | 95% Hazard Ratio Confidence Limits | |
|---|---|---|---|---|---|---|
| age | -0.026184 | 0.010404 | 0.011841 | 0.9742 | 0.9545 | 0.9942 |
| weight | -0.005896 | 0.015738 | 0.707919 | 0.9941 | 0.9639 | 1.0253 |
| treat | 1.579579 | 0.397665 | 7.12e-05 | 4.8529 | 2.2259 | 10.5802 |
| marker | 0.009984 | 0.003394 | 0.003265 | 1.0100 | 1.0033 | 1.0168 |
| treat*marker | -0.007383 | 0.002128 | 0.000522 | 0.9926 | 0.9885 | 0.9968 |

We have tested all covariates using Cox model including interactions between treatment and biomarker. There was a high interaction effect between the treatment and the biomarker at the significance level $\alpha$ =0.05. We therefore have reason to believe that the biomarker level can be the predictive biomarker and we need to use Martingale residuals and CART to try to find whether there can be some cut point to distinguish how patients behave on treatment with respect to their characteristics.

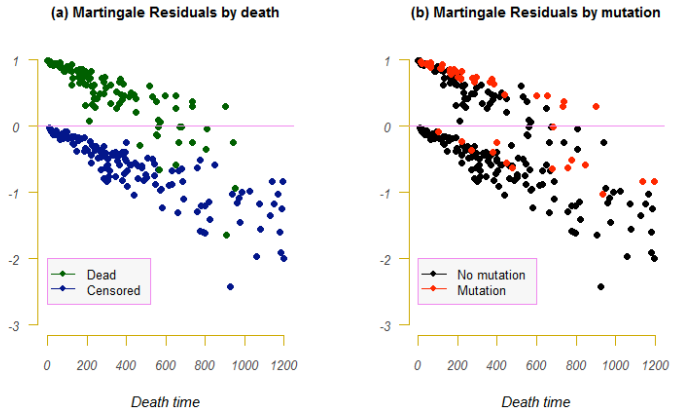


**Figure 4**: Plotting Martingale residuals to detect a predictive factor

**Figure 4(a)** and **Figure 4(b)** show the Martingale residuals by mutation group and by death group respectively

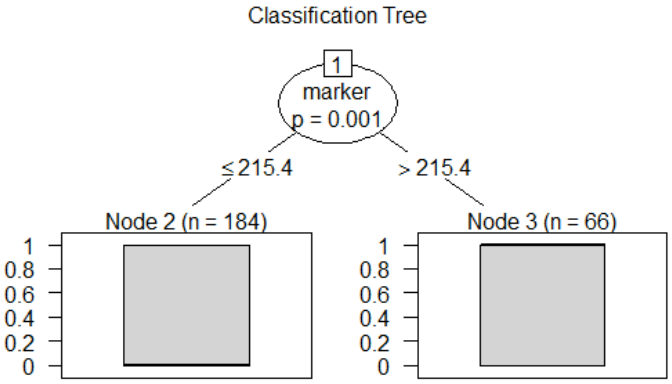


**Figure 5**: Classification tree

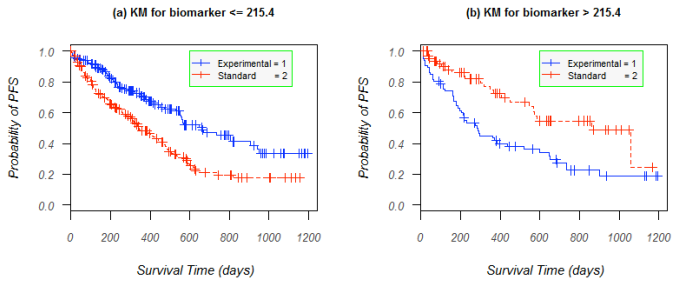


**Figure 6**: Kaplan-Meier curves for Progression-free survival

Our target (positive Martingale residuals) represents 41% of the Experimental sub-dataset. Using the CART, we can therefore get cut point values to be used in our database. The most important variable is "biomarker" and was used to split the experimental sub-data set. The cut point value is 215.4, which is close to our predefined cut point of 225.65. We can then estimate the Kaplan-Meier curves according to treatment groups and see whether there is any difference. The difference is well shown on the figure 6 where one can say that the level of marker can be the predictive for progression-free survival benefit between the Experimental and Standard treatments: PFS seems to be longer on Standard treatment when there is mutation and shorter on Experimental treatment when there is mutation

## DISCUSSION

In this work, we focus our attention on predictive biomarkers in cancer clinical trials. The identification of predictive factors is of great interest in Medicine. Clinicians want to know which therapy will be effective in a particular patient [10]. Predictive biomarkers, which predict the likely response of patients to specific treatments, require more extensive data for validation, specifically large randomized clinical trials and meta-analysis [5]. We use the term predictive to describe an interaction between a factor and a treatment [11]. Interactions of high order are the ones that are difficult to identify.

The main question is why a patient under the new therapy lives longer or shorter than the average or than patients under the standard treatment. Several reasons are possible: (1) it could be by chance; (2) the patient could have a better or a worse prognosis than the average or (3) there could be some patients with a positive or negative reaction to the new treatment [10], hence the role of predictive factor for the difference in survival [12].

The main objective of our study was to assess the interaction effect when we have a continuous factor using methods described in the literature. We chose to use the Martingale residuals approach combined with the classification and regression tree. This method proposed by Ulm et al. (2006) suggests that it is possible to detect also interactions of high order [10].

We applied this approach on simulated data. The idea of this procedure is to test whether a good model which describes well the patients on standard treatment (say, prognostic model) can also fit with the patients on the new treatment. Patients whose outcomes are not well described by the prognostic model are the ones we are interested in. The deviation from the prognostic model is measured by Martingale residuals and we look for similarities among those patients using the CART. Patients who have event after predicted time by the model have large negative Martingale residuals. These patients may benefit from the new treatment. On the other hand there will be patients who have event before predicted time. These patients will have positive Martingale residuals and may be harmed

by the new therapy [10]; all censored patients will have Martingale residuals which are negative or equal to zero.
In the study, we analyzed simulated data where the experimental treatment was overall superior to the standard treatment. However, it was possible to identify a subgroup of patients who seem to be harmed by the new therapy or in other words patients who do not benefit from the new treatment.
One should keep in mind that theses analysis have to be taken into account with additional considerations as the results might be due the chance. Therefore, it is the best advice to report them sceptically as hypotheses to be investigated in other studies and to validate them [13]. Assmann et al in 2000 state that the investigators should be cautious when undertaking subgroup analysis: the findings are always exploratory and one should avoid over interpretation [14]. There can be also the effect of small sample size across subgroups and not really a clinical difference among treatment groups.
With CART, we can get a cut point to be used in splitting patients among 2 different sub groups with respect to the reaction on the treatment. In practice, a cut point is needed to know which patient can benefit from the treatment and which patient can be harmed by the treatment.
The determination of a cut point is based to a single fitted model, and the generalization of the cut point can bring some issues. There is a potential of over fitting if an optimal cut point is determined in one dataset. We need to know whether the method is robust enough so that the cut point can be accepted as optimal; otherwise, the cut point should be restricted to data which have been used to estimate it.
The method of Martingale residuals combined with CART is useful when we have a continuous biomarker variable and when we think there is a possibility of high order interactions. A simple Cox model is useful to detect interactions, but in practice, we need also to identify which patient can benefit from the new treatment. We can therefore estimate hazard ratios among different subgroups of patients and compare them.
We can not therefore rule out that the results are influenced by the play of chance. There is a need to do many simulations to get a distribution (rather than a single estimate) for the parameters of interest. Many simulations could confirm the robustness of estimates and thus, we recommend another study to pursue this work. No consensus yet exists on processes or standards for pragmatic evaluation and adoption of biomarkers and surrogate end points in the absence of robust statistical validation [5].

## CONCLUSION

This study shows that the Martingale residuals combined with the CART can be used to identify interaction effect between treatment and the biomarker. We simulated data where the biomarker was a continuous variable and got cut points using CART procedure. Estimated cut points were close to those expected. We can not rule out the effect of chance in the estimation of optimal cut point as we analyzed each time a singe simulation. Another study is needed in order to make a large number of simulations and get a distribution of parameters: this can give an idea on how the approach is robust and how an optimal cut point can be estimated. It is also important that the methods be applied on real data and be compared to other available

## REFERENCES


1. Lesko, L.J.S., Atkinson AJ. Use of biomakers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. An Rev Pharm Toxic. 2001 ;41:347-366.
2. Buyse M. Towards the validation of statistically reliable biomarkers. Eur J Cancer Supplements. 2007 ;589-95.
3. Mandrekar, S.J., Sargent DJ. Clinical Trial Designs for Predictive Biomarker Validation: Theoretical Considerations and Practical Challenges. J Clin Oncol. 2009 ;27(24):4027-4034.
4. Bonetti, M, David Zahrieh, D., Cole, B.F., Gelber RD. A small sample study of the STEPP approach to assessing treatment–covariate interactions in survival data. Statistics in Medicine. 2009 ;28(8):1255-1268.
5. Buyse, M., Sargent DJ. Biomarkers and surrogate end points—the challenge of statistical validation. Clin Oncol. 2010 ;7:309-317.
6. Breiman, L., Friedman, J.H., Olshen, R.A., Stone CG. Classification and Regression Trees. Belmont, California, USA: Wadsworth International Group; 1984.
7. De'Ath, G., Fabricius KE. Classification and Regression Trees: A Powerful Yet Simple Technique For Ecological Data Analysis. Ecology. 2000 ;81(11):3178–3192.
8. Therneau T. A Package for Survival Analysis in S. 2002 ;
9. Dalgaard P. Introductory Statistics with R. New York: Springer; 2002.
10. Ulm, K., Kriner, M., Eberle S. Statistical Methods to Identify Predictive Factors. In: Crowley, J. Ankerst D, editor(s). Handbook of Statistics in Clinical Oncology. London: Chapman & Hall/CRC; 2006. p. 289-382.
11. Royston, P., Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Stat Med. 2004 ;23(16):2509-2525.
12. Sargent, D.J., Conley, B.A., Allegra C. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol. 2005 ;23:2020-2027.
13. Byar D. Assessing apparent treatment: covariate interactions in randomized clinical trials. Statistics in Medicine. 1985 ;4255-263.
14. Assmann SF. Subgroups analysis and other (mis)uses of baseline data in clinical trials. The Lancet. 2000 ;355:1064-1068.