# MathDigest

Research Bulletin of Institute for Mathematical Research

VOLUME 3 NUMBER 2 ISSN : 1985-2436

JUN 2010 JAMADILAKHIR 1431H

## Research Articles

UPM

UNIVERSITI PUTRA MALAYSIA

BERILMU BERBAKTI

Research Bulletin of Institute for Mathematical Research

MathDigest

## :: EDITORIAL BOARD

## CONTENTS ::

# Optimal Cutoff Point for Dichotomization of Continuous Independent Variable in Competing Risks Data Analysis

Abdul Kudus and Noor Akma Ibrahim

Department of Statistics, Universitas Islam Bandung
Jl. Purnawarman No. 63, Bandung 40116, Indonesia
&
Laboratory of Applied and Computational Statistics
Institute for Mathematical Research
Universiti Putra Malaysia

akudus69@yahoo.com

## Abstract

In the time-to-event analysis, particularly for competing risk data analysis, it may be desirable to dichotomize continuous predictor variable(s) since the functional form between the predictor variable and competing risk survival time data need not be assumed. There may be thresholds for the predictor variables and the statistical inference might be more robust when the predictor variables are dichotomized. By transforming a continuous predictor variable into a categorical variable, usually binary, we will come up with a more interpretable model. Maximum Gray's statistic method is considered for the cutoff point determination in this paper. This statistic is useful for comparing cumulative incidence function for the main cause of interest, therefore we will have two groups with different subdistribution survival times.

## Introduction

In the applied statistics fields, it may be desirable for the independent continuous variables in descriptive, univariate, or multivariate analysis to be viewed as categorical variables. The reasons to dichotomize such variables are because the functional form between the independent and dependent variables need not be assumed and there may be thresholds for the independent variables.

Association between independent variable and time to event of a particular type in competing risk setting may be inferred via relative risks. For a continuous independent variable the relative risk will indicate the magnitude of increased risk of the independent variable. This interpretation assumes that the functional form between independent and dependent variables is known. However, for many independent variables, its functional relationship to time of event of a particular type is unknown.

If the continuous independent variable is dichotomized, a general functional relationship can be represented depending on the number of categories selected for that variable. In a medical setup, the results from a study which dichotomizes a continuous independent variable will show the relative risk by comparing the risk of getting a particular type of event in specific categorical levels of the independent variable with the risk of getting a particular type of event for a selected reference category. This interpretation may be easier to comprehend by the physician and patient.

Some independent variables might have thresholds in predicting the outcomes. That is, the relationship between an independent variable and an outcome variable might not be represented adequately by a continuous function. Dichotomizing a continuous independent variable with thresholds might allow us to investigate the association between independent variables and the outcome with more validity.

Thus, for the above reasons, many researchers have dichotomized continuous variables in many areas of applied statistics studies. However, "What is the correct method of dichotomization?" is a critical question, because there are so many different ways to determine cutoff point for continuous independent variable dichotomization. Some cutoff points are decided on the basis of prior biological or physiological knowledge, and others are based on outcome-dependent methods. There is no clear rule to specify which dichotomization is the best and results of analysis from a variety of dichotomizations may be different.

Most researchers prefer to use prior knowledge for deciding on cutoff points. These cutoff points may be obtained from biological observations, clinical studies, or physicians' experiences. However, there are some problems in using prior knowledge. The first problem is that people do not always know whether the thresholds or reference ranges are from scientific sources or not. The second problem is that these thresholds cannot always be applied to people with different characteristics. Therefore, when a dichotomization is used in a study, the existing criteria for its original selection might not be appropriate for the new and different situations. Problem in using prior knowledge to determine the appropriate dichotomization in some specific areas of applied statistics is that many risk factors have not been studied, so the cutoff point (or cutoff points) for them are not clear.

We propose the use of outcome-oriented cutoff points in competing risk data analysis as have been done for survival data by Mandrekar *et al.* [4]. Statistical methods applied to decide on cutoff points are based on the characteristic of the time to event of main interest. We use two-sample test from Gray [2] to find an optimal cutoff point which does not require the development of a regression model as studied by Ibrahim et al. [3].

## Outcome-Oriented Cutoff Point Determination in Competing Risk Data Analysis

The method of cutoff point determination based on Gray's test statistic for comparing the cumulative incidence is not straight forward due to the statistical issues made complicated by the presence of multiple events. Here, each subject may fail due to one of several possible causes called competing risks. A competing risk can be defined as an event whose occurrence precludes the occurrence of other events under examination. Some examples of competing risks is cause-specific mortality, such as death from heart disease and death from cancer, where deaths from other causes (for example old age) are the competing risks.

Interest is often on estimating the rate of occurrence of the competing risks and comparing these rates between groups of subject and modeling the effect of some other factors on the rate of the competing risks. Here, we will focus our attention on comparing probabilities of a specific event in two distinct groups through Gray's test.

*Competing Risks Data*

Consider competing risks data where each subject may fail due to one of $J$ ($J \geq 2$) causes. Occurrence of one event precludes observation of the other events (it is assumed that subject can fail only from one cause).

The latent failure time approach is one of the ways to describe competing risks data. Here, competing risks are represented by a set of positive random variables $X_1,\ldots,X_J$ with $X_j$ being a potential (unobservable) time to occurrence of the $j^{\text{th}}$ competing risk. We observe the time at which each subject fails from any cause, $T = \min(X_1,\ldots,X_J)$, and an indicator $\delta$ indicating which of the $J$ risks caused the failure, i.e., $\delta = j$ if $T = X_j$.

A fundamental parameter in competing risks data analysis is the *cumulative incidence function* which is defined as follows:

$$F_j(t) = P(T \leq t, \delta = j) = \int_0^{\alpha} S_T(u)\,\alpha_j(u)\,du,$$
$$j = 1,\ldots,J \qquad (1)$$

where $S_T$ is the overall survival function of $T$, that is, $S_T(t) = P(T \geq t)$ and $\alpha_j$ is the *cause specific hazard rate* for risk $j$, defined by

$$\alpha_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = j \mid T \geq t)}{\Delta t},$$
$$j = 1,\ldots,J \qquad (2)$$

Note that the value of $F_j(t)$ depends not only on the rate at which the specific cause of interest is occurring, but also on the rates at which all the competing risks occur. $F_j(t)$ is called a "subdistribution" function because it is not a true distribution function due to its properties: it is non-decreasing, $F_j(0) = 0$ and $F_j(\infty) = P(\delta = j) < 1$. These curves have a straightforward interpretation. They are probabilities of experiencing death from the $j$th cause in the setting where competing risks are acknowledged to exist.

For further development, it is convenient to introduce the counting process notation. A formal and rigorous survey of counting processes and their applications can be found in [1]. Here, we will introduce the notation and approaches to be used in the rest of the paper.

Define the process $Y^i(t) = I(T_i \geq t)$ as an indicator of $i$ being at risk just before time $t$. The total number of subjects at risk at time $t$ is $Y(t)=\sum_{i=1}^{n} Y^i(t)$, where $i = 1,\dots, n$. Consider a counting process:

$$N_j^i(t)=I(T_i \le t, \delta_i = j) \tag{3}$$

Note that $N_j^i(t)$ is a step function, which is zero until $i$ dies from cause $j$ and then jumps to one. The process $N_j(t)=\sum_{i=1}^{n} N_j^i(t)$ is also a counting process which simply counts the number of failures of type $j$ in the sample at or prior to time $t$. Throughout, the subscript replaced by "•" will denote summation over that index. After adopting this notation, the total number of failures by time $t$ is

$$N_\bullet(t)=\sum_{j=1}^{J} N_j(t).$$

In the counting process notation, the data $(T_i, \delta_i)$, i = 1,…, n, are represented by $\{Y^i(\cdot), N_j^i(\cdot)\}$, i = 1,..., n, j = 1,..., J. Based on these data we can estimate some quantities in the competing risk data analysis.

*Gray's Test for Comparing Cumulative Incidence*

Suppose there are K independent groups of subjects, where $k^{th}$ group consists of $n_k$ subjects and $n=\sum_{k=1}^{K} n_k$ . Each subject may fail due to one of J ($J \ge 2$) competing causes. Indeed, it is enough to consider the case where there are only two types of failure. This does not place any restriction on the generality of the results, since when there are more than two types of failure, all types other than the type of interest can be combined into one "other" category while analyzing the event of interest. The failure type of special interest is taken to be type 1. In general, data will be right censored. For the $i^{th}$ subject in group k, $i = 1,\dots, n_k$, $k = 1,\dots, K$, let $T_{ik}$ be the failure time and $\delta_{ik}$ be the cause of the removal from the study:

$$\delta_{ik} = \begin{cases} 0, & \text{if subject was censored,} \\ 1, & \text{if subject failed from cause of interest,} \\ 2, & \text{if subject failed from other causes.} \end{cases}$$

The pairs $(T_{ik}, \delta_{ik})$ from different subjects in a group are assumed to be independent and identically distributed. However, it is not assumed that the underlying processes leading to failures of different types are acting independently for a given subject. Censoring mechanism will be considered to be independent of competing risks acting in the population.

Notation used in the previous section will be extended to accommodate indicator of a group being considered. That is $F_{jk}$ will be used to define subdistribution or cumulative incidence function for failures of type $j$ in group $k$,

$$F_{jk}(t)=P(T_{ik} \le t, \delta_{ik} = j) \tag{4}$$

Counting process notation will be extended in a similar fashion. Let $i$ be the index for subject belonging to group $k$, then we define $Y_k^i(t)=I(T_{ik} \ge t)$ as an indicator of being at risk just before time $t$ and indicator of experiencing failure from cause $j$ by time $t$ is $N_{jk}^i(t)=I(T_{ik} \le t, \delta_{ik} = j)$.

Define

$$N_{jk}(t)=\sum_{i=1}^{n_k} N_{jk}^i(t) \tag{5}$$

and

$$Y_k(t)=\sum_{i=1}^{n_k} Y_k^i(t) \tag{6}$$

Then the $N_{jk}(t)$ is the number of failures of type $j$ by time $t$ and $Y_k$ is the number of subjects still at risk just prior to $t$ in group $k$. The process $N_{j\bullet}(t)=\sum_{k=1}^{K} N_{jk}(t)$ counts the number of failures of type $j$ in all samples by time $t$ and the number of subjects still at risk just prior to $t$ in the pooled sample is given by $Y_\bullet(t)=\sum_{k=1}^{K} Y_k(t)$.

The cumulative incidence function is the primary measure summarizing the likelihood of a specific event in the competing risks setting. Differences in the cumulative incidence curves would reflect differences in the probabilities of a specific event being observed in distinct populations in the presence of other competing risks. Next we will present Gray's technique to compare the cumulative incidence functions.

Without loss of generality we will assume that there are only two types of failure ($J = 2$). The failure type of special interest is taken to be type 1. Consider a problem of comparing the cumulative incidence functions for the cause of interest among $K$ ($K \ge 2$) populations. Inference will be based on a sample of size $n$.

The hypothesis of interest is:

$H_0 : F_{11}(t) = \dots = F_{1K}(t) = F_1^0(t)$, for all $t \le \tau$

versus
$H_A$ : at least one of the $F_{1k}(t)$'s is different for some $t \leq \tau$,

where $F_1^0(.)$ is an unspecified subdistribution function. Inference is on the cumulative incidence functions for all time points less than $\tau$, which is usually taken to be the largest time on study. The $F_{jk}(t)$'s are assumed to be continuous with subdensities $f_{jk}(t)$.

Gray [2] developed a class of test statistics for making comparisons between cumulative incidence functions. The test statistic is based on the (improper) random variable, $X_{ik}^*$, $i = 1,\ldots, n_k$, $k = 1,\ldots, K$. This random variable is defined by

$$X_{ik}^* = \begin{cases} T_{ik}, & \text{if } \delta_{ik} = 1, \\ \infty, & \text{if } \delta_{ik} > 1. \end{cases} \tag{7}$$

Then $P(X_{ik}^* \leq t) = P(T_{ik} \leq t, \delta_{ik} = 1) = F_{1k}(t)$ and the hazard rate for $X_{ik}^*$ is $\gamma_{1k}(t)$ given by

$$\gamma_{1k}(t) = \frac{dF_{1k}(t)/dt}{1 - F_{1k}(t)} = \frac{f_{1k}(t)}{1 - F_{1k}(t)} \tag{8}$$

Let $\hat{F}_{1k}$ be the estimated cumulative incidence function for cause 1 on sample $k$ and $\hat{F}_1^0(t)$ be a similar estimator based on the pooled sample. Let $\hat{S}_k(t-)$ be the left-hand limit of the Kaplan-Meier estimate of the overall survival function in sample $k$ obtained by considering failure from any cause as an event. $\hat{S}_k(t-)$ is defined to be 0 when $Y_k(t) = 0$ and the convention 0/0 = 0 is employed.

The $K$ sample statistic will be defined by assigning a score to each group which compares subdistribution hazard $\gamma_{1k}(t)$ for each group to a combined estimate of this hazard under the null hypothesis. Define

$$R_k(t) = \frac{I(\tau_k \geq t) Y_k(t) \left(1 - \hat{F}_{1k}(t-)\right)}{\hat{S}_k(t-)} \tag{9}$$

The quantity $\tau_k$ represents the largest time on study in group $k$. An estimate of the cumulative subdistribution hazard function for the cause of interest in sample $k$, $\Gamma_{1k}(t) = \int_0^t \gamma_{1k}(u)du$, is given by

$$\hat{\Gamma}_{1k}(t) = \int_0^t \frac{d\hat{F}_{1k}(u)}{1 - \hat{F}_{1k}(u-)} = \int_0^t \frac{dN_{1k}(u)}{R_k(u)}, \quad \text{for } t \leq \tau_k \tag{10}$$

The expression for $\hat{\Gamma}_{1k}$ suggests taking

$$\hat{\Gamma}_1^0(t) = \int_0^t \frac{dN_{1\bullet}(u)}{R_\bullet(u)} \tag{11}$$

as an estimator for $\Gamma_1^0$, the null value of $\Gamma_{1k}$. This estimator is consistent under the null hypothesis.

$K$ sample tests are based on scores of the form

$$Z_k = \int_0^{\tau_k} W_k(t) \left\{ d\hat{\Gamma}_{1k} - d\hat{\Gamma}_1^0 \right\}, \tag{12}$$

where $W_k(\cdot)$ is suitably chosen weight function. When the null hypothesis is true, $\mathbf{Z} = (Z_1,\ldots,Z_k)'$ has an asymptotic $K$-variate normal distribution with zero mean and covariance matrix $\Sigma$ which can be consistently estimated by $\hat{\Sigma}$ with components given by

$$\begin{aligned} \hat{\sigma}_{jj'}^2 = & \sum_{k=1}^{K} \int_0^{\tau_j \wedge \tau_{j'}} a_{jk}(t) a_{j'k} h_k^{-1} d\hat{F}_1^0(t) \\ & + \sum_{k=1}^{K} \int_0^{\tau_j \wedge \tau_{j'}} b_{2jk}(t) b_{2j'k} h_k^{-1} d\hat{F}_{2k}(t) \end{aligned} \tag{13}$$

where

$$a_{jk}(t) = d_{jk}(t) + b_{1jk}(t),$$

$$b_{ljk}(t) = \left[ I(l = 1) - \frac{1 - \hat{F}_1^0(t)}{\hat{S}_k^0(t-)} \right] \left[ c_{jk}(\tau_j) - c_{jk}(t) \right],$$

$$c_{jk}(t) = \int_0^t d_{jk}(u) d\hat{\Gamma}_1^0(u),$$

$$d_{jk}(t) = \frac{W_j(t) \left[ I(j = k) - \frac{\hat{h}_k(t)}{\hat{h}_\bullet(t)} \right]}{n(1 - \hat{F}_1^0(t))} \tag{14}$$

Here, $\hat{h}_k(t) = \dfrac{I(t \leq \tau_k) Y_k(t)}{n \hat{S}_k(t-)}$ .

In practice the weight function $W_k(t)$ is generally of the form $L(t)R_k(t)$, for some function $L(t)$. In this case, $\sum_{k=1}^{K} Z_k = 0$, so only $K - 1$ of the scores are linearly independent. An appropriate $K$-sample test statistic can then be formed by using a quadratic form consisting of $K - 1$ components of $Z$ and their estimated variance-covariance matrix $\hat{\Sigma}_0$ :

$$\chi^2 = (Z_1(\tau),\ldots,Z_{K-1}(\tau))\hat{\Sigma}_0^{-1}(Z_1(\tau),\ldots,Z_{K-1}(\tau))' \qquad (15)$$

When the null hypothesis is true, this statistic has an asymptotic chi-squared distribution with $K - 1$ degrees of freedom.

The form of the test statistic (12) is clearest when only two groups are being compared. For this case it is proposed that the test is based on a score of the form

$$\int_0^\tau W(t)\left[\frac{d\hat{F}_{11}(t)}{1-\hat{F}_{11}(t-)} - \frac{d\hat{F}_{12}(t)}{1-\hat{F}_{12}(t-)}\right], \qquad (16)$$

where $W(\cdot)$ is a weight function. This statistic compares weighted averages of the subdistribution hazards $f_{1k}/(1-F_{1k})$ in two groups. With the $W_k(t)$ in (12) being of the form $L(t)R_k(t)$, and setting $W(t) = L(t)R_1(t)R_2(t)/[R_1(t) + R_2(t)]$ in (16), it can be verified that (12) has the desirable property of reducing to (16) when only two groups are being compared.

## Simulation

Competing risk data were generated using the absolutely continuous bivariate exponential $(\lambda_0,\lambda_1,\lambda_2)$ distribution with probability density function:

$$f(t_1,t_2) = \begin{cases} \dfrac{\lambda_1\lambda(\lambda_2\lambda_0)}{\lambda_1+\lambda_2}\exp[-\lambda_1 t_1 - (\lambda_2+\lambda_0)t_2], \text{if } t_1 < t_2 \\ \dfrac{\lambda_2\lambda(\lambda_1\lambda_0)}{\lambda_1+\lambda_2}\exp[-\lambda_2 t_2 - (\lambda_1+\lambda_0)t_1], \text{if } t_1 > t_2 \end{cases}$$

Censoring time was generated from the exponential distribution with parameter $\lambda = 1$ and 3 corresponding to "light" and "heavy" censoring.

A single covariate $X_i = i$ ; $i = 1,\ldots,100$ was considered, and the failure time distribution was $F_1$ if $i \leq 50$ and $F_2$ if $i > 50$. This scenario reveals that the true cutoff point for independent variable $X$ is 50. We have four scenarios in comparing $F_1$ and $F_2$ with their corresponding parameter values as presented in Table 1. The simulation was run 1000 times.

**Table 1:** Model for the Cutoff point Simulation

| Model | $F_1$<br>$\lambda_0,\lambda_1,\lambda_2$ | $F_2$<br>$\lambda_0,\lambda_1,\lambda_2$ |
|---|---|---|
| $A_1$ | 0,1,4 | 0,4,1 |
| $A_2$ | 1,1,4 | 1,4,1 |
| $A_3$ | 0,1,1 | 0,1,1 |
| $A_4$ | 1,1,1 | 1,1,1 |

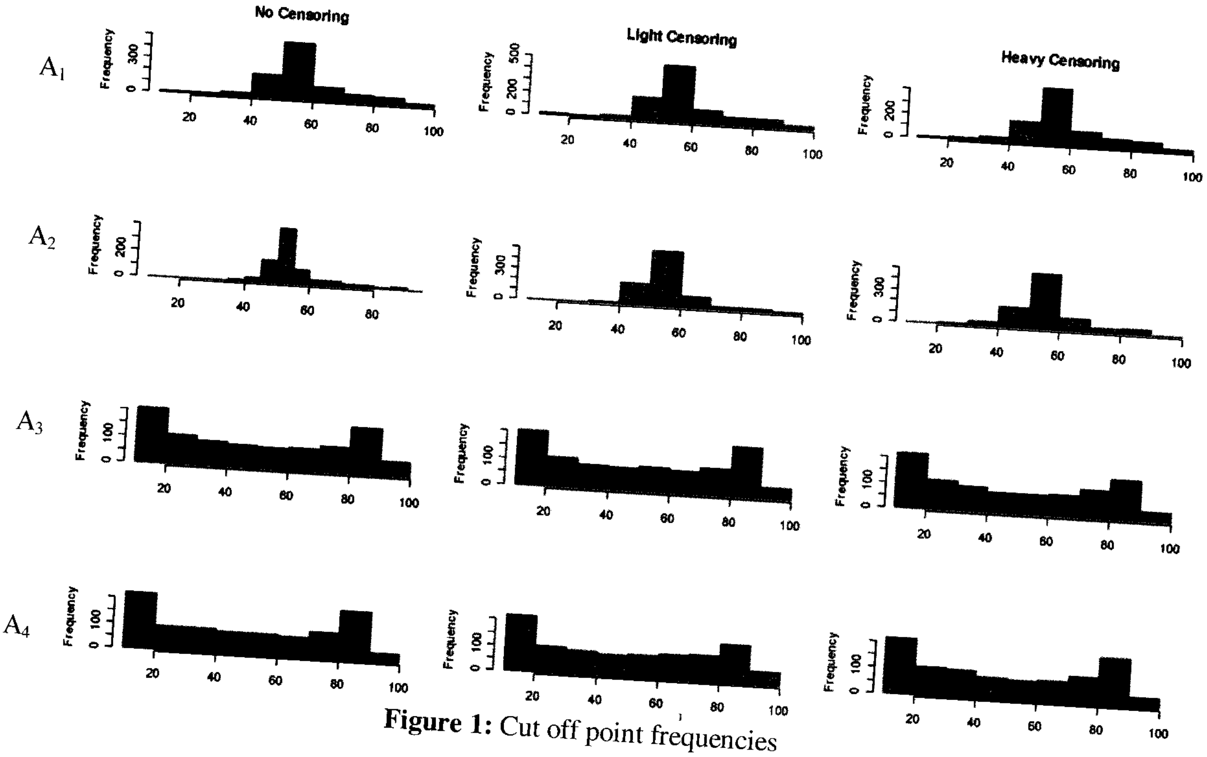

**Figure 1:** Cut off point frequencies

Simulation result is illustrated in Figure 1. We can see that scenarios $A_1$ and $A_2$ show desirable result in capturing the true cutoff point. The result is inline with our expectation, since the value of parameter of $\lambda_1$ and $\lambda_2$ are exchanged. However scenarios $A_3$ and $A_4$ have only slightly different value of $\lambda_0$ in which $F_1$ and $F_2$ is slightly indistinguishable. The two last rows of Figure 1 show that the method cannot capture the true cutoff point, resulting in a much lower and higher cutoff points. This shows the occurrence of end-cut-preference as studied by Torgo [6].

## Illustration

We used the data described in example I.3.1 in [1] which consist of $n$ = 205 observations for patients with malignant melanoma (cancer of the skin) who had a radical operation performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. Through this operation the tumor was completely removed together with the skin within a distance of about 2.5 cm around it. All patients were followed until the end of 1977, that is, it was noted if and when any of patients died. Note that the survival time was known only for those patients who died before the end of 1977. The rest of the patients were censored at the duration in the study obtained then. Two causes of failure were: (1) death from malignant melanoma; and (2) death from other causes. In example VII.2.5 of [1] it was demonstrated that tumor thickness only had an effect on failure of cause 1. It is interesting to find out a cutoff point in tumor thickness risk factor based on survival time of cause 1.

The data file can be obtained from http://www.pubhealth.ku.dk/~pka/MalignantMelanoma.dat and contains the variables id, time, thick, sex and cause. Here id identifies the patient, time is the failure time of the patient (days after radical operation), thick is the tumor thickness, sex is 0 for females, 1 for males and finally cause attains the value 1 if the patient died from malignant melanoma, 2 if the patient died of other causes and 0 if the patient was right censored. Rosthøj *et al.* [5] used these data for cumulative incidence function estimation using Cox regression model for competing risk data.

We can use the median tumor thickness as a simple cutoff point determination which is 1.94. In addition, if we use the information about survival time, then the outcome-oriented cutoff point determination can be used.

Next, we consider the lowess smoothed plot of the survival time of cause 1 (death from malignant melanoma) to determine a cutoff point for the patient's tumor thickness. This is an exploration tool for cutoff point determination and we only took into account for failure type 1 regardless of failure type 2 and censored data.

The display of both the smooth fit and the individual survival time of cause 1 provide insight into the influence of specific individuals on the estimate of the functional form. Figure 2 suggests that treating tumor thickness as linear is inappropriate. The smoothed curve is almost constant up to about 5 mm and decrease almost linearly up to about 10 mm. This suggests that patient's tumor thickness can be coded as an indicator variable in competing risk regression model.

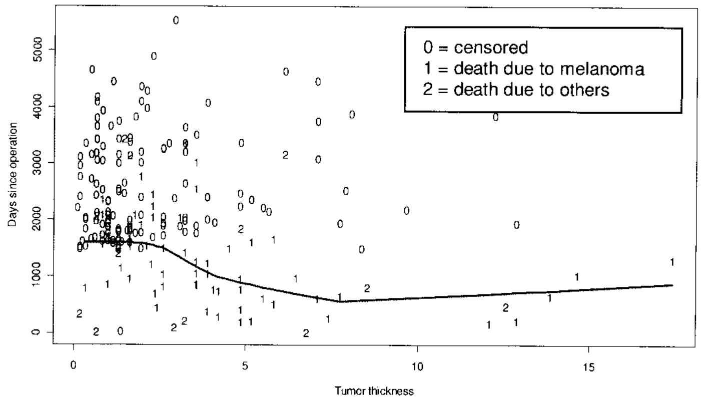

**Figure 2:** Plot of survival time of cause 1 (death from malignant melanoma) versus tumor thickness and lowess smooth

We now use the maximum Gray's statistic for dichotomizing patients into high or low risk groups for survival time based on the patient's tumor thickness and also assess the significance of the cutoff point. There were 64 distinct tumor thickness, any of which can be a potential cut point. The maximum value of Gray's statistic occured at tumor thickness cutoff point 2.17 with Gray's statistic 29.19 (see Figure 3). We used permutation test to evaluate the significance of this maximum Gray's statistic. Figure 3 (right panel) shows that the $p$-value $< 0.001$. This suggests that the cutoff point is significant and that tumor thickness is related to survival time of cause 1.

After we have found the cutoff point, we separated the data into two groups and further analysis was employed. We estimated the cumulative incidence function for both groups. We can see that the two groups resulted by dichotomizing at cutoff point 2.17 is better separated than the others. The cumulative incidence functions of both groups are quite different (see Figure 4). The thicker tumor group (thickness > 2.17) has the higher risk in death from malignant melanoma compared to the thinner ones (thickness < 2.17).

In the case of survival time for cause 2 which only consists of 14 distinct thickness values, we cannot find a good cutoff point due to the absence of potential cutoff points. The highest p-value (0.0786) suggests that the cutoff point obtained is not significant and tumor thickness is not related to survival time of other causes.

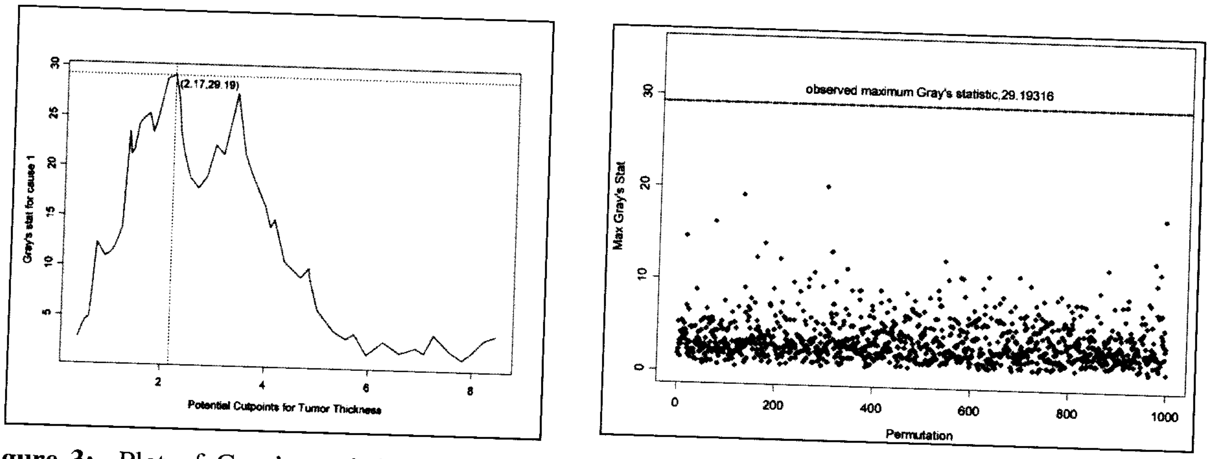

**Figure 3:** Plot of Gray's statistic versus distinct tumor thickness potential cutoff points (left) and permutation test for its maximum value.

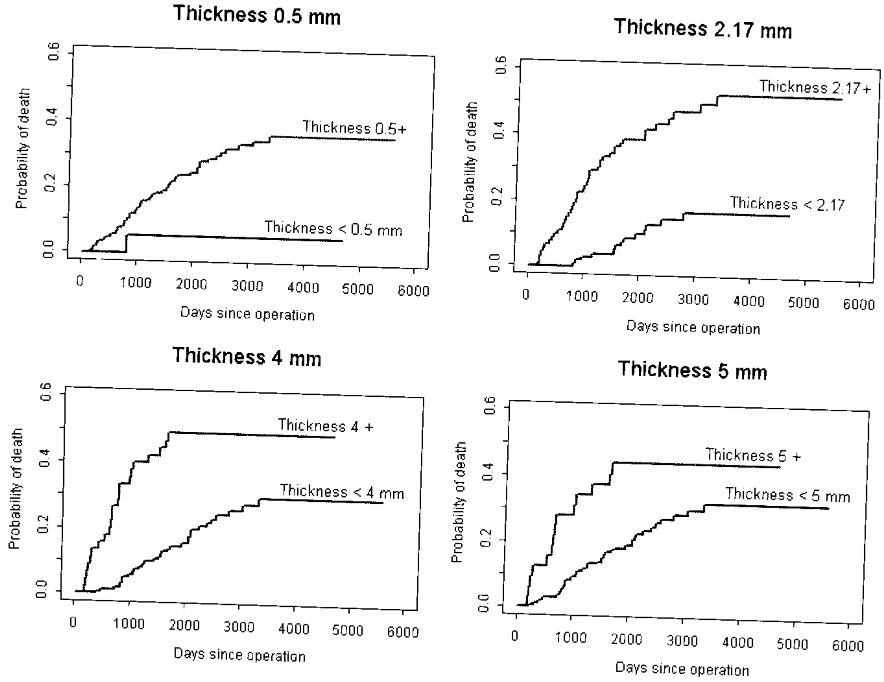

**Figure 4:** The estimated cumulative incidence function of malignant melanoma for two groups of patients.

## Some Limitations

We have only focused on the dichotomization of a continuous covariate with the assumption that such a dichotomization is possible from biological point of view, however, in reality, more than one cutoff point may exist. The obtained cutoff point(s) may differ across studies depending on several factors including which data are used and therefore the results may not be comparable. Lastly, there is always the possibility of loss in information from dichotomizing a continuous covariate, possible loss of power to detect actual significance that can sometimes lead to biased estimates in regression settings.

## Conclusion

We have proposed an approach for dichotomization of independent continuous variable in competing risk data analysis based on Gray's two-sample statistic. We have shown that this method has the ability to capture the true cutoff point as studied through simulation. Finally we provided an application of our method to separate malignant melanoma patients based on their tumor thickness. The study of cutoff point methodology is very important in the health science field. Implications for preventive medical attention are obvious. Simple, but yet accurate, guidelines for physician and other practitioners is provides for easy implementation.

## References

[1] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. 1993. Statistical Models Based on Counting Processes. Springer, New York.

[2] Gray, R.J. 1988. A class of *K*-sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*, **16**, 1141-1154.

[3] Ibrahim, N. A., Kudus, A., Daud, I., and Abu Bakar, M. R. 2009. Cutpoint Determination Methods in Competing Risks Subdistribution Model. *Journal of Quality Measurement and Analysis*. **5**, 103-117.

[4] Mandrekar, J. N., Mandrekar, S. J. and Cha, S. S. 2003. Cutpoint Determination Methods in Survival Analysis Using SAS®. *SUGI 28 Proceedings*. http://www2.sas.com/proceedings/sugi28/261-28.pdf

[5] Rosthøj, S., Andersen, P. K. and Abildstrom, S. Z. 2004. SAS Macros for Estimation of the Cumulative Incidence Functions Based on a Cox Regression Model for Competing Risks Survival Data. *Computer Methods and Programs in Biomedicine,* **74**, 69-75.

[6] Torgo, L. 2001. A Study on End-Cut Preference in Least Squares Regression Trees. Proceedings of the 10th Portuguese Conference on Artificial Intelligence. *Lecture Notes in Computer Science*, **2258**, 104-115. Springer-Verlag: London.