UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Facoltà di Medicina e Chirurgia

Scuola di Dottorato in Scienze Mediche Sperimentali e Cliniche

Dottorato di Ricerca in Epidemiologia e Biostatistica

XXV ciclo

Cut-point finding methods for continuous biomarkers

Tutor: Prof.ssa Maria Grazia Valsecchi

Tesi di Dottorato di:

Matteo Rota

Matricola 062958

Anno Accademico 2011-2012

To my family who supported me over these years

To all the people who helped me during my PhD program

ABSTRACT

My PhD dissertation deals with statistical methods for cut-point finding for continuous biomarkers. Categorization is often needed for clinical decision making when dealing with diagnostic (or prognostic) biomarkers and a dichotomous or censored failure time outcome. This allows the definition of two or more prognostic risk groups, or also patient's stratifications for inclusion in randomized clinical trials (RCTs).

We investigate the following cut-point finding methods: minimum P-value, Youden index, concordance probability and point closest to-(0,1) corner in the ROC plane. We compare them by assuming both Normal and Gamma biomarker distributions, showing whether they lead to the identification of the same true cut-point and further investigating their performance by simulation. Within the framework of censored survival data, we will consider here new estimation approaches of the optimal cut-point, which use a conditional weighting method to estimate the true positive and false positive fractions. Motivating examples on real datasets are discussed within the dissertation for both the dichotomous and censored failure time outcome.

In all simulation scenarios, the point closest-to-(0,1) corner in the ROC plane and concordance probability approaches outperformed the other methods. Both these methods showed good performance in the estimation of the optimal cut-point of a biomarker. However, to improve results communicability, the Youden index or the concordance probability associated to the estimated cut-point could be reported to summarize the

associated classification accuracy. The use of the minimum P-value approach for cut-point finding is not recommended because its objective function is computed under the null hypothesis of absence of association between the true disease status and X. This is in contrast with the presence of some discrimination potential of the biomarker X that leads to the dichotomization issue.

The investigated cut-point finding methods are based on measures, i.e. sensitivity and specificity, defined conditionally on the outcome. My PhD dissertation opens the question on whether these methods could be applied starting from predictive values, that typically represent the most useful information for clinical decisions on treatments. However, while sensitivity and specificity are invariant to disease prevalence, predictive values vary across populations with different disease prevalence. This is an important drawback of the use of predictive values for cut-point finding.

More in general, great care should be taken when establishing a biomarker cut-point for clinical use. Methods for categorizing new biomarkers are often essential in clinical decision-making even if categorization of a continuous biomarker is gained at a considerable loss of power and information. In the future, new methods involving the study of the functional form between the biomarker and the outcome through regression techniques such as fractional polynomials or spline functions should be considered to alternatively define cut-points for clinical use. Moreover, in spite of the aforementioned drawback related to the use of predictive values, we also think that additional new methods for cut-point finding should be developed starting from predictive values.

CONTENTS

CHAPTER 1

INTRODUCTION 1

1.1 BIOMARKERS IN CLINICAL RESEARCH

A panel of experts of the National Institute of Health (NIH) defined a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention” [1]. The word “biomarker” covers a broad range of biochemical entities, such as nucleic acids and proteins, as well as cell lines or biophysical characteristics of tissues.

The biomarker revolution started about 20 years ago and it is still growing [2]. Advances in experimental sciences have led scientists to discover a huge number of biomarkers, but only some of them have been validated and currently used in the standard medical practice. Exploring the relationship between new biomarkers and health outcomes can have potentially strong effects on the biomedical community, leading to new etiological discoveries, as well as increased diagnostic capabilities [2]. Biomarkers can be broadly classified into three categories: *screening* (hereafter also defined as diagnostic), *prognostic* and *predictive* [3].

Screening (or *diagnostic*) *biomarkers* play an important role in cancer research [4], allowing to detect cancer at early stages, and helping to increase the chances of care. Classical examples of screening biomarkers are the prostate-specific antigen (PSA) test for

the early detection of prostate cancer, or the fecal occult blood test (FOBT) for the early detection of colorectal cancer.

Prognostic biomarkers identify patients with different risks of having a specific outcome, regardless of the type of treatment (Figure 1.1 Panel (a)). From Figure 1.1 Panel (a), we can see that the magnitude of the effect of a new treatment T versus a standard treatment S on the clinical response is not affected by the presence of a positive (or negative) value of the biomarker. Prognostic biomarkers have utility both in the management of patients, and within the early stages of pharmaceutical development, such as target discovery or target development [5], and also to define patients' stratifications for inclusion in randomized clinical trials (RCTs). As an example, the Amyloid β peptide (AB) 1-42 is a prognostic biomarker in patients diagnosed with amnestic Mild Cognitive Impairment (aMCI), predicting progression to Alzheimer's disease [6].

Predictive biomarkers identify patients likely to have a favorable outcome from a specific treatment (Figure 1.1 Panel (b)). From Figure 1.1 Panel (b), we can note that the magnitude of the effect of a new treatment T versus a standard treatment S on the clinical response depends upon the presence of a positive (or negative) value of the biomarker. For example, P450s enzymes - expressed at higher levels in the tumor cells than in the surrounding normal tissues - offer therapeutic options through the activation of prodrugs specifically in cancer cells, and avoiding undesirable systemic effects [7].

After this short introduction aimed to overview the basic concepts of biomarkers, we focus our attention to prognostic modeling from the biostatistics perspective in order to introduce the objective of this PhD dissertation.

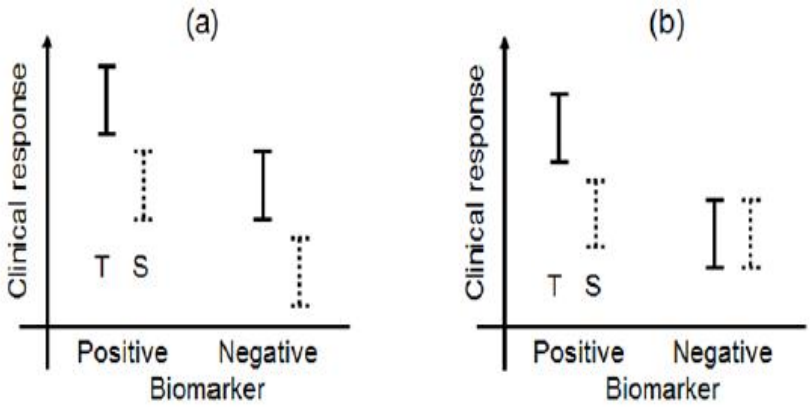

Figure 1.1 Panel (a): Prognostic biomarker. **Panel (b):** Predictive biomarker. T and S denote, respectively, test and standard treatment [3].

1.2 PROGNOSTIC MODELING

Prognosis simply means foreseeing, predicting, or estimating the probability of risk of future conditions [8]. Prognosis is central to medicine, and it is often used to direct diagnostic pathways, to inform patients and to predict outcome. The outcome can be identified as a specific event such as death, relapse or disease progression. However, prognostication in medicine is not limited to those who are ill. For example, the Apgar score is aimed to assess in a simply way the health of newborn children immediately after birth, or also cardiovascular risk profiles are aimed to predict heart diseases in general population.

The biostatistician contribution is of primary importance in the conduction of prognostic researches [9]. A prognostic model is a statistical tool aimed to investigate the probability of development of a considered outcome in a given population composed by subjects with different demographic and clinical characteristics. We are in the era of the "Evidence-Based Medicine" (EBM) [10], defined as "the conscientious, explicit and judicious

use of current best evidence in making decisions about the care of individual patients". As a consequence, more and more emphasis is placed on prognostic modeling. We could simply detect through a PubMed search an exponential growth of the number of articles published after 1970 dealing with prognostic modeling (Figure 1.2) [9].

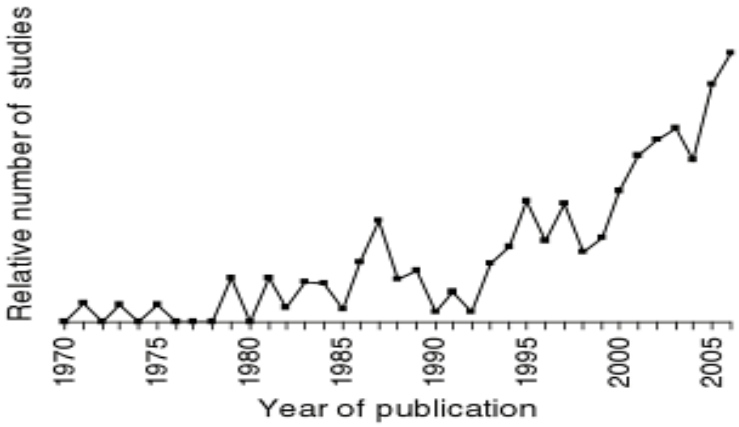

Figure 1.2 Studies in PubMed with the term "prognostic model" in the title, from [9].

The biostatistician expertise in this area ranges widely, from study design to statistical analysis of data [11]. Within this macro area, one main challenge is the definition of threshold values, hereafter defined as *cut-points*, for new biomarkers or prognostic indexes.

In the absence of an a priori clinical consensus, often the commonly used method for cut-point finding is to use a non-data (outcome) driven method [12], i.e. splitting the population into equal sizes groups based on the quartiles of the biomarker distribution. A recent review of 47 studies [12] found that risk groups were developed from prognostic models in 76% of the included studies, and 25% of these studies involved non-data (outcome) driven methods for cut-point finding.

We are dealing with a continuous prognostic factor X and we want to define a data-driven method for determining an optimal cut-point for X. This is essential for the

clinical decision-making process, allowing the definition of two or more prognostic risk groups, or also patient's stratifications for inclusion in randomized clinical trials (RCTs).

1.3 RATIONALE

This PhD dissertation is aimed to compare methods for cut-point finding in the presence of a dichotomous outcome (i.e., diseased/non-diseased or relapse/non-relapse and so on), and to define new methods for cut-point finding in a framework of possibly censored failure time data.

Once a continuous variable has been identified as being related to the outcome by showing some association with the outcome in a regression framework, a first explorative graphical analysis is needed to examine the form of the relationship between the two variables. This can detect the presence of a potential cut-point or provide a range of values within of which the search should be restricted (Figure 1.3).

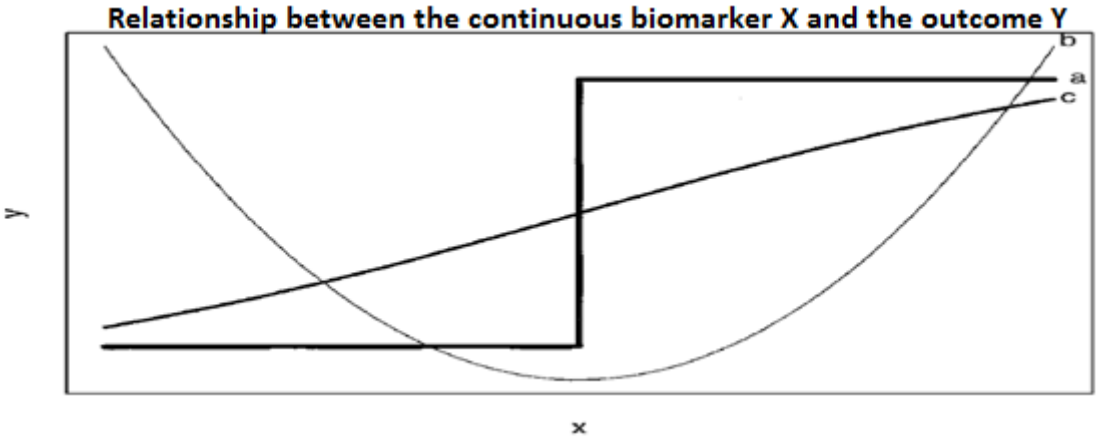

Figure 1.3 Line (a): step function model with a single cut-point. **Line (b)**: if the underlying relationship is non-monotonic, a single cut-point is not plausible. **Line (c)**: no cut-point is apparent, and reducing X to a dichotomous variable could result in a loss of information.

Figure 1.3 lines (a) and (c) show monotonic functions; the former function represents an ideal situation where a cut-point is suggested, while the latter describes a common situation where a cut-point model does not provide the best fit. Figure 1.3 line (b) shows a non-monotonic function for which a single cut-point is not plausible.

Whether or not the graphical examination suggests an ideal cut-point, a data-driven (outcome) method that allows to systematically search a cut-point is needed to define risk groups for the clinical decision making process.

The thesis is mainly focused on the comparison of the performance of cut-point finding methods both in the framework of dichotomous and right censored survival outcome. Within the framework of dichotomous outcome, several methods are indistinctly used by researcher aimed to find cut-points: minimum P-value [13], Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane approach [18]. So far these methods have not been compared theoretically in order to establish which one has the best performance, and thus there is a lack of evidence on which method should be preferred within the biostatistics community. Moreover, within the framework of censored survival data, we will consider here new estimation approaches of the optimal cut-point, which use a conditional weighting method to estimate the true positive and false positive fractions [17]. Motivating examples are discussed thorough the dissertation.

The thesis is structured as follows. Chapter 2 describes the methodological work along with clinical applications related to the cut-point finding task for a dichotomous outcome, while Chapter 3 describes the analogous scenario and an application in the context of censored failure time data. Chapter 4 deals with conclusions and future research perspectives. Last, ad hoc software developed in this research is reported in appendices.

CHAPTER 2

CUT-POINT FINDING FOR A DICHOTOMOUS OUTCOME 2

2.1 THE RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE

In the presence of a continuous response test or a biomarker X, it is often necessary to identify a cut-point value to classify subjects as testing positive, considered at higher risk of being diseased, from those testing negative, considered at lower risk of being diseased (Figure 2.1).

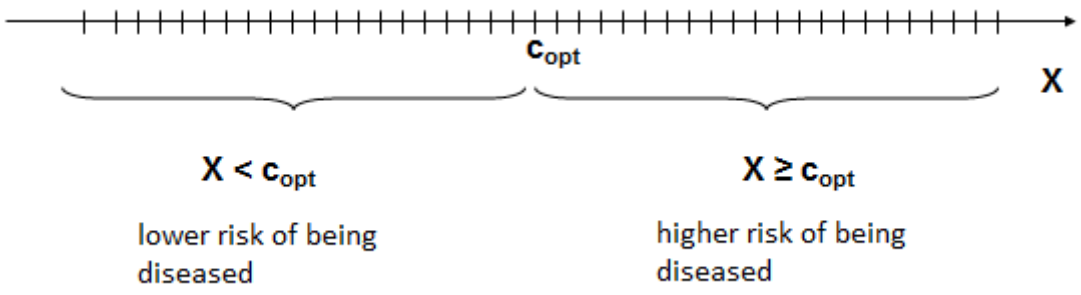

Figure 2.1 The optimal cut-point value c_{opt} for a biomarker X on a continuous scale.

Let X denote a continuous biomarker which is supposed to be related to the binary outcome (true disease status), where D stands for the presence of disease and $\overline{D}$ for the absence of disease.

The **true positive fraction** TPF(c) and the **false positive fraction** FPF(c) are respectively defined, at any given possible cut-point c of X, as:

$$TPF(c) = P(X > c \mid D) = S_D(c) \tag{2.1}$$

$$FPF(c) = P(X > c \mid \bar{D}) = S_{\bar{D}}(c) \tag{2.2}$$

In this setting, the **receiver operating characteristic curve** (ROC) is often the starting point for determining the optimal cut-point c_{opt}. The ROC curve is a plot of the TPF(c) (2.1) - or sensitivity (probability of correctly classifying diseased subjects) - and the FPF(c) (2.2) - or 1-specificity (probability of incorrectly classifying non–diseased subjects) - for all possible cut-point values of X (Figure 2.2) [16].

The ROC curve is a monotone increasing function in the positive quadrant, lying between the corners (FPF(c)=0, TPF(c)=0) and (FPF(c)=1, TPF(c)=1), that could be mathematically expressed as follow:

$$ROC(.) = \{FPF(c), TPF(c), c \in (-\infty, +\infty)\} \tag{2.3}$$

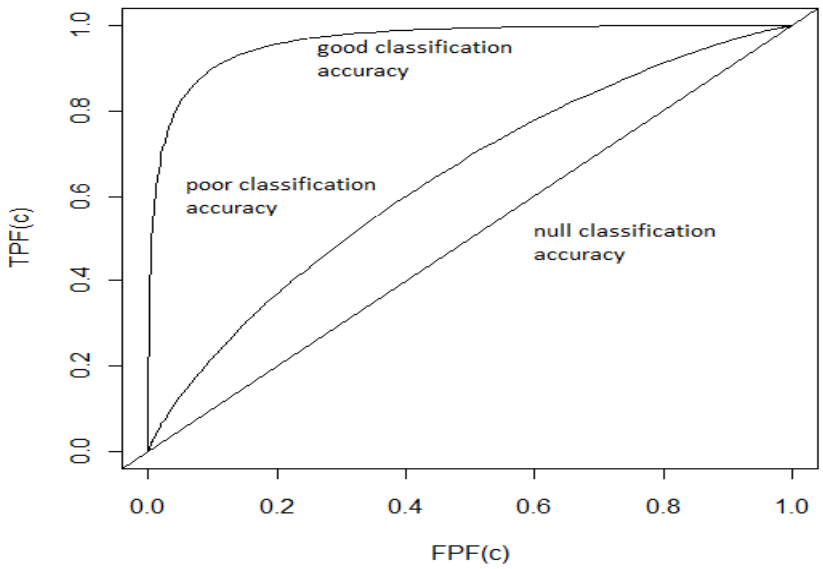

Figure 2.2 Examples of ROC curves with different classification accuracies.

As the candidate cut-point c of X increases, both TPF(c) and FPF(c) decrease. In particular, $\lim_{c\to+\infty} \mathrm{TPF}(c) = 0$ and $\lim_{c\to+\infty} \mathrm{FPF}(c) = 0$. Conversely, as the candidate cut-point c of X decreases, $\lim_{c\to-\infty} \mathrm{TPF}(c) = 1$ and $\lim_{c\to-\infty} \mathrm{FPF}(c) = 1$.

As you can see from Figure 2.2, the ROC curve of a biomarker with a good classification accuracy lyes near the left and upper border of the positive quadrant; conversely, the bisecting line represents a ROC curve of a uselss biomarker, i.e. a biomarker having TPF(c)=FPF(c) at any c of X, and thus a classification accuracy not better than a random guess. Better biomarkers have ROC curves closer to the upper left corner.

The area under the ROC curve (AUC) is the most widely used summary measure of classification accuracy, and it is mathematically defined as follows: $\mathrm{AUC} = \int_0^1 \mathrm{ROC}(t)dt$. So, the perfect biomarker has the value AUC=1.0, while the useless biomarker has AUC=0.5. The AUC is also equal to the probability that the biomarker results from a randomly selected pair of diseased and non-diseased are correctly ordered, namely $P(X|D > X|\overline{D})$.

This paragraph introduces the fundamental concepts of the ROC curve theory needed to start the core of this PhD dissertation.

2.2 METHODS BACKGROUND AND AIMS

Two methods based on the ROC curve are commonly used to determine the optimal cut-point for X: the Youden index [14] and the point closest-to-(0,1) corner in the ROC plane approach [18]. The Youden index method selects the optimal cut-point that maximizes the Youden function, defined as the difference between TPF(c) (2.1) and FPF(c) (2.2) over all possible cut-point values of X [14]. The point closest-to-(0,1) corner in the ROC plane

approach selects the optimal cut-point for X as the one minimizing the Euclidean distance between some point on the ROC curve and the (0,1) point, that represents the ideal situation with maximum sensitivity (TPF=1) and specificity (1-FPF=0) [18]. Perkins and Schisterman [18] showed that these ROC-based methods may not lead to the same optimal cut-point, and they advocated for the use of the Youden index method [14]. In fact, this latter is a commonly used measure of overall classification accuracy, while the point closest-to-(0,1) corner in the ROC plane approach involves the minimization of a quadratic term that does not own a clinical meaning per se.

A third approach to determine the optimal cut-point for X relies on the maximization of the diagnostic odds ratio (DOR) function [19], defined as the ratio between the odds of TPF and FPF over all possible cut-point values of X. Böhning et al. [20] have recently showed that the DOR strategy is no longer recommended since it might easily lead to the choice of cut-point values on the boundary of the parameter range of X. In fact, when considered within a parametric homoscedastic Normal distribution scenario, the log-DOR is a convex function (Figure 2.3). Conversely, this deficient behaviour is not present for the Youden index method, which should be preferred to the DOR approach [20].

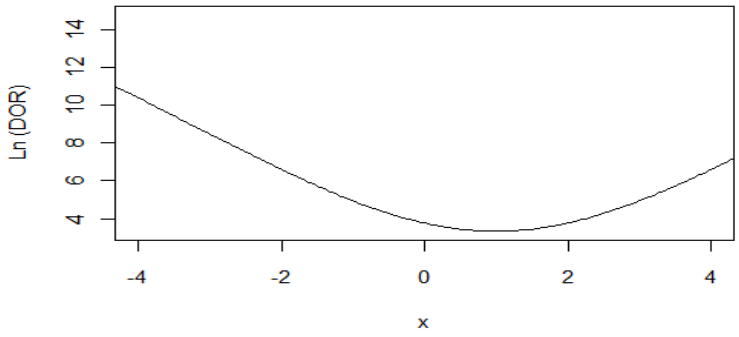

Figure 2.3 The DOR function is convex if considered in a Normal distribution scenario [20].

A more recent simulation study [15] compared the performance of the Youden index method and the point closest-to-(0,1) corner in the ROC plane approach with a new concordance probability criterion based on the choice of the optimal cut-point that achieves the maximum of the product of sensitivity (TPF) and specificity (1-FPF). This latter method performed well within the proposed scenarios of Normal and Gamma biomarker distributions [15].

Indeed, within the class of ROC-based methods, the Youden index seems the most commonly used and sound intuitive criterion for cut-point finding [18, 20]. The recently proposed concordance probability [15] has a meaningful interpretation as well. It corresponds to the probability of being below or beyond the cut-point for any random pair of non-diseased and diseased subjects, respectively.

However, none have so far compared the performance of the Youden index, the concordance probability and the point closest-to-(0,1) corner in the ROC plane methods with the minimum P-value approach [13]. This latter is based on the exploration of all observed values of X to identify the optimal cut-point that best separates the two risk groups (i.e., diseased and non-diseased) according to the maximum achievable value of the Chi-square statistic on the association between X and the binary outcome variable (true disease status).

We aimed to compare theoretically the minimum P-value approach, the Youden index method, the concordance probability criterion and the point closest-to-(0,1) corner in the ROC plane approach in order to investigate whether they lead or not to the identification of the same optimal cut-point through a theoretical study aimed to investigate the relationship between these methods. Moreover, we further investigate the performance of these methods by a simulation study.

2.3 METHODS FOR CUT-POINT FINDING FOR A DICHOTOMOUS OUTCOME

We review theoretically the minimum P-value approach [13], the Youden index method [14], the concordance probability criterion [15] and the point closest-to-(0,1) corner in the ROC plane approach [18].

Let us consider a sample of n_D diseased subjects and $n_{\bar{D}}$ non-diseased subjects, where $N = n_D + n_{\bar{D}}$. For any c of X, we can define the following classification matrix:

	$X \le c$	$X > c$	
$\bar{D}$	n_{11}	n_{12}	$n_{\bar{D}}$
D	n_{21}	n_{22}	n_D
			N

(2.4)

The sample estimates of the classification probabilities (2.1) and (2.2) can be obtained from matrix (2.4) at a given c of X as $\hat{S}_D(c) = n_{22} / n_D$ and $\hat{S}_{\bar{D}}(c) = n_{12} / n_{\bar{D}}$.

2.3.1 MINIMUM P-VALUE APPROACH

The minimum P-value approach [13] is based on a systematic search of the optimal cut-point that achieves the minimum of the P-value of the Chi-square test statistic on the absence of association between the resulting dichotomized biomarker and the binary outcome, or, in other words, the maximum of the associated Chi-square statistic on matrix (2.4) over all possible cut-point values c of X.

Thus, according to this method, the optimal cut-point $\hat{c}_{\chi^2}$ is the c of X maximizing the one degree of freedom Chi-square statistic computed from matrix (2.4) as:

$$\chi_1^2(c) = \frac{N(n_{11}n_{22} - n_{12}n_{21})^2}{n_{\bar{D}}n_D(n_{11}+n_{21})(n_{12}+n_{22})} \tag{2.5}$$

In practice, the search of the optimal cut-point $\hat{c}_{\chi^2}$ excludes the boundary of X, where either $n_{11}+n_{21}=0$ or $n_{12}+n_{22}=0$, in order to avoid an undefined Chi-square statistic (2.5).

Yet the Chi-square statistic (2.5) can be expressed as a population quantity starting from its typical expression:

$$CHI_1^2(c) = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(O_{ij}(c) - E_{ij}(c))^2}{E_{ij}(c)} \tag{2.6}$$

by writing the observed classification matrix (2.4) in terms of classification probabilities (2.1) and (2.2).

Here, $O_{ij}(c)$ and $E_{ij}(c)$ from (2.6) are random variables representing, respectively, the observed and expected population counts in a classification matrix of the type (2.4) (where i, j=1,2 are the row and column indexes), for a sampling design with n_D diseased and $n_{\bar{D}}$ non-diseased subjects.

Let us consider the counts $O_{ij}(c)$ from (2.6). The observed number of non-diseased subjects presenting X≤c (i.e., n_{11} in matrix (2.4)) is represented by the random count $n_{\bar{D}}P(X \le c \mid \bar{D})$, or from (2.2), $n_{\bar{D}}(1-S_{\bar{D}}(c))$. A similar argument follows also for the other random counts $O_{ij}(c)$ from (2.6), leading to the following observed classification matrix:

	X ≤ c	X > c	
$\overline{D}$	$n_{\overline{D}}(1-S_{\overline{D}}(c))$	$n_{\overline{D}}S_{\overline{D}}(c)$	$n_{\overline{D}}$
D	$n_D(1-S_D(c))$	$n_D S_D(c)$	n_D
			N

In order to derive the expected counts $E_{ij}(c)$ from (2.6), we need to write the expected classification matrix under the assumption of absence of association between the resulting dichotomized biomarker X and the binary outcome. Under this assumption, the expected count of non-diseased subjects presenting X≤c is $n_{\overline{D}}P(X \le c)$, where P(X≤c) is the complement to one of the positive fraction (PF), that can be written as a weighted average of $S_D(c)$ and $S_{\overline{D}}(c)$ as:

$$PF(c) = P(X > c) = \frac{n_D S_D(c) + n_{\overline{D}} S_{\overline{D}}(c)}{n_D + n_{\overline{D}}}$$

By arguing similarly, we obtain the following expected classification matrix:

	X ≤ c	X > c	
$\overline{D}$	$n_{\overline{D}} \frac{n_D + n_{\overline{D}} - n_D S_D(c) - n_{\overline{D}} S_{\overline{D}}(c)}{n_D + n_{\overline{D}}}$	$n_{\overline{D}} \frac{n_D S_D(c) + n_{\overline{D}} S_{\overline{D}}(c)}{n_D + n_{\overline{D}}}$	$n_{\overline{D}}$
D	$n_D \frac{n_D + n_{\overline{D}} - n_D S_D(c) - n_{\overline{D}} S_{\overline{D}}(c)}{n_D + n_{\overline{D}}}$	$n_D \frac{n_D S_D(c) + n_{\overline{D}} S_{\overline{D}}(c)}{n_D + n_{\overline{D}}}$	n_D
			N

Now, considering the single term $\frac{(O_{ij}(c) - E_{ij}(c))^2}{E_{ij}(c)}$ for i, j=1,2 from (2.6), and using a little algebra, we obtain:

$$\frac{(O_{11}(c) - E_{11}(c))^2}{E_{11}(c)} = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} \frac{n_D}{n_D + n_{\bar{D}} - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)}$$

$$\frac{(O_{12}(c) - E_{12}(c))^2}{E_{12}(c)} = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} \frac{n_D}{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}$$

$$\frac{(O_{21}(c) - E_{21}(c))^2}{E_{21}(c)} = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} \frac{n_{\bar{D}}}{n_D + n_{\bar{D}} - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)}$$

$$\frac{(O_{22}(c) - E_{22}(c))^2}{E_{22}(c)} = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} \frac{n_{\bar{D}}}{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}$$

Since $\frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}}$ is common to all the terms, we can write:

$$CHI_1^2(c) = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij}(c) - E_{ij}(c))^2}{E_{ij}(c)} = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} g(c)$$

where:

$$g(c) = \frac{n_D + n_{\bar{D}}}{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)} + \frac{n_D + n_{\bar{D}}}{n_D + n_{\bar{D}} - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)}$$

Thus, we obtain the Chi-square function in population as:

$$\mathrm{CHI}_1^2(c) = \frac{(S_D(c) - S_{\bar{D}}(c))^2}{\frac{1}{n_D} + \frac{1}{n_{\bar{D}}}} \left(\frac{n_D + n_{\bar{D}}}{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)} + \frac{n_D + n_{\bar{D}}}{n_D + n_{\bar{D}} - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)} \right)$$

and after some little algebra, we obtain:

$$\mathrm{CHI}_1^2(c) = \frac{\left(S_D(c) - S_{\bar{D}}(c)\right)^2}{\left(\frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}} \right) \left(1 - \frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}} \right) \left(\frac{1}{n_D} + \frac{1}{n_{\bar{D}}} \right)} \qquad (2.7)$$

Since we carried out a huge number of hypothesis tests, one for each potential cut-point value c of X, the probability of obtaining a significant result is inflated. The minimum P-value P_{min} associated to the optimal cut-point $\hat{c}_{\chi^2}$ determined through this method should be corrected by using the formulae developed by Miller and Siegmund [13]:

$$P_{ms} = \varphi(z)\left(z - \frac{1}{z}\right)\log\frac{\varepsilon_{high}(1-\varepsilon_{low})}{\varepsilon_{low}(1-\varepsilon_{high})} + 4\frac{\varphi(z)}{z} \qquad (2.8)$$

where z is the $(1 - p_{min}/2)$ standard Normal percentile, while ε_{low} and ε_{high} represents the sample proportions of cut-point values less than or greater than the smallest and largest cut-point considered in the analysis, respectively. It is worthwhile to state that the standard Bonferroni correction method is not appropriate in this setting since it does not hold the independence assumption between the $\chi_1^2(c)$ test statistics.

2.3.2 YOUDEN INDEX METHOD

The Youden index (J) [14] is the maximum achievable value of the Youden function J(c), defined as the difference between the population quantities TPF(c) (2.1) and the FPF(c) (2.2):

$$J(c) = S_D(c) - S_{\bar{D}}(c) \tag{2.9}$$

J(c) ranges between 0 if $S_D(c) = S_{\bar{D}}(c)$, and 1 in the ideal case where $S_D(c) = 1$ and $S_{\bar{D}}(c) = 0$. J could also be seen as the maximum vertical distance between the ROC curve and the diagonal chance line representing a useless biomarker (Figure 2.4). It can be interpreted as the net gain of the true positive fraction with respect to the false positive fraction.

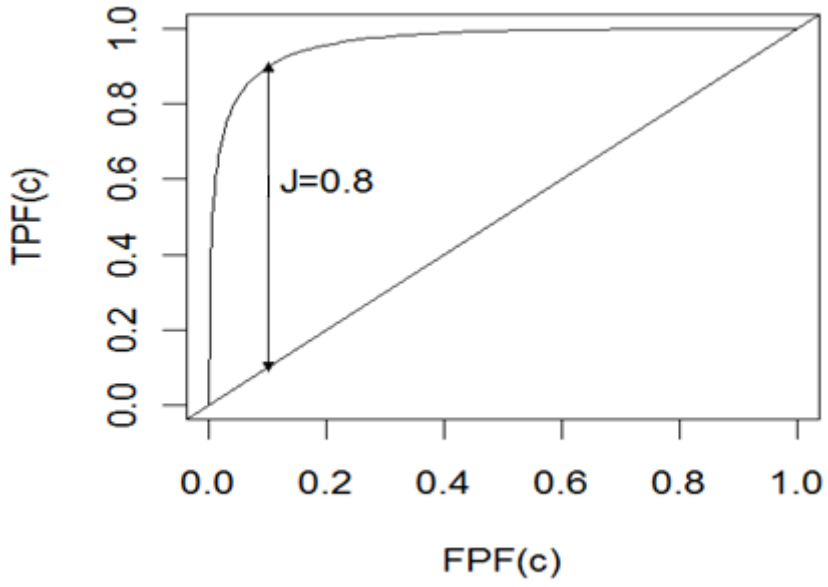

Figure 2.4 The graphical interpretation of the Youden index J.

Following this approach, the optimal cut-point $\hat{c}_J$ is the c that achieves the maximum of the Youden function $\hat{J}(c) = \hat{S}_D(c) - \hat{S}_{\bar{D}}(c)$ over all possible cut-point values c of X.

2.3.3 CONCORDANCE PROBABILITY METHOD

The concordance probability [15] objective function could be defined as the product of the population quantities TPF(c) (2.1) and the complement to one of the FPF(c) (2.2):

$$CZ(c) = S_D(c) \cdot (1 - S_{\bar{D}}(c)) \tag{2.10}$$

CZ(c) ranges between 0 if $S_D(c) = 0$ or $1 - S_{\bar{D}}(c) = 0$, and 1 in the ideal case where $S_D(c) = 1$ and $1 - S_{\bar{D}}(c) = 1$, i.e. $S_{\bar{D}}(c) = 0$. CZ(c) could be also expressed as the area of a rectangle on the ROC curve of width $1 - S_{\bar{D}}(c)$ and height $S_D(c)$ varying c (Figure 2.5, area of the dotted rectangle) and interpreted as the probability of being below or beyond the cut-point for any random pair of non-diseased and diseased subjects. CZ(c) evaluates the classification accuracy of the dichotomous biomarker at cut-point c.

Following this approach, the optimal cut-point $\hat{c}_{CZ}$ is the c that achieves the maximum of the concordance probability function $\hat{CZ}(c) = \hat{S}_D(c) \cdot (1 - \hat{S}_{\bar{D}}(c))$ over all possible cut-point values c of X.

2.3.4 POINT CLOSEST-TO-(0,1) CORNER IN THE ROC PLANE APPROACH

The objective function of this ROC-based method [18] could be easily defined by applying the Euclidean distance formulae between the point on the ROC plane defined by the population quantities (2.1) and (2.2) and the point (0,1):

$$ER(c) = \sqrt{(S_D(c) - 1)^2 + (S_{\bar{D}}(c))^2} \tag{2.11}$$

According to this method, the optimal cut-point $\hat{c}_{ER}$ is the c that achieves the minimum of the objective function $\hat{ER}(c)=\sqrt{(\hat{S}_D(c)-1)^2+(\hat{S}_{\bar{D}}(c))^2}$ over all possible cut-point values c of X (Figure 2.5, length of the thin line segment).

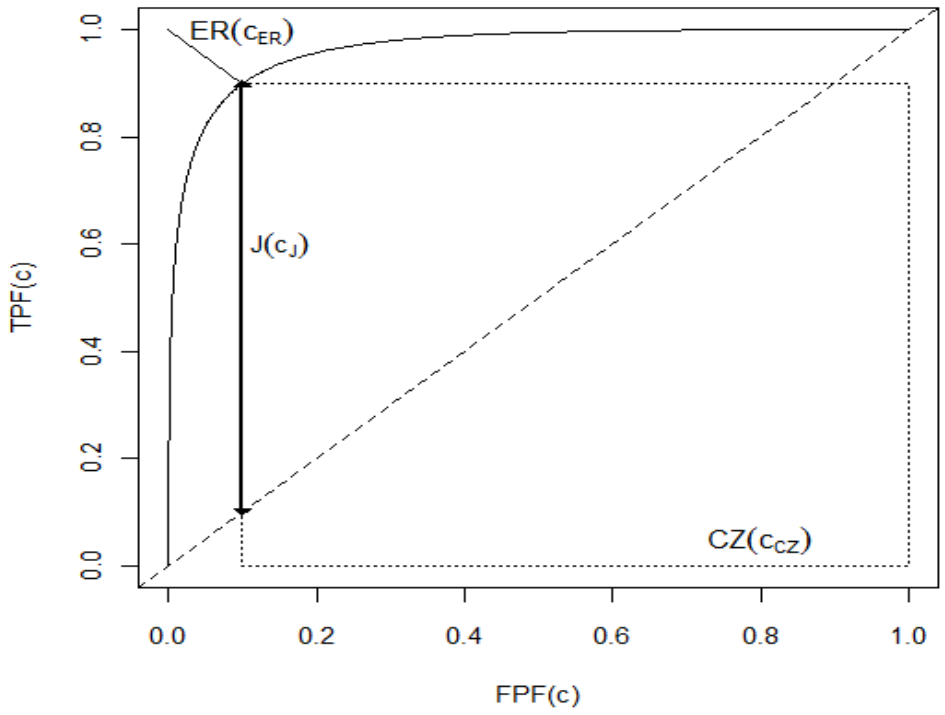

Figure 2.5 Comparison of ROC-based methods for cut-point finding: the Youden index J [14], the concordance probability CZ [15] and the distance from the ideal biomarker ER [18].

2.4 THEORETICAL CONSIDERATIONS ON THE RELATIONSHIP BETWEEN DIFFERENT APPROACHES

The minimum P-value approach and the Youden index method are mathematically related. In fact, from the population quantities (2.7) and (2.9), we can write

$$CHI_1^2(c) = J(c)^2 / \left[\left(\frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}} \right) \left(1 - \frac{n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)}{n_D + n_{\bar{D}}} \right) \left(\frac{1}{n_D} + \frac{1}{n_{\bar{D}}} \right) \right].$$

Thus, in other words, the Youden function (2.9) is the square root of the numerator of the Chi-square function (2.7).

However, despite this mathematical similarity, we show here within a parametric scenario that the associated cut-point finding methods do not necessarily identify the same true cut-point. We describe the special Normal homoscedastic case where the three considered ROC-based methods (2.9), (2.10) and (2.11) identify the same true cut-point c_{ROC}. Moreover, if the design is balanced ($n_D = n_{\bar{D}}$), this equality also holds for the minimum P-value approach (2.7), leading to a common cut-point c_{opt}. When $n_D < n_{\bar{D}}$, as often typical in clinical applications, the true cut-point c_{χ^2} underlying the minimum P-value approach (2.7) is greater than the one associated to the ROC-based methods, being $c_{\chi^2} > c_{ROC}$.

2.4.1 THE GAUSSIAN PARAMETRIC SCENARIO

Let us assume that X is normally distributed in the diseased and non-diseased populations, respectively as $X_D \sim N(\mu_D, \sigma_D = 1)$ and $X_{\bar{D}} \sim N(\mu_{\bar{D}} = 0, \sigma_{\bar{D}} = 1)$. This means

that $S_D(c) = 1 - \Phi(c - \mu_D)$ and $S_{\bar{D}}(c) = 1 - \Phi(c)$, where Φ denotes the standard Normal distribution function. If μ_D is set equal to {0.51, 1.05, 1.68, 2.56}, the corresponding maximum values of the Youden function (2.9) are $J(c_J)$={0.2, 0.4, 0.6, 0.8}, the corresponding maximum values of the concordance probability (2.10) are $CZ(c_{CZ})$={0.36, 0.49, 0.64, 0.81}, while the minimum values of the objective function (2.11) of the point closest-to-(0,1) corner in the ROC plane approach are $ER(c_{ER})$={0.57, 0.42, 0.28, 0.14}. This set of values of μ_D ensures a wide variety of classification accuracies, ranging from a poor one (J=0.2 and CZ=0.36) to a high one (J=0.8 and CZ=0.81) [21]. Conversely, the maximum values of the Chi-square function (2.7) of the minimum P-value approach depend on the size of the sample. For example, by considering $n_D = n_{\bar{D}}$=100, we obtain $CHI_1^2(c_{\chi^2})$={8, 32, 72, 128}, and considering n_D=50, $n_{\bar{D}}$=150, we obtain $CHI_1^2(c_{\chi^2})$={6, 26, 64, 121}.

Under the aforementioned Normal homoscedastic distribution scenario, we investigate whether the optimal cut-point c_J underlying the Youden index method (2.9) corresponds also to the true cut-points c_{χ^2}, c_{CZ} and c_{ER} of the minimum P-value (2.7), concordance probability (2.10) and point closest-to-(0,1) corner in the ROC plane (2.11) methods, respectively.

The distribution functions $F_D(c) = 1 - S_D(c)$ of X_D and $F_{\bar{D}}(c) = 1 - S_{\bar{D}}(c)$ of $X_{\bar{D}}$ are indeed absolutely continuous. The cut-point optimizing (2.9) can be obtained by studying the first derivative of J(c), $\frac{\partial J(c)}{\partial c} = -f_D(c) + f_{\bar{D}}(c)$, where $f_D(c) = \partial F_D(c)/\partial c$ and

$f_{\bar{D}}(c) = \partial F_{\bar{D}}(c)/\partial c$ are the Normal probability density functions of diseased and non-diseased subjects, respectively. By a little algebra, it follows that $c_J = \mu_D/2$ is a root of $\frac{\partial J(c)}{\partial c} = 0$.

Moreover, it can be proven in case of a balanced design ($n_D = n_{\bar{D}}$) that c_J is a maximum point of the Chi-square function (2.7), since the first derivative of $CHI_1^2(c)$ vanishes at $c = c_J$, arguing as follows.

Let us consider the sign of the first derivative $\frac{\partial CHI_1^2(c)}{\partial c}$ at the maximum point c_J of the Youden function (2.9). Since the objective function (2.7) is a ratio, we just need to study the numerator of $\frac{\partial CHI_1^2(c)}{\partial c}$, that by some algebra can be written as:

$$\begin{aligned}&\left(S_D(c) - S_{\bar{D}}(c)\right)\cdot\\&\left[\begin{aligned}&2\left(-f_D(c) + f_{\bar{D}}(c)\right)\left(n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)\right)\left(n_D F_D(c) + n_{\bar{D}} F_{\bar{D}}(c)\right)\\&+\left(S_D(c) - S_{\bar{D}}(c)\right)\left(n_D f_D(c) + n_{\bar{D}} f_{\bar{D}}(c)\right)\left(n_D F_D(c) + n_{\bar{D}} F_{\bar{D}}(c) - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)\right)\end{aligned}\right]\end{aligned} \quad (2.12)$$

We showed above that $f_D(c_J) = f_{\bar{D}}(c_J)$, since $\frac{\partial J(c)}{\partial c}|_{c=c_J} = -f_D(c_J) + f_{\bar{D}}(c_J) = 0$. Consequently, the term $2\left(-f_D(c) + f_{\bar{D}}(c)\right)\left(n_D S_D(c) + n_{\bar{D}} S_{\bar{D}}(c)\right)\left(n_D F_D(c) + n_{\bar{D}} F_{\bar{D}}(c)\right)$ in (2.12) vanishes at $c = c_J$. Now, since $S_D(c) - S_{\bar{D}}(c)$ and $n_D f_D(c) + n_{\bar{D}} f_{\bar{D}}(c)$ are indeed positive at any c of X, we can conclude that the sign of (2.12) at $c = c_J$ depends only on the term $n_D F_D(c) + n_{\bar{D}} F_{\bar{D}}(c) - n_D S_D(c) - n_{\bar{D}} S_{\bar{D}}(c)$. Using the equalities demonstrated by Liu [15], i.e. $F_D(c_J) = S_{\bar{D}}(c_J)$ and $F_{\bar{D}}(c_J) = S_D(c_J)$, and by some algebra, the aforementioned term (2.12) can be written as:

$$(n_{\bar{D}} - n_D)(S_D(c_J) - S_{\bar{D}}(c_J)) \quad (2.13)$$

By considering that $S_D(c_J) - S_{\bar{D}}(c_J) > 0$, the sign of expression (2.13) depends only on the values of $(n_{\bar{D}} - n_D)$. Thus, under a balanced scenario ($n_D = n_{\bar{D}}$), the true cut-point of the Youden index method $c_J = \mu_D/2$ also maximizes the objective function $CHI_1^2(c)$ (2.7) since the first derivative function vanishes. If we consider a non-balanced scenario with $n_D < n_{\bar{D}}$, we can observe that $\partial CHI_1^2(c)/\partial c|_{c=c_J} > 0$, and thus $c_{\chi^2} > c_J$.

Furthermore, under the aforementioned Normal homoscedastic distribution scenario, it can be also easily proven that the true cut-point $c_J = \mu_D/2$ is also the maximum point underlying the concordance probability and the point closest-to-(0,1) corner in the ROC plane objective functions (2.10) and (2.11). In fact, it can be shown that $\frac{\partial CZ(c)}{\partial c}|_{c=c_J} = 0$ when $\frac{f_D(c)}{f_{\bar{D}}(c)} = \frac{S_D(c)}{F_{\bar{D}}(c)}$ and $\frac{\partial ER^2(c)}{\partial c}|_{c=c_J} = 0$ when $\frac{f_D(c)}{f_{\bar{D}}(c)} = \frac{S_{\bar{D}}(c)}{F_D(c)}$.

These equations are verified using the aforementioned equalities described by Liu [15].

Figure 2.6 shows the population objective functions (2.7), (2.9), (2.10) and (2.11) in case of homoscedastic Normal distributions, with μ_D=2.56, i.e. J=0.8 and CZ=0.81, for both a balanced ($n_D = n_{\bar{D}}$=100, Panel A) and a non-balanced ($n_D = 50 < n_{\bar{D}} = 150$, Panel B) design. In the balanced homoscedastic scenario, the objective functions (2.7), (2.9), (2.10) and (2.11) reach their maximum in correspondence of the same true cut-point, i.e. $c_{opt} = \mu_D/2$ (see above). Analytically, this common cut-point occurs at the intersection between the Normal probability density functions of diseased (i.e., $f_D(c)$) and non-diseased subjects (i.e., $f_{\bar{D}}(c)$). Conversely, in the non-balanced homoscedastic case, the three ROC-based methods lead to

the same true cut-point c_{ROC}, which differs from the one based on the minimum P-value approach method, being $c_{\chi^2} > c_{ROC}$. In particular, the true cut-point c_{χ^2} of the minimum P-value approach increases with increasing prevalence of the disease in the sample as shown above.

2.4.2 THE GAMMA PARAMETRIC SCENARIO

Let us now assume that X is Gamma distributed in the diseased population as $X_D \sim G(\alpha_D = 2.5, \beta_D)$ and in the non-diseased population as $X_{\bar{D}} \sim G(\alpha_{\bar{D}} = 1.5, \beta_{\bar{D}} = 1)$. This implies that $S_D(c) = \frac{1}{\Gamma(2.5)(\beta_D)^{2.5}} \int_c^{\infty} x^{1.5} e^{-x/\beta_D} dx$ and $S_{\bar{D}}(c) = \frac{1}{\Gamma(1.5)} \int_c^{\infty} x^{0.5} e^{-x} dx$. If β_D is set equal to {0.79, 1.22, 1.97, 3.82}, the corresponding maximum values of the Youden function (2.9) are $J(c_J)$={0.2, 0.4, 0.6, 0.8}, the corresponding maximum values of the concordance probability (2.10) are $CZ(c_{CZ})$={0.36, 0.49, 0.64, 0.81} while the minimum values of the objective function (2.11) of the point closest-to-(0,1) corner in the ROC plane approach are $ER(c_{ER})$={0.57, 0.42, 0.28, 0.14}. As in the Normal distribution case, the corresponding maximum values of the Chi-square function (2.7) depend on the size of the sample; by considering $n_D = n_{\bar{D}}$=100, $CHI_1^2(c_{\chi^2})$={9, 32, 72, 128}. We set the Gamma parameters to ensure a wide variety of classification accuracies, ranging from a poor one (J=0.2 and CZ=0.36) to a high one (J=0.8 and CZ=0.81) [21]. In these scenarios, the objective functions (2.7), (2.9), (2.10) and (2.11) do not reach the maximum in correspondence of the same true cut-point (Figure 2.7), and a closed form for the true cut-point cannot be derived for the investigated methods.

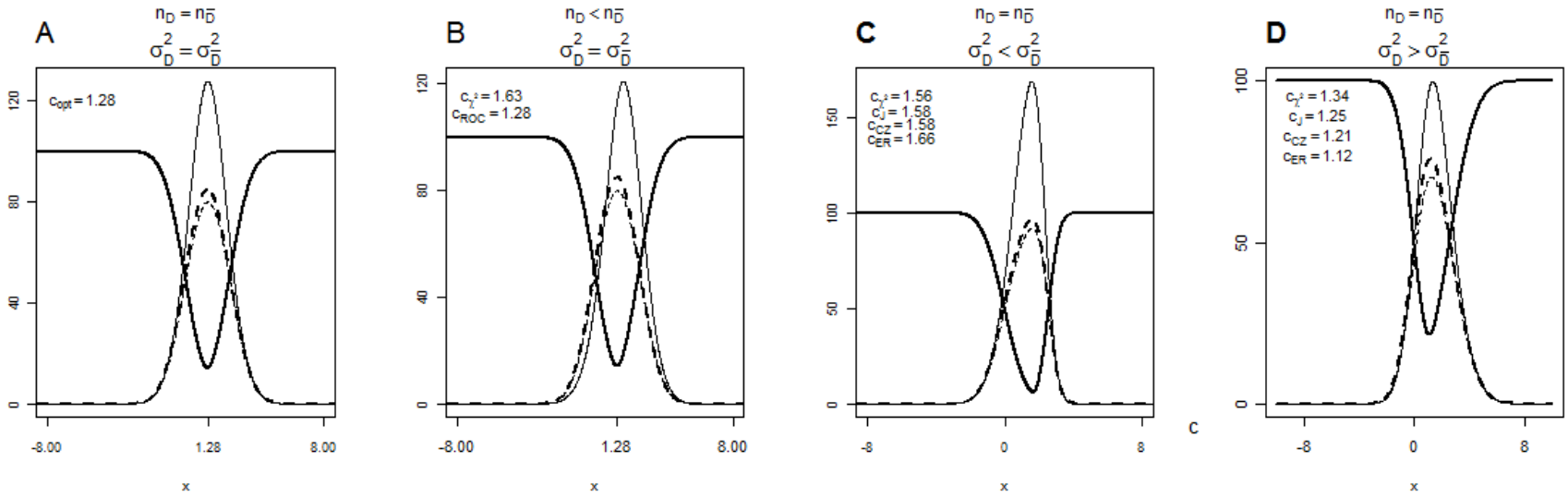

Figure 2.6 Population objective functions $CHI_1^2(c)$ (thin line), J(c) (dashed line), CZ(c) (thick dashed line) and ER(c) (thick line). The homoscedastic Normal distribution scenario with a balanced design ($n_D = n_{\bar{D}}$=100) is represented in **Panel A**, the non-balanced (n_D=50< $n_{\bar{D}}$=150) design is represented in **Panel B**. In both cases the underlying distributions are $X_D \sim N(\mu_D = 2.56, \sigma_D = 1)$ and $X_{\bar{D}} \sim N(\mu_{\bar{D}} = 0, \sigma_{\bar{D}} = 1)$, leading to J=0.8, CZ=0.81. The heteroscedastic Normal distribution scenarios with a balanced design are represented in **Panels C** and **D.** In the first case $X_D \sim N(\mu_D = 2.56, \sigma_D = 0.5)$ and $X_{\bar{D}} \sim N(\mu_{\bar{D}} = 0, \sigma_{\bar{D}} = 1)$, leading to J=0.92 and CZ=0.92, while in the second case $X_D \sim N(\mu_D = 2.56, \sigma_D = 1.5)$, $X_{\bar{D}} \sim N(\mu_{\bar{D}} = 0, \sigma_{\bar{D}} = 1)$, leading to J=0.70, CZ=0.72. In all panels the values of the objective functions $CHI_1^2(c)$, J(c) and ER(c) are multiplied by 100 whereas CZ(c) is multiplied by 110.

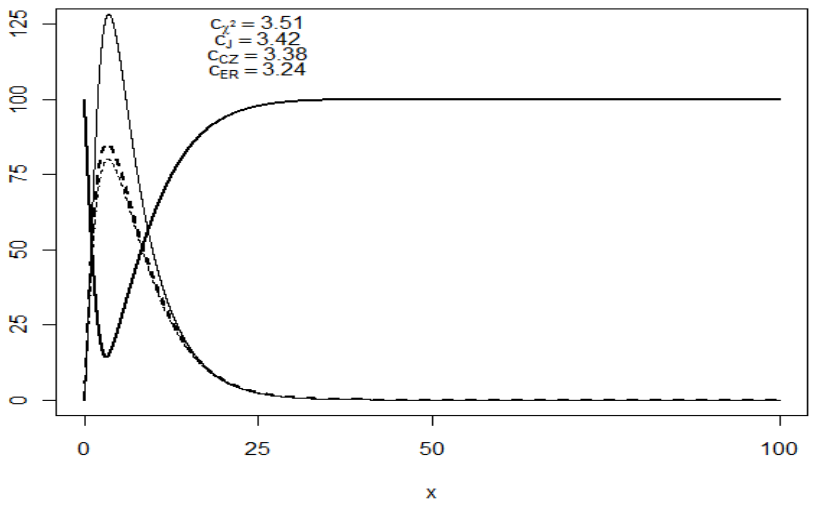

Figure 2.7 Population objective functions (thin line), J(c) (dashed line), CZ(c) (thick dashed line) and ER(c) (thick line) for the Gamma distribution scenario with $X_D \sim G(\alpha_D = 2.5, \beta_D = 3.82)$ and $X_{\bar{D}} \sim G(\alpha_{\bar{D}} = 1.5, \beta_{\bar{D}} = 1)$, i.e. J=0.8 and CZ=0.81.

2.5 SIMULATION PROTOCOL

We compare the performance of the minimum P-value (2.7) and the three ROC-based methods, Youden index (2.9), concordance probability (2.10) and point closest-to-(0,1) corner in the ROC plane (2.11), in the estimation of the optimal cut-point. We consider first the Normal homoscedastic scenario with balanced design ($n_D = n_{\bar{D}}$), where all the investigated methods identify theoretically the same true cut-point c_{opt}. Second, we analyse the non-balanced (specifically, $n_D < n_{\bar{D}}$) Normal case, where the ROC-based methods lead to the same true cut-point c_{ROC}, which differs from the true cut-point c_{χ^2} underlying the

minimum P-value approach. Last, the Gamma case is also considered. All simulations are performed under the same parametric scenarios introduced in paragraph 2.4 (page 20).

We generate 1000 samples of size $n_D = n_{\bar{D}}$=50, $n_D = n_{\bar{D}}$ =100 and $n_D = n_{\bar{D}}$=200. Moreover, in the non-balanced designs ($n_D < n_{\bar{D}}$), we generate 1000 samples of size n_D=50, $n_{\bar{D}}$=100 and n_D=50, $n_{\bar{D}}$=150 and n_D=50, $n_{\bar{D}}$=200. For each sample, we determine by numerical maximization the optimal cut-points estimates $\hat{c}_{\chi^2}$, $\hat{c}_J$, $\hat{c}_{CZ}$ and $\hat{c}_{ER}$ for the minimum P-value [13] and the three ROC-based methods, Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18], respectively. The relative bias of each method is computed by $E[(\hat{c}_\cdot - c_\cdot)/c_\cdot]$, while mean square error (MSE) is also determined as $E[(\hat{c}_\cdot - c_\cdot)^2]$. R simulation code is reported in Appendix 1.

2.5.1 BOOTSTRAP RESAMPLING TECHNIQUE

We apply the bootstrap resampling technique [22] to estimate the standard deviation and the confidence interval (CI) for the optimal cut-point. Within each of the 1000 generated samples, random sampling with replacement is used to draw 200 bootstrap samples to calculate the bootstrap estimate $\hat{c}_B$ (B=1, …, 200). Then, we apply the basic percentile method, taking the 0.025 and 0.975 percentiles of the $\hat{c}_B$ bootstrap distribution in order to construct a 95% CI of the optimal cut-point for each simulated sample.

The standard deviation SD_B of the 200 bootstrap cut-point estimates is used as the estimator of the standard deviation for the cut-point for each simulated sample.

Within each of the aforementioned scenarios, the CI for the cut-point is subsequently evaluated by computing coverage probability and mean length. The coverage probability is the proportion of times that the bootstrap confidence interval contains the true cut-point, while the mean length is a measure of the precision of the confidence interval around the estimated cut-point. The wider is the confidence interval, the higher is the uncertainty related to the estimated cut-point. R code for bootstrap resampling is shown in Appendix 2.

2.5.2 SIMULATION RESULTS

The results of the balanced design under Normal homoscedastic distributions are shown in Table 2.1. The relative bias of the investigated methods is small on all levels of classification accuracy. While the minimum P-value [13], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18] methods generally underestimate the true cut-point c_{opt}, the Youden index method [14] overestimates c_{opt}. By comparing the MSEs, it can be noticed that the point closest-to-(0,1) corner in the ROC plane and concordance probability methods have better performance than the other methods. Indeed, the MSE is inversely related to sample size. The performance of all methods improves with increasing classification accuracy.

Table 2.2 shows the bootstrap standard deviation, coverage probability and mean length of the 95% bootstrap CI for the cut-point for the balanced design under Normal homoscedastic distributions. The SD_B of the minimum P-value approach is systematically greater than the SD_B of the three ROC-based methods. 95% bootstrap CIs are narrower when considering the scenarios with better classification accuracies, i.e. J of 0.6

Table 2.1 Relative Bias and Mean Square Error (MSE) of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Normal homoscedastic balanced scenario[†].

			Sample sizes	Minimum P-value		Youden Index		Concordance probability		Point closest-to-(0,1) corner	
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	c_{opt}	$n_D = n_{\bar{D}}$	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE
0.2	**0.36**	0.25	50	0.0582	0.4807	0.0826	0.2332	0.0230	0.0733	-0.0261	0.0510
			100	-0.1217	0.4399	0.0850	0.1647	0.0346	0.0418	0.0254	0.0285
			200	0.0328	0.3259	0.0252	0.1044	-0.0291	0.0264	-0.0050	0.0184
0.4	**0.49**	0.52	50	-0.0013	0.2590	0.0517	0.1254	0.0047	0.0683	-0.0201	0.0434
			100	0.0221	0.1845	0.0372	0.0812	-0.0047	0.0390	-0.0041	0.0221
			200	0.0210	0.1225	0.0138	0.0513	0.0001	0.0250	-0.0045	0.0144
0.6	**0.64**	0.84	50	-0.0030	0.1332	0.0416	0.0835	-0.0108	0.0559	-0.0214	0.0355
			100	0.0067	0.0876	0.0237	0.0508	-0.0042	0.0356	-0.0042	0.0219
			200	0.0072	0.0571	0.0222	0.0354	-0.0012	0.0242	-0.0040	0.0138
0.8	**0.81**	1.28	50	-0.0221	0.0893	0.0043	0.0645	-0.0183	0.0520	-0.0254	0.0404
			100	-0.0097	0.0558	0.0120	0.0408	-0.0104	0.0341	-0.0112	0.0248
			200	-0.0055	0.0354	0.0054	0.0268	-0.0098	0.0223	-0.0118	0.0137

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size. [†] $X_D \sim N(\mu_D, 1)$, $X_{\bar{D}} \sim N(0,1)$. [‡] The levels of J and CZ are achieved by μ_D= 0.51, 1.05, 1.68, 2.56, respectively.

Table 2.2 Bootstrap standard deviation, coverage probability and mean length of the 95% confidence interval estimation of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Normal homoscedastic balanced scenario[†].

		Sample sizes	Minimum P-value			Youden Index			Concordance probability			Point closest-to-(0,1) corner		
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	$n_D = n_{\bar{D}}$	SD_B	Coverage	Length	SD_B	Coverage	Length	SD_B	Coverage	Length	SD_B	Coverage	Length
0.2	**0.36**	50	0.6166	0.978	2.1259	0.4428	0.963	1.5435	0.2733	0.960	0.9705	0.2329	0.957	0.8260
		100	0.5802	0.979	2.0000	0.3888	0.965	1.3715	0.2156	0.954	0.7598	0.1819	0.949	0.6397
		200	0.5229	0.972	1.8157	0.3214	0.973	1.1441	0.1678	0.965	0.5981	0.1404	0.962	0.5004
0.4	**0.49**	50	0.4493	0.962	1.5668	0.3289	0.962	1.1573	0.2523	0.960	0.8878	0.2050	0.960	0.7185
		100	0.4082	0.962	1.4333	0.2759	0.954	0.9746	0.2016	0.951	0.7115	0.1597	0.945	0.5677
		200	0.3402	0.963	1.2038	0.2213	0.956	0.7880	0.1581	0.958	0.5624	0.1233	0.953	0.4393
0.6	**0.64**	50	0.3306	0.952	1.1468	0.2728	0.944	0.9381	0.2400	0.937	0.8331	0.1959	0.926	0.6854
		100	0.2852	0.960	0.9999	0.2226	0.955	0.7823	0.1898	0.950	0.6716	0.1483	0.943	0.5247
		200	0.2379	0.971	0.8315	0.1842	0.958	0.6494	0.1545	0.957	0.5440	0.1171	0.950	0.4127
0.8	**0.81**	50	0.2712	0.922	0.9088	0.2504	0.922	0.8480	0.2350	0.911	0.8036	0.2052	0.894	0.7020
		100	0.2234	0.950	0.7814	0.1996	0.946	0.7012	0.1874	0.938	0.6610	0.1537	0.919	0.5350
		200	0.1829	0.963	0.6411	0.1594	0.960	0.5636	0.1505	0.952	0.5339	0.1174	0.944	0.4165

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size. [†] $X_D \sim N(\mu_D, 1)$, $X_{\bar{D}} \sim N(0,1)$. [‡] The levels of J and CZ are achieved by μ_D= 0.51, 1.05, 1.68, 2.56, respectively.

and 0.8, and CZ of 0.64 and 0.81. Moreover, the point closest-to-(0,1) corner in the ROC plane method achieves the narrowest CIs in all the analysed scenarios. Coverage probabilities are close to the nominal level.

Table 2.3 shows the results of the unbalanced homoscedastic Normal case. It can be noticed that the true c_{χ^2} depends on the prevalence of the disease in the sample, $n_D/(n_D + n_{\bar{D}})$. In particular, a reduction of the prevalence determines a systematic shift to the right of the true c_{χ^2}, while the three ROC-based methods lead to the same true cut-point c_{ROC}, which differs from the one based on the minimum P-value approach method, being $c_{\chi^2} > c_{ROC}$. The relative bias of the investigated methods is small on all levels of classification accuracy. The point closest-to-(0,1) corner in the ROC plane method achieves the lowest MSE of the estimates of the optimal cut-point, too.

Table 2.4 shows the bootstrap standard deviation, coverage probability and mean length of the 95% bootstrap CI for the cut-point for the unbalanced design under Normal homoscedastic distributions. Even in this scenario, the point closest-to-(0,1) corner in the ROC plane method achieves the lowest SD_B of the estimates of the optimal cut-point. Coverage probabilities are also close to the nominal level.

Table 2.3 Relative Bias and Mean Square Error (MSE) of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Normal homoscedastic unbalanced scenario[†].

				Sample sizes		Minimum P-value		Youden Index		Concordance probability		Point closest-to-(0,1) corner	
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	c_{χ^2}	c_{ROC}	n_D	$n_{\bar{D}}$	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE
0.2	**0.36**	0.39	0.25	50	100	0.0371	0.4496	0.2420	0.1956	0.1248	0.0550	0.0967	0.0390
		0.46		50	150	-0.0108	0.4358	0.0609	0.1928	0.1426	0.0545	0.1303	0.0403
		0.51		50	200	-0.0079	0.4001	0.2237	0.1758	0.1760	0.0536	0.1513	0.0381
0.4	**0.49**	0.75	0.52	50	100	0.0566	0.2118	0.0895	0.0973	0.0559	0.0530	0.0354	0.0331
		0.87		50	150	0.0377	0.1915	0.0669	0.0908	0.0701	0.0538	0.0530	0.0328
		0.96		50	200	0.0336	0.1669	0.0852	0.0905	0.0573	0.0475	0.0555	0.0307
0.6	**0.64**	1.09	0.84	50	100	0.0074	0.1042	0.0352	0.0667	0.0166	0.0477	0.0093	0.0287
		1.23		50	150	0.0161	0.1063	0.0414	0.0613	0.0231	0.0440	0.0289	0.0288
		1.33		50	200	0.0165	0.0929	0.0428	0.0654	0.0378	0.0443	0.0452	0.0262
0.8	**0.81**	1.50	1.28	50	100	-0.0105	0.0650	0.0166	0.0546	0.0030	0.0488	0.0048	0.0335
		1.63		50	150	-0.0098	0.0588	0.0344	0.0508	0.0265	0.0443	0.0238	0.0310
		1.72		50	200	-0.0138	0.0576	0.0269	0.0474	0.0199	0.0404	0.0228	0.0271

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size.

[†] $X_D \sim N(\mu_D, 1)$, $X_{\bar{D}} \sim N(0,1)$. [‡] The levels of J and CZ are achieved by μ_D= 0.51, 1.05, 1.68, 2.56, respectively.

Table 2.4 Bootstrap standard deviation, coverage probability and mean length of the 95% confidence interval estimation of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Normal homoscedastic unbalanced scenario[†].

		Sample sizes		Minimum P-value			Youden Index			Concordance probability			Point closest-to-(0,1) corner		
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	n_D	$n_{\bar{D}}$	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length
0.2	**0.36**	50	100	0.6048	0.978	2.0550	0.4270	0.968	1.5055	0.2525	0.962	0.8875	0.2132	0.955	0.7544
		50	150	0.5808	0.971	1.9671	0.4149	0.955	1.4502	0.2378	0.948	0.8414	0.1999	0.944	0.7127
		50	200	0.5568	0.974	1.8945	0.4041	0.969	1.4227	0.2350	0.951	0.8229	0.1978	0.948	0.7010
0.4	**0.49**	50	100	0.4360	0.959	1.5255	0.3081	0.958	1.0866	0.2298	0.949	0.8073	0.1843	0.946	0.6547
		50	150	0.4116	0.950	1.4354	0.3012	0.959	1.0562	0.2234	0.952	0.7875	0.1782	0.945	0.6332
		50	200	0.3850	0.954	1.3330	0.2948	0.960	1.0353	0.2180	0.943	0.7661	0.1740	0.937	0.6136
0.6	**0.64**	50	100	0.3139	0.960	1.0980	0.2513	0.954	0.8852	0.2185	0.952	0.7725	0.1737	0.942	0.6130
		50	150	0.3055	0.947	1.0643	0.2424	0.953	0.8521	0.2107	0.943	0.7419	0.1671	0.929	0.5877
		50	200	0.2928	0.965	1.0214	0.2369	0.948	0.8303	0.2047	0.939	0.7159	0.1608	0.924	0.5649
0.8	**0.81**	50	100	0.2422	0.926	0.8437	0.2245	0.946	0.7757	0.2135	0.940	0.7364	0.1788	0.931	0.6220
		50	150	0.2336	0.929	0.8105	0.2187	0.929	0.7598	0.2083	0.928	0.7234	0.1740	0.914	0.6054
		50	200	0.2263	0.952	0.7889	0.2111	0.926	0.7354	0.2019	0.929	0.7035	0.1669	0.925	0.5879

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size. [†] $X_D \sim N(\mu_D, 1)$, $X_{\bar{D}} \sim N(0,1)$. [‡]The levels of J and CZ are achieved by μ_D= 0.51, 1.05, 1.68, 2.56, respectively.

Under a Gamma distribution assumption with a balanced design (Table 2.5), it can be noticed that the theoretical true cut-points c_{χ^2}, c_J, c_{CZ} and c_{ER} are all different. However, we fail to recognize a systematic ordering between these theoretical cut-points. Concerning the relative performance of the investigated methods, we note a negligible relative bias in the estimate of the optimal cut-point, except for the minimum P-value approach in the scenario with poor classification accuracy, J=0.2. In all the analysed cases, the point closest-to-(0,1) corner in the ROC plane method achieves the lowest SD_B of the estimates of the optimal cut-point (Table 2.6). Coverage probabilities are also close to the nominal level.

Summarizing, in all the analysed simulation scenarios, the point closest-to-(0,1) corner in the ROC plane and concordance probability approaches show a better performance in the estimation of the optimal cut-point than the minimum P-value and Youden index methods. This finding is also confirmed by the bootstrap standard deviation and mean length of the 95% bootstrap CI.

Table 2.5 Relative Bias and Mean Square Error (MSE) of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Gamma balanced scenario [†].

						Sample sizes	Minimum P-value		Youden Index		Concordance probability		Point closest-to-(0,1) corner	
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	c_{χ^2}	c_J	c_{CZ}	c_{ER}	$n_D = n_{\bar{D}}$	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE
0.2	**0.36**	0.80	1.12	1.35	1.38	50	0.3826	0.5237	0.1377	0.2329	0.0113	0.0741	-0.0009	0.0587
						100	0.2517	0.3075	0.0920	0.1502	0.0048	0.0460	-0.0025	0.0322
						200	0.1524	0.1624	0.0631	0.0978	0.0069	0.0299	0.0035	0.0230
0.4	**0.49**	1.73	1.79	1.81	1.82	50	0.0288	0.4599	0.0211	0.2158	-0.0004	0.1229	-0.0102	0.0760
						100	0.0411	0.3651	0.0255	0.1509	0.0091	0.0785	0.0038	0.0496
						200	0.0277	0.2429	0.0097	0.0961	0.0032	0.0445	0.0023	0.0291
0.6	**0.64**	2.54	2.45	2.41	2.36	50	0.0037	0.4681	0.0263	0.2905	-0.0027	0.1960	-0.0059	0.1245
						100	0.0151	0.3227	0.0229	0.1941	0.0030	0.1271	-0.0012	0.0712
						200	-0.0020	0.2135	0.0114	0.1185	-0.0067	0.0716	-0.0012	0.0434
0.8	**0.81**	3.51	3.42	3.38	3.24	50	-0.0114	0.5465	0.0082	0.4705	-0.0319	0.3629	-0.0167	0.2474
						100	-0.0067	0.3467	0.0065	0.2647	-0.0146	0.2259	-0.0031	0.1387
						200	-0.00004	0.2174	0.0023	0.1642	-0.0078	0.1388	-0.0050	0.0820

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size. [†] $X_D \sim G(2.5, \beta_D)$, $X_{\bar{D}} \sim G(1.5,1)$. [‡] The levels of J and CZ are achieved by β_D= 0.79, 1.22, 1.97, 3.82, respectively.

Table 2.6 Bootstrap standard deviation, coverage probability and mean length of the 95% confidence interval estimation of the minimum P-value [*], Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. The Gamma balanced scenario [†].

		Sample sizes	Minimum P-value			Youden Index			Concordance probability			Point closest-to-(0,1) corner		
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	$n_D = n_{\bar{D}}$	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length	SD_B	Coverage	Mean length
0.2	**0.36**	50	0.6808	0.942	2.3320	0.4722	0.959	1.6308	0.2956	0.954	1.0255	0.2545	0.946	0.8855
		100	0.5465	0.956	1.8764	0.3882	0.975	1.3572	0.2282	0.954	0.8054	0.1944	0.955	0.6811
		200	0.4216	0.964	1.4473	0.3166	0.973	1.1203	0.1812	0.965	0.6414	0.1537	0.967	0.5432
0.4	**0.49**	50	0.6553	0.963	2.2558	0.4730	0.954	1.6585	0.3494	0.942	1.2364	0.2821	0.934	1.0047
		100	0.5770	0.978	1.9986	0.3932	0.965	1.3847	0.2799	0.954	0.9906	0.2212	0.952	0.7882
		200	0.4931	0.965	1.7300	0.3159	0.960	1.1199	0.2206	0.955	0.7864	0.1717	0.943	0.6116
0.6	**0.64**	50	0.6177	0.944	2.1299	0.5072	0.947	1.7606	0.4347	0.940	1.5301	0.3484	0.924	1.2140
		100	0.5310	0.958	1.8537	0.4171	0.958	1.4667	0.3499	0.951	1.2320	0.2700	0.946	0.9504
		200	0.4537	0.965	1.5791	0.3402	0.968	1.1963	0.2784	0.957	0.9900	0.2069	0.949	0.7395
0.8	**0.81**	50	0.6548	0.885	2.1745	0.6133	0.891	2.0568	0.5611	0.894	1.9164	0.4839	0.880	1.6490
		100	0.5467	0.934	1.8958	0.4976	0.942	1.7317	0.4574	0.933	1.6032	0.3629	0.936	1.2554
		200	0.4442	0.947	1.5672	0.3946	0.946	1.3886	0.3651	0.945	1.2932	0.2801	0.943	0.9866

[*] The theoretical maximum values of the Chi-square statistic are not shown since they depend on the sample size. [†] $X_D \sim G(2.5, \beta_D)$, $X_{\bar{D}} \sim G(1.5,1)$. [‡] The levels of J and CZ are achieved by β_D= 0.79, 1.22, 1.97, 3.82, respectively.

2.6 APPLICATION EXAMPLES

Two application examples of cut-point finding are discussed within this paragraph. The first one deals with the cut-point finding task for the continuous remodeling index (CRI), defined in electrocardiology as a function of signal-averaged P-wave duration and left atrial anteroposterior diameter at end-diastole [23]. Moreover, we use a subset of data from a nested case-control study on prostate cancer [24] aimed to find a cut-point value for the serum prostate specific antigen (PSA) velocity [25].

2.6.1 THE CUT-POINT FINDING FOR THE CONTINUOUS REMODELING INDEX

Atrial fibrillation (AF) is a frequent arrhythmia associated with almost all heart diseases. An AF episode is defined paroxysmal (ParAF) when it spontaneously stops within 7 days from the onset and persistent (PerAF) when pharmacological or electrical interventions are needed for interruption. Atrial fibrillation episodes are often interrupted with electrical or pharmacologic cardioversion within 7 days from the onset. In such a case, it is not possible to know the natural history of the arrhythmia. This raises the question on whether the episode in patients who underwent intervention within 7 days from the onset would have resolved spontaneously.

A continuous remodeling index (CRI) aimed to separate patients presenting ParAF from those with sustained PerAF (i.e., AF episodes interrupted 7 days or more after the onset) has been recently developed (Figure 2.8) [23]. Briefly, this index is computed for each AF patient using the individual values of the P-wave duration (SAPWd) and the left atrium diameter

(LADd) as CRI=-12+0.13·LADd (mm)+0.05·SAPWd (msec). SAPWd is an expression of the intra-atrial conduction time, while LADd is a measure of atrial dimension. The modifications of intra-atrial conduction time and of atrial dimensions are characteristics of atrial remodeling.

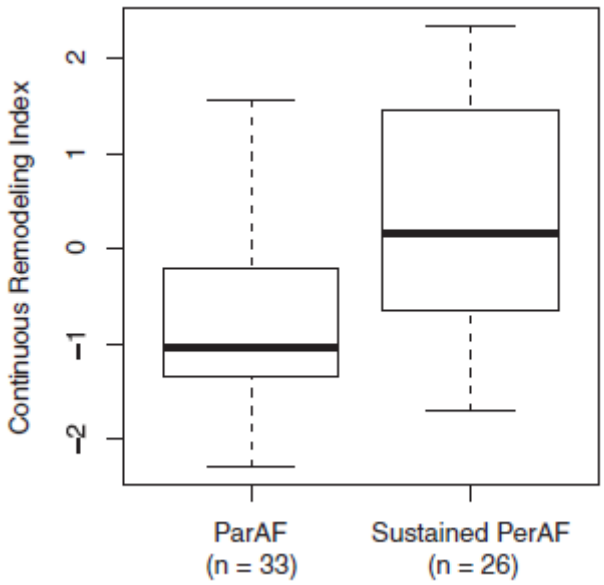

Figure 2.8 Box-plot representation of the CRI.

Table 2.7 reports some descriptive statistics of SAPWd and LADd for the ParAF and sustained PerAF groups. We can note that SAPWd is significantly longer in sustained PerAF than in ParAF patients (153 ± 15 vs. 142 ± 13 milliseconds, P = 0.004) while LADd is larger in sustained PerAF vs. ParAF patients (43 ± 6 vs 38 ± 5 mm, P = 0.002).

Figure 2.9 Panel A shows the ROC curve of the continuous remodeling index, along with the indication of the Youden index J, the concordance probability CZ and the distance from the (0,1) corner ER. The points on the Normal QQ plot graphs of ParAF (Figure 2.9 Panel B) and sustained PerAF (Figure 2.9 Panel C) fall approximately on the straight diagonal line.

This index could be considered normally distributed; as a confirmation, the Shapiro-Wilk non-parametric Normal distribution test returns a P-value of 0.29 and 0.16, respectively for the ParAF and sustained PerAF groups. The theoretical distributional scenario of the CRI is shown in Figure 2.9 Panel D. We tested the homoscedasticity assumption through the Bartlett's test, which returns a P of 0.26.

Table 2.7 Descriptive statistics[*] of the application example of cut-point finding for the CRI, from [23].

	ParAF (n=33)	**Sustained PerAF** (n=26)	**P-value** [**]
LADd (mm)	38 (5)	43 (6)	0.002
SAPWd (msec)	142 (13)	153 (15)	0.004
CRI [***]	-0.75 (0.93)	0.36 (1.15)	0.0001

[*] Values are expressed as mean (standard deviation).
[**] Computed from the t test for continuous variables
[***] CRI= -12+0.13·LADd (mm)+0.05·SAPWd (msec)

A cut-point c is needed to dichotomize the CRI result to define a criterion to divide AF episodes into PerAF or ParAF, so that the criteria identifies a PerAF condition when CRI is greater than c, and a ParAF condition when the CRI is less or equal than c. Since we are analysing a relatively balanced design (33 vs. 26 patients) with Normal homoscedastic distributions, the investigated methods lead to the same cut-point $\hat{c}_{opt}$ = -0.9 (Table 2.8). We also estimate the 95% bootstrap CI estimate for the cut-point for each of the investigated methods (Table 2.8), finding respectively for the minimum P-value, Youden index,

concordance probability and point closest-to-(0,1) corner in the ROC plane methods the following 95% CI estimates: (-1.12, 2.34), (-1.18, 0.46), (-1.12, 0.46) and (-1.09, 0.24). Indeed, the point closest-to-(0,1) corner in the ROC plane method achieves the narrowest CI. This is consistent with the results of the Normal homoscedastic simulation study.

Table 2.8 Estimated cut-point along with 95% bootstrap CI for the CRI [23].

Method of cut-point finding	Estimated cut-point	95% Bootstrap confidence interval
Minimum P-value [13]	-0.9	-1.12 - 2.34
Youden index [14]	-0.9	-1.18 - 0.46
Concordance probability [15]	-0.9	-1.12 - 0.46
Point closest-to-(0,1) corner in the ROC plane [18]	-0.9	-1.09 - 0.24

The binary RI classifies subjects as at risk of being sustained PerAF of ParAF depending on whether CRI is greater than -0.9 or less or equal than -0.9. Subsequently, a multivariate prognostic model was built up by adjusting the aforementioned binary RI by sex and age [23]. The model AUC was equal to 0.81. The model predicted a probability of approximately 85% (95% CI, 59%-96%) of having a sustained PerAF episode for males aged 69 years or older presenting CRI greater than −0.9 (RI = 1), whereas for females younger than 69 years presenting CRI less or equal than −0.9 (RI = 0), the predicted probability was 3% (95% CI, 0.3%-18%) [23].

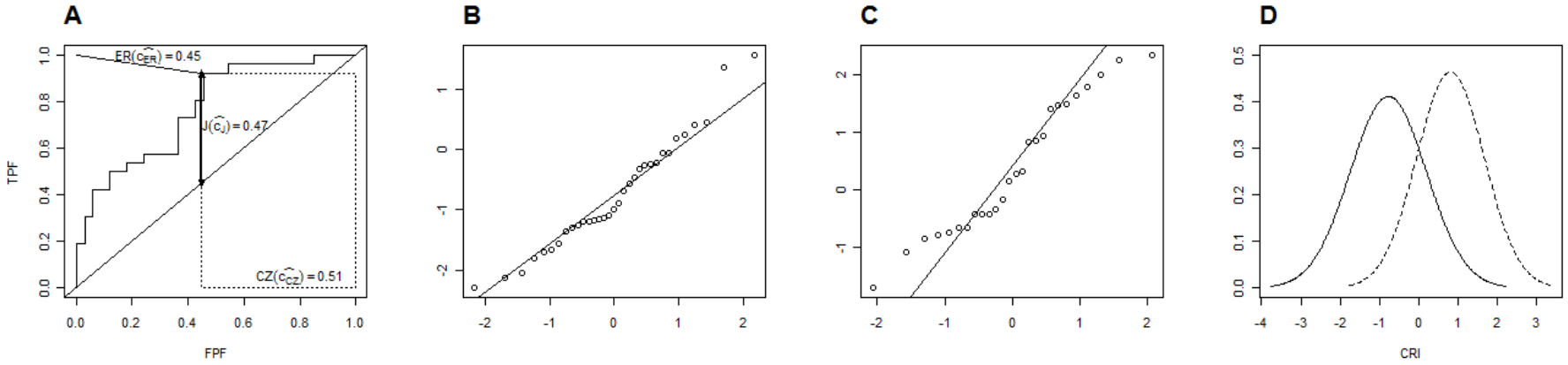

Figure 2.9 Descriptive analysis of the continuous remodeling index (CRI). **Panel A.** ROC curve depicting the Youden index J, the concordance probability CZ and and the distance from the (0,1) corner ER. **Panel B.** Normal QQ-plot among paroxysmal atrial fibrillation patients (ParAF). **Panel C.** Normal QQ-plot among persistent atrial fibrillation patients (PerAF). **Panel D.** Density functions among ParAF (solid line) and PerAF (dashed line).

2.6.2 THE CUT-POINT FINDING FOR THE PSA VELOCITY

We use data from a nested case-control study [24] (available for free download at *http://labs.fhcrc.org/pepe/book/#datasets*). This study aimed to analyse longitudinally the variation of total serum levels of PSA for 71 cases of prostate cancer and 70 controls matched to cases on date of birth and number of serum samples available for analysis.

PSA velocity (PSAV) measures how quickly PSA levels increase over time before prostate cancer diagnosis and it may have advantages over a single PSA measurement in differentiating between men with prostate cancer and controls [25]. Several methods are available to compute PSAV [26]. Given that PSA measurements are separated by a sufficient long time period, in this case study we decide to compute PSAV through the rate of PSA change between the first and last PSA measurement.

Table 2.9 shows some descriptive statistics of prostate cancer cases and controls. Our analyses are restricted to 57 prostate cancer cases and 60 controls reporting at least two PSA measurements. Mean ages are similar between men diagnosed with cancer compared with controls (62 vs. 60, P=0.07), while initial PSA levels (5.8 ng/ml vs. 1.2 ng/ml, P<0.0001), last PSA levels before diagnosis (18.1 ng/ml vs. 1.7 ng/ml, P<0.0001) and mean number of total PSA tests (4.4 vs. 4.8, P=0.0006) are different between cases and controls. PSAV is indeed higher in prostate cancer cases than in controls (3.7 ng/ml/year vs. 0.3 ng/ml/year, P<0.0001).

Figure 2.10 Panel A shows the ROC curve of the PSA velocity, along with the indication of the Youden index J, the concordance probability CZ and the distance from the (0,1) corner

ER. The distributions of the PSAV within controls (Figure 2.10 Panel B) and within the prostate cancer group (Figure 2.10 Panel C) are positive skewed, with some overlapping.

Table 2.9 Descriptive statistics * of the application example of cut-point finding for the PSAV, from [24].

	Prostate cancer cases (n=57)	**Matched controls** (n=60)	**P-value** **
Age (yr)	62 (6)	60 (5)	0.07
Initial PSA (ng/ml)	5.8 (7.8)	1.2 (0.6)	<0.0001
Last PSA before diagnosis (ng/ml)	18.1 (30.0)	1.7 (1.3)	<0.0001
No. of PSA tests	4.4 (1.2)	4.8 (1.1)	0.0006
PSAV (ng/ml/year)	3.7 (3.7)	0.3 (0.3)	<0.0001

* Values are expressed as mean (standard deviation).

** Computed from the t test for continuous variables.

A cut-point c for the PSAV is needed to discriminate suspicious prostate cancer. Here, the minimum p-value approach selects as cut-point $\hat{c}_{\chi^2}$ = 0.76 ng/ml/year, while the three ROC-based methods identifies the same cut-point $\hat{c}_{ROC}$ = 0.63 ng/ml/year (Table 2.10). This finding is indeed consistent with the behaviour of the c_{χ^2} with respect to the c_{ROC} observed theoretically within the Gamma distribution simulation scenario described in Table 2.5.

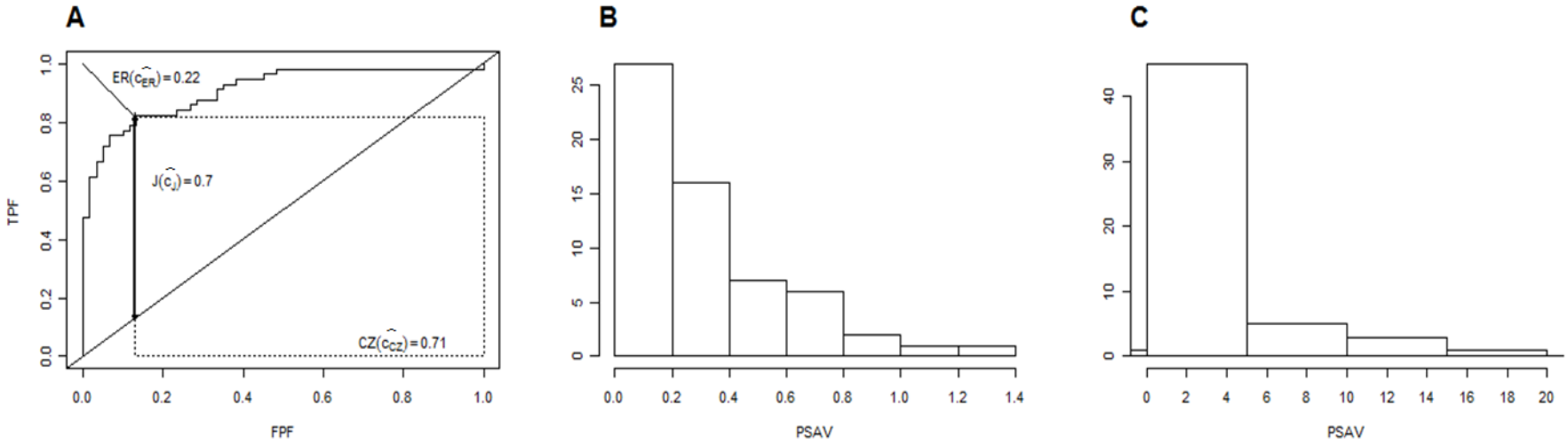

Figure 2.10 Descriptive analysis of the prostate specific antigen velocity (PSAV). **Panel A**. ROC curve depicting the Youden index J, the concordance probability CZ and the distance from the (0,1) corner ER. **Panel B.** Histogram of the PSAV within the control group. **Panel C.** Histogram of the PSAV within the prostate cancer group.

In fact, for relatively high values of classification accuracy, in this case J=0.7, c_{χ^2} tends to be greater than c_{ROC}.

The bootstrap 95% CI estimates of the cut-point are (0.34, 1.04), (0.34, 0.94), (0.39, 0.82) and (0.39, 0.76), respectively for the minimum P-value, Youden index, concordance probability and point closest-to-(0,1) corner in the ROC plane methods. This latter approach achieves the narrowest confidence interval (Table 2.10). This is consistent with the Gamma simulation results presented in Table 2.6.

Table 2.10 Estimated cut-point along with 95% bootstrap CI for the PSAV [24].

Method of cut-point finding	Estimated cut-point	95% Bootstrap confidence interval
Minimum P-value [13]	0.76	0.34 - 1.04
Youden index [14]	0.63	0.34 - 0.94
Concordance probability [15]	0.63	0.39 - 0.82
Point closest-to-(0,1) corner in the ROC plane [18]	0.63	0.39 - 0.76

CHAPTER 3

CUT-POINT FINDING FOR CENSORED FAILURE TIME OUTCOME

3

3.1 BASIC CONCEPTS OF SURVIVAL ANALYSIS

Classical survival analysis focuses on the time elapsed from an initiating event to an outcome event, or endpoint, of interest. Classical examples comprise time from disease diagnosis to death, or from an epidemiologic perspective, time from exposure to a risk factor to disease development. Generically, the time from the initiating event to the endpoint of interest is denoted as *survival time*, even when the endpoint is something different from death [27, 28].

Standard statistical methods such as ordinary linear regression are not applicable in a framework of survival data since when the study ends the endpoint of interest has occurred for some individuals, but not for others. Hence, there are some incomplete observations named as *censored survival times* [27, 28]. In fact, we can see from Figure 3.1 that subjects 1, 2, 4 and 6 enter the study at different calendar times and are then followed until the event occurs. The survival time of subject 5 is censored due to loss to follow-up, while subject 3 is followed until the closure of the study without having the event of interest, i.e.

administrative censoring. Notice that the censoring present in this framework occurs at the right-hand side, and so is called *right-censoring*. Moreover, in Figure 3.1 each subject enters the study at different calendar times, a phenomenon also known as *staggered entry* [27].

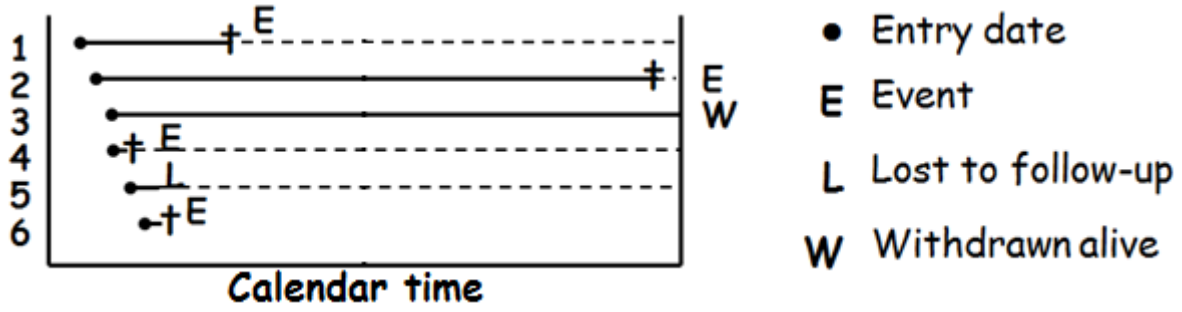

Figure 3.1 The Classical framework of survival data.

Formally, we assign to each individual - i = 1, ... , N - a time to the event Z_i, and a censoring time C_i. Both Z_i and C_i are non-negative random variables. δ is a (0,1) random variable such that if δ=1 the event of interest occurs during the study period, while δ=0 identifies one of the following situations: a person survives until the study ends (administrative censoring), a person is lost to follow-up or a person withdraws during the study period (censored survival time). The random variable $T_i=\min(Z_i,C_i)$ is such that $T_i=Z_i$ when $\delta_i=1$, while $T_i=C_i$ when $\delta_i=0$. We must be able to reasonably assume that the survival time Z is independent of the censoring time C. This is called *non-informative (*or *random) censoring* [27, 28].

The survival function S(t):

$$S(t) = P(Z > t) \tag{3.1}$$

gives the proportion of individuals for which the event has not yet occurred by time t, or alternatively, the probability that a subject survives longer than the specified time t. S(t) is

monotonically decreasing, with S(t=0)=1 and $\lim_{t\to+\infty} S(t) = 0$, since as t increases over time, more and more individuals will experience the event of interest [27, 28].

By definition, the cumulative density function of the random variable Z evaluated at time t is $F(t) = P(Z \leq t) = 1 - S(t)$, while the probability density function of Z is $f(t) = \frac{d}{dt}F(t) = \lim_{\Delta t\to 0} \frac{P(t\leq Z<t+\Delta t)}{\Delta t}$ [27, 28].

The hazard rate ʎ(t) is defined as:

$$\lambda(t) = \lim_{\Delta t\to 0} \frac{P(t \leq Z < t + \Delta t | Z \geq t)}{\Delta t} \tag{3.2}$$

and it represents the instantaneous probability of occurrence of the considered event at time t per unit time, given survival up to time t. The hazard function is always non-negative and it has no upper bound. The hazard function is sometimes called conditional failure rate.

Following the Bayes' rule, it could be derived after some algebra that $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\ln(S(t))$. As a consequence, the survival function S(t) (3.1) could be written as $S(t) = e^{-\int_0^t \lambda(u)du} = e^{-\Lambda(t)}$, being $\Lambda(t) = \int_0^t \lambda(u)\,du$ the cumulative hazard function, i.e. the total hazard an individual is exposed to up to time t [27, 28].

To estimate the survival function (3.1), the non-parametric maximum likelihood, parametric and semi-parametric methods could be used. To the aim of this thesis, only the non-parametric Kaplan-Meier estimator [29] is shown.

3.1.1 THE KAPLAN-MEIER ESTIMATOR

The non-parametric Kaplan-Meier estimator [29] is the most widely used estimator for estimating the survival function (3.1). Let us consider a sample of N individual observations

for which we order the J distinct event times, $t_{(1)} < t_{(2)} < \ldots < \ldots < t_{(j)} < \ldots < t_{(J)}$, with J≤N. At the time t of each event, we denote the number of patients at risk by $n_1, n_2, \ldots, n_j, \ldots, n_J$, and with $d_1, d_2, \ldots, d_j, \ldots, d_J$ the number of events among them.

The Kaplan-Meier estimator $\hat{S}(t)$ [29] of the survival time S(t) (3.1) is:

$$\hat{S}(t) = \prod_{j|t_{(j)} \leq t} \frac{n_j - d_j}{n_j} \tag{3.3}$$

where n_j denotes the number of subjects still at risk at the beginning of the j-th time unit and d_j the number of events during the j-th time unit. $\hat{S}(t)$ is a non-increasing step function that changes only at event times, and it is a right continuous function. Moreover, censored subjects influence indirectly only the height of the steps by eroding the number of at risk patients.

The variance of the Kaplan-Meier estimator (3.3) developed by Greenwood [30] is:

$$\mathrm{Var}\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum_{i=1}^{j-1} \frac{d_i}{n_i(n_i - d_i)} \tag{3.4}$$

According to formula (3.4), the precision of the survival estimate tends to decrease as the number of at risk subjects decreases, and consequently the variance increases.

3.2 CUT-POINT FINDING METHODS FOR CENSORED FAILURE TIME DATA

Let X denotes a continuous biomarker which is supposed to be related to the binary outcome in time (true disease status). We consider that subjects are free of the event when the biomarker X is measured at baseline, i.e. time zero. The interest is on whether the event can occur within a predefined follow-up interval [0,τ], where τ is a meaningful time point for

the clinical problem. The event (disease) indicator could be rewritten as $D=I(Z\leq\tau)$, where Z is the survival time from the biomarker measurement to the development of the event.

In this setting, the TPF and the FPF are:

$$\mathrm{TPF}(c) = \mathrm{P}(X > c | Z \leq \tau) = S_{Z\leq\tau}(c) \tag{3.5}$$

$$\mathrm{FPF}(c) = \mathrm{P}(X > c | Z > \tau) = S_{Z>\tau}(c) \tag{3.6}$$

As described above, Z_i denotes the time to development of the event (disease), C_i the censoring time, $T_i=\min(Z_i,C_i)$ the observed time, with $\delta_i=1$ if $T_i=Z_i$ and $\delta_i=0$ if $T_i=C_i$. Independence between Z and C is assumed. The observed data in a sample of N subjects is $\{(X_i, T_i, \delta_i, D_i); i=1, \ldots, N\}$, where:

- D_i is equal to 1 (hereafter denoted by D) if $T_i \leq \tau$ and $\delta_i=1$;
- D_i is equal to 0 (hereafter denoted by $\overline{D}$) if $T_i > \tau$ regardless of δ_i;
- D_i is missing if $T_i \leq \tau$ and $\delta_i=0$; i.e., censored subjects by τ, for whom it is not possible to know whether they would experience or not the disease within time τ.

Let us consider a sample of N subjects. For any c of X, we can define the following classification matrix:

	$X \leq c$	$X > c$	
$\overline{D}$	$n_{0\overline{D}}$	$n_{1\overline{D}}$	$n_{\overline{D}}$
D	n_{0D}	n_{1D}	n_{D}
Missing	n_{0C}	n_{1C}	n_{C}
			N

(3.7)

where $n_{0D} = \sum_{i=1}^{N} I(T_i \leq \tau)\, \delta_i I(X_i \leq c)$, $n_{1D} = \sum_{i=1}^{N} I(T_i \leq \tau)\, \delta_i I(X_i > c)$, $n_{0\bar{D}} = \sum_{i=1}^{N} I(T_i > \tau)\, I(X_i \leq c)$, $n_{1\bar{D}} = \sum_{i=1}^{N} I(T_i > \tau)\, I(X_i > c)$. Moreover, the disease status is missing for n_C=n_{0C}+n_{1C} censored subjects, where $n_{0C} = \sum_{i=1}^{N} I(T_i \leq \tau)\, (1 - \delta_i) I(X_i \leq c)$ and $n_{1C} = \sum_{i=1}^{N} I(T_i \leq \tau)\, (1 - \delta_i) I(X_i > c)$.

In this setting, the estimation of the TPF (3.5) and FPF (3.6) is not intuitive for the lack of information on the n_C censored subjects on whom the disease status is unknown. Several methods could be applied to supply to the lack of information on the disease status for censored subjects: weighting by inverse probability, direct estimation by imputation, indirect estimation by conditional probability and indirect estimation by Bayes theorem. It has been shown in the recent work of Antolini and Valsecchi [17] that these four different estimators of the TPF (3.5) and FPF (3.6) are equivalent. So, within this thesis, we only theoretically show the indirect estimation of the TPF (3.5) and FPF (3.6) by means of the Bayes theorem [31].

3.2.1 INDIRECT ESTIMATION OF TPF AND FPF BY BAYES THEOREM

Formally, starting from a partition $\{A_1, \ldots, A_k\}$ of a given sample space S and considering a separate event B, the Bayes theorem [31] states that:

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}, \text{for a fixed } j = 1, \ldots, k \tag{3.8}$$

By applying formula (3.8), the TPF (3.5) and the FPF (3.6) could be easily written as:

$$TPF(c) = S_{Z \leq \tau}(c) = \frac{(1 - P(Z > \tau|X > c))P(X > c)}{(1 - P(Z > \tau|X \leq c))P(X \leq c) + (1 - P(Z > \tau|X > c))P(X > c)} \tag{3.9}$$

$$\mathrm{FPF}(c) = S_{Z>\tau}(c) = \frac{P(Z > \tau | X > c)P(X > c)}{P(Z > \tau | X \leq c)P(X \leq c) + P(Z > \tau | X > c)P(X > c)} \quad (3.10)$$

Within formulae (3.9) and (3.10), two survival probabilities could be identified: $S_{X>c}(\tau) = P(Z > \tau | X > c)$ and $S_{X\leq c}(\tau) = P(Z > \tau | X \leq c)$. These probabilities can be estimated through the non-parametric Kaplan-Meier estimator (3.3) separately in the sample of subjects who have the biomarker value X greater or less or equal than c, respectively. The two probabilities $P(X > c)$ and $P(X \leq c) = 1 - P(X > c)$ could be easily estimated by $p = \frac{\sum_{i=1}^{N} I(X_i > c)}{N}$ and 1-p.

Consequently, the TPF sample estimates of (3.9) is

$$\hat{S}_{Z\leq\tau}(c) = \frac{\left(1 - \hat{S}_{X>c}(\tau)\right) p}{\left(1 - \hat{S}_{X\leq c}(\tau)\right) (1 - p) + \left(1 - \hat{S}_{X>c}(\tau)\right) p}$$

while the sample estimates of the FPF (3.10) is

$$\hat{S}_{Z>\tau}(c) = \frac{\hat{S}_{X>c}(\tau)\, p}{\hat{S}_{X\leq c}(\tau)\, (1 - p) + \hat{S}_{X>c}(\tau)\, p}$$

It could be observed that the two estimators (3.9) and (3.10) allows for the presence of biomarker-dependent censoring because the aforementioned survival functions $S_{X>c}(\tau)$ and $S_{X\leq c}(\tau)$ are estimated conditional on the biomarker value. Moreover, it can be proven by a little algebra that the two estimators (3.9) and (3.10) corresponds to the estimators obtained using the nearest neighbor approach described by Haegerty et al. [32].

The Youden index [14], the concordance probability [15] and the point closest-to-(0,1) corner in the ROC plane [18] methods could be easily extended to the censored failure time scenario. Conversely, we do not extend the minimum P-value method [13] to the censored

failure time outcome scenario since we showed within the binary outcome scenario (Chapter 2) a poor performance and a lack of clinical meaning of its objective function.

3.2.2 YOUDEN INDEX METHOD FOR CENSORED FAILURE TIME DATA

The Youden function J(c) [14] as a function of c of X is the difference between the population quantities $S_{Z\leq\tau}(c)$ (3.9) and $S_{Z>\tau}(c)$ (3.10):

$$J(c) = S_{Z\leq\tau}(c) - S_{Z>\tau}(c) \tag{3.11}$$

J(c) takes values between 0 when $S_{Z\leq\tau}(c) = S_{Z>\tau}(c)$ and 1 when $S_{Z\leq\tau}(c) = 1$ and $S_{Z>\tau}(c) = 0$. The Youden index J is defined as the maximum of the Youden function (3.11). Graphically, J represents the maximum vertical distance between the ROC curve and the diagonal chance line representing a useless biomarker. It can be interpreted as the net gain of the true positive fraction with respect to the false positive fraction (Figure 2.4 page 17).

The optimal cut-point $\hat{c}_J$ is the c maximizing the Youden function $\hat{J}(c) = \hat{S}_{Z\leq\tau}(c) - \hat{S}_{Z>\tau}(c)$ over all possible cut-point values c of X.

3.2.3 CONCORDANCE PROBABILITY METHOD FOR CENSORED FAILURE TIME DATA

The concordance probability [15] function is the product of the population quantities $S_{Z\leq\tau}(c)$ (3.9) and the complement to one of the $S_{Z\leq\tau}(c)$ (3.10) over all possible cut-point values c of X:

$$CZ(c) = S_{Z\leq\tau}(c) \cdot (1 - S_{Z>\tau}(c)) \tag{3.12}$$

CZ(c) ranges between 0 if $S_{Z\leq\tau}(c) = 0$ or $S_{Z>\tau}(c) = 1$, and 1 in the ideal case where $\hat{S}_{Z\leq\tau}(c) = 1$ and $S_{Z>\tau}(c) = 0$. CZ(c) could be also expressed as the area of a rectangle on the ROC curve of width $1 - S_{Z>\tau}(c)$ and height $S_{Z\leq\tau}(c)$ varying c (Figure 2.5 page 19, area of the dotted rectangle) and interpreted as the probability of being below or beyond the cut-point for any random pair of non-diseased and diseased subjects.

Following this approach, the optimal cut-point $\hat{c}_{CZ}$ is the maximum point of the concordance probability function $\widehat{CZ}(c) = \hat{S}_{Z\leq\tau}(c) \cdot (1 - \hat{S}_{Z>\tau}(c))$ over all possible cut-point values c of X.

3.2.4 POINT CLOSEST-TO-(0,1) CORNER IN THE ROC PLANE APPROACH FOR CENSORED FAILURE TIME DATA

The point closest-to-(0,1) corner in the ROC plane [18] objective function could be defined by applying the Euclidean distance between the point on the ROC curve defined by the population quantities $S_{Z\leq\tau}(c)$ (3.9) and $S_{Z>\tau}(c)$ (3.10) and the point (0,1) representing the perfect biomarker:

$$ER(c) = \sqrt{(S_{Z\leq\tau}(c) - 1)^2 + S_{Z>\tau}(c)^2} \quad (3.13)$$

Following this approach, the optimal cut-point $\hat{c}_{ER}$ is the c that achieves the minimum of the objective function $\widehat{ER}(c) = \sqrt{(\hat{S}_{Z\leq\tau}(c) - 1)^2 + \hat{S}_{Z>\tau}(c)^2}$ over all possible cut-point values c of X (Figure 2.5 page 19, length of the thin line segment).

3.3 SIMULATION STUDY PROTOCOL

We conduct a simulation study within a censored failure time outcome scenario to compare the performance of the Youden index (3.11), concordance probability (3.12) and point closest-to-(0,1) corner in the ROC plane (3.13) methods in the estimation of the optimal cut-point.

The time-to-event Z is generated according to a parametric exponential model. Z is an exponential distributed random variable with density function $f(t) = \lambda e^{-\lambda t}$, being λ a real number greater than zero. Consequently, the survival function S(t) (3.1) could be written as $S(t) = 1 - \int_0^t f(u)\,du = e^{-\lambda t}$. So, it could be easily derived that we are dealing with a constant hazard parametric model [28], being $\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$. For our simulation scenario, we set the constant hazard λ=2, so that the survival function (3.1) is parametrically modeled by $S(t) = e^{-2t}$.

To simulate independent censoring, the censoring time C is generated according to a uniform distribution in the interval [0,b]. When we consider the scenario with a true cases fraction of 50%, b is set equal to 2, 1 and 0.66 time units in order to achieve different censoring levels, i.e. 12%, 25% and 38%. Within the non-balanced scenario, b is set equal to 0.67, 0.50 and 0.40 time units to achieve a censoring fraction of 25% when considering different cases fractions, i.e. 33%, 25% and 20%, respectively. For each subjects i=1, …, N, the observed time $T_i=\min(Z_i,C_i)$ is then assigned, being $\delta_i=1$ if $T_i=Z_i$ and $\delta_i=0$ if $T_i=C_i$.

The biomarker X is separately generated for diseased and non-diseased subjects from Gaussian and Gamma distributions as described below in paragraphs 3.3.1 and 3.3.2.

3.3.1 THE GAUSSIAN SCENARIO

The biomarker X is generated from a Normal distribution $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, \sigma_{Z\leq\tau}=1)$ for diseased subjects - i.e. subjects with a true survival time Z from the biomarker measurement to the development of disease less or equal than τ - and from a standard Normal distribution $X_{Z>\tau}\sim N(\mu_{Z\leq\tau}=0, \sigma_{Z\leq\tau}=1)$ for non-diseased subjects - i.e. subjects with a true survival time Z from the biomarker measurement to the development of disease greater than τ. This means that $S_{Z\leq\tau}(c)=1-\Phi(c-\mu_{Z\leq\tau})$ and $S_{Z>\tau}(c)=1-\Phi(c)$, where Φ denotes the standard Normal distribution function.

As in the binary outcome scenario described in Chapter 2, $\mu_{Z\leq\tau}$ is set equal to {0.51, 1.05, 1.68, 2.56} in order to achieve different values of the Youden function (3.11), i.e. $J(c_J)$={0.2, 0.4, 0.6, 0.8}, of the concordance probability function (3.12), i.e. $CZ(c_{CZ})$={0.36, 0.49, 0.64, 0.81}, and of the objective function (3.13) of the point closest-to-(0,1) corner in the ROC plane approach, i.e. $ER(c_{ER})$={0.57, 0.42, 0.28, 0.14}. This set of values of $\mu_{Z\leq\tau}$ ensures a wide variety of classification accuracies, ranging from a poor one (J=0.2 and CZ=0.36) to a high one (J=0.8 and CZ=0.81) [21].

Within this scenario, the objective functions (3.11), (3.12), (3.13) reach their maximum in correspondence of the same true cut-point, i.e. $c_{opt}=\mu_{Z\leq\tau}/2$ (see paragraph 2.4.1 page 20 for the formal derivation). Analytically, this common cut-point occurs at the intersection between the Normal probability density functions of diseased, i.e. $f_{Z\leq\tau}(c)$, and non-diseased subjects, i.e. $f_{Z>\tau}(c)$.

3.3.2 THE GAMMA SCENARIO

The biomarker X is generated from a Gamma distribution $X_{Z\leq\tau}\sim G(\alpha_{Z\leq\tau}=2.5,\beta_{Z\leq\tau})$ for diseased subjects - i.e. subjects with a survival time Z from the biomarker measurement to the development of disease less or equal than τ - and $X_{Z>\tau}\sim G(\alpha_{Z>\tau}=1.5,\beta_{Z>\tau}=1)$ for non-diseased subjects – i.e. subjects with a survival time Z from the biomarker measurement to the development of disease greater than τ. This implies that $S_{Z\leq\tau}(c)=\frac{1}{\Gamma(2.5)(\beta_{Z\leq\tau})^{2.5}}\int_c^{+\infty}x^{1.5}e^{-x/\beta_{Z\leq\tau}}\,dx$ and $S_{Z>\tau}(c)=\frac{1}{\Gamma(1.5)}\int_c^{+\infty}x^{0.5}e^{-x/\beta_{Z>\tau}}\,dx$. If $\beta_{Z\leq\tau}$ is set equal to {0.79, 1.22, 1.97, 3.82}, the corresponding maximum values of the Youden function (3.11) are $J(c_J)$={0.2, 0.4, 0.6, 0.8}, the corresponding maximum values of the concordance probability (3.12) are $CZ(c_{CZ})$={0.36, 0.49, 0.64, 0.81} while the minimum values of the objective function (3.13) of the point closest-to-(0,1) corner in the ROC plane approach are $ER(c_{ER})$={0.57, 0.42, 0.28, 0.14}. We set the Gamma parameters to ensure a wide variety of classification accuracies, ranging from a poor one (J=0.2 and CZ=0.36) to a high one (J=0.8 and CZ=0.81) [21].

In these scenarios, the objective functions (3.11), (3.12), (3.13) do not reach the maximum in correspondence of the same true cut-point (Figure 2.7 page 26). Moreover, a closed form for the true cut-point cannot be derived for the investigated methods.

We generate 1000 samples of size N=100, N=200 and N=400 with a cases fraction of 50% and three different censoring levels, i.e. 12%., 25% and 38%. Moreover, in the non-balanced design, we generate 1000 samples of size N=150, N=200 and N=250 with different cases fractions, i.e. 33%, 15% and 20%, and a censoring level of 25%. For each sample, we

determine by numerical maximization the optimal cut-points estimates $\hat{c}_J$, $\hat{c}_{CZ}$ and $\hat{c}_{ER}$ for the Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18] methods, respectively. The relative bias of each method is computed by $E[(\hat{c}_{\cdot} - c_{\cdot})/c_{\cdot}]$, while mean square error (MSE) is also determined as $E[(\hat{c}_{\cdot} - c_{\cdot})^2]$. R simulation code is reported in Appendix 3.

3.3.3 BOOTSTRAP RESAMPLING TECHNIQUE

We apply the bootstrap resampling technique to estimate the standard deviation and the confidence interval (CI) for the optimal cut-point [22]. Following the Efron and Tibshirani's procedure [33], random sampling with replacement is used to draw 200 bootstrap samples from the empirical distribution function of $\{(X_i, T_i, \delta_i, D_i); i=1, \ldots, N\}$ in order to calculate the bootstrap estimate $\hat{c}_B$ (B=1, …, 200). Then, we apply the basic percentile method, taking the 0.025 and 0.975 percentiles of the $\hat{c}_B$ bootstrap distribution in order to construct a 95% CI of the optimal cut-point within each of the 1000 generated samples.

Each bootstrap sample contributes one cut-point estimate, so that the standard deviation of the 200 cut-point estimates is used as the bootstrap estimator of the standard deviation (SD_B) for the estimated cut-point. Within each of the aforementioned scenarios, the CI for the cut-point for each of the investigated methods is subsequently evaluated by computing coverage probability and mean length. The coverage probability is the proportion of times that the bootstrap confidence interval contains the true cut-point, while the mean length is a measure of the precision of the confidence interval around the estimated cut-

point. The wider is the confidence interval, the higher is the uncertainty related to the estimated cut-point.

3.3.4 SIMULATION RESULTS

We compare the performance of the Youden index (3.11), concordance probability (3.12) and point closest-to-(0,1) corner in the ROC plane (3.13) methods, in the estimation of the optimal cut-point. We consider the normal homoscedastic scenario with balanced and non-balanced designs, where all the investigated methods identify theoretically the same true cut-point c_{opt}. The Gamma case is also considered. All simulations are performed under the same parametric scenarios introduced above.

The results of the design with a Gaussian biomarker with a cases and controls fraction of 50% are shown in Tables 3.1, 3.2 and 3.3 for different censoring levels, i.e. 12%, 25% and 38%, respectively. The relative bias of the investigated methods is small on all levels of classification accuracy and it increases as the censoring level increases. By comparing the MSEs, it can be noticed that the point closest-to-(0,1) corner in the ROC plane and concordance probability methods have better performance than the Youden index method. Indeed, the MSE is inversely related to sample size and it increases as the censoring level increases. The performance of the investigated methods improves with increasing classification accuracy.

Table 3.1 Relative Bias and Mean Square Error (MSE) of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Normal scenario: cases and controls fraction of 50%, censoring level of 12%[†].

Cases fraction=Controls fraction=50%, Censoring level=12%				**Youden Index**		**Concordance probability**		**Point closest-to-(0,1) corner**	
$J(c_{opt})$ [‡]	**$CZ(c_{opt})$ [‡]**	**c_{opt}**	**N**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**
0.2	**0.36**	0.25	100	0.1168	0.2226	0.0687	0.0719	0.0769	0.0513
			200	0.0696	0.1750	0.0546	0.0501	0.0589	0.0360
			400	0.0813	0.1137	0.0507	0.0273	0.0404	0.0199
0.4	**0.49**	0.52	100	0.0584	0.1295	0.0346	0.0724	0.0356	0.0462
			200	0.0038	0.0870	0.0009	0.0462	0.0032	0.0279
			400	0.0073	0.0536	0.0070	0.0270	0.0100	0.0147
0.6	**0.64**	0.84	100	0.0311	0.0884	0.0350	0.0675	0.0259	0.0421
			200	0.0147	0.0575	0.0201	0.0408	0.0152	0.0227
			400	0.0086	0.0332	0.0015	0.0240	-0.0009	0.0138
0.8	**0.81**	1.28	100	0.0399	0.0727	0.0379	0.0669	0.0291	0.0465
			200	0.0167	0.0435	0.0149	0.0383	0.0115	0.0241
			400	0.0033	0.0290	0.0021	0.0253	0.0027	0.0148

[†] $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, 1)$, $X_{Z>\tau}\sim N(0,1)$. [‡]The levels of J and CZ are achieved by $\mu_{Z\leq\tau}$= 0.51, 1.05, 1.68, 2.56, respectively.

Table 3.2 Relative Bias and Mean Square Error (MSE) of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Normal scenario: cases and controls fraction of 50%, censoring level of 25%[†].

Cases fraction=Controls fraction=50%, Censoring level=25%				**Youden Index**		**Concordance probability**		**Point closest-to-(0,1) corner**	
$J(c_{opt})$ [‡]	**$CZ(c_{opt})$ [‡]**	**c_{opt}**	**N**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**
0.2	**0.36**	0.25	100	0.1172	0.2448	0.1078	0.0798	0.1016	0.0565
			200	0.0777	0.1951	0.0796	0.0525	0.0561	0.0388
			400	0.1123	0.1196	0.0541	0.0304	0.0408	0.0214
0.4	**0.49**	0.52	100	0.0490	0.1332	0.0327	0.0771	0.0339	0.0489
			200	0.0200	0.0946	0.0101	0.0522	0.0031	0.0294
			400	0.0042	0.0573	0.0037	0.0274	0.0075	0.0157
0.6	**0.64**	0.84	100	0.0491	0.0947	0.0428	0.0428	0.0355	0.0355
			200	0.0185	0.0605	0.0237	0.0447	0.0192	0.0237
			400	0.0058	0.0367	0.0055	0.0255	0.0004	0.0148
0.8	**0.81**	1.28	100	0.0384	0.0786	0.0373	0.0722	0.0299	0.0505
			200	0.0149	0.0464	0.0140	0.0401	0.0081	0.0259
			400	0.0053	0.0314	0.0064	0.0271	0.0024	0.0158

[†] $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, 1)$, $X_{Z>\tau}\sim N(0,1)$. [‡]The levels of J and CZ are achieved by $\mu_{Z\leq\tau}$= 0.51, 1.05, 1.68, 2.56, respectively.

Table 3.3 Relative Bias and Mean Square Error (MSE) of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Normal scenario: cases and controls fraction of 50%, censoring level of 38%[†].

Cases fraction=Controls fraction=50%, Censoring level=38%				**Youden Index**		**Concordance probability**		**Point closest-to-(0,1) corner**	
$J(c_{opt})$ [‡]	**$CZ(c_{opt})$ [‡]**	**c_{opt}**	**N**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**
0.2	**0.36**	0.25	100	0.1347	0.2592	0.1042	0.0894	0.1225	0.0627
			200	0.0547	0.2084	0.0638	0.0572	0.0652	0.0413
			400	0.0932	0.1395	0.0177	0.0350	0.0200	0.0245
0.4	**0.49**	0.52	100	0.0595	0.1418	0.0460	0.0841	0.0438	0.0518
			200	0.0106	0.1039	0.0082	0.0541	0.0024	0.0339
			400	0.0085	0.0641	-0.0021	0.0314	0.0009	0.0173
0.6	**0.64**	0.84	100	0.0628	0.1094	0.0599	0.0794	0.0492	0.0494
			200	0.0258	0.0653	0.0256	0.0492	0.0212	0.0260
			400	0.0042	0.0408	0.0052	0.0287	0.0012	0.0160
0.8	**0.81**	1.28	100	0.0449	0.0883	0.0411	0.0792	0.0282	0.0587
			200	0.0132	0.0503	0.0108	0.0462	0.0073	0.0297
			400	0.0039	0.0367	0.0041	0.0323	0.0030	0.0185

[†] $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, 1)$, $X_{Z>\tau}\sim N(0,1)$. [‡]The levels of J and CZ are achieved by $\mu_{Z\leq\tau}$= 0.51, 1.05, 1.68, 2.56, respectively.

Table 3.4 shows the results of the Normal case when considering different cases fractions and a censoring level of 25%. The relative bias of the investigated methods is small on all levels of classification accuracy, except for the scenario with a low classification accuracy (J=0.2 and CZ=0.36). As above, the point closest-to-(0,1) corner in the ROC plane and concordance probability methods outperform the Youden index method. The MSE is lower for the point closest-to-(0,1) corner in the ROC plane method, too.

Table 3.5 shows the bootstrap standard deviation, coverage probability and mean length of the 95% bootstrap CI for the cut-point for the Normal scenario considering different cases fractions and censoring levels. The SD_B of the point closest-to-(0,1) corner in the ROC plane approach is lower than the SD_B of the Youden index and concordance probability methods. Coverage probabilities are close to the nominal level. 95% bootstrap CIs are narrower when considering the scenarios with better classification accuracies, i.e. J of 0.6 and 0.8.

The results of the Gamma distribution scenario considering a cases and controls fraction of 50% and a censoring level of 25% are shown in Table 3.6. It can be noticed that the theoretical true cut-points c_J, c_{CZ} and c_{ER} are all different (please also see Figure 2.7 page 26). For what regards the relative performance of the investigated methods, we note a negligible relative bias in the estimate of the optimal cut-point. As above, the concordance probability and the point closest-to-(0,1) corner in the ROC plane methods outperform the Youden index method. The MSE is inversely related to sample size and it is lower for the point closest-to-(0,1) corner in the ROC plane estimator. Moreover, it seems that the Youden index, concordance probability and point closest-to-(0,1) corner in the ROC plane methods perform better when considering low classification accuracy scenarios, i.e. J of 0.2 and 0.4.

Table 3.4 Relative Bias and Mean Square Error (MSE) of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Normal scenario: different cases fractions, censoring level of 25%†.

Different cases fractions, Censoring level=25%					**Youden Index**		**Concordance probability**		**Point closest-to-(0,1) corner**	
$J(c_{opt})$ ‡	**$CZ(c_{opt})$ ‡**	**c_{opt}**	**N**	**Cases fraction**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**	**Relative Bias**	**MSE**
0.2	**0.36**	0.25	150	33%	0.2188	0.2089	0.1921	0.0611	0.1749	0.0440
			200	25%	0.0973	0.2125	0.1758	0.0666	0.1822	0.0482
			250	20%	0.1872	0.1871	0.2246	0.0592	0.2164	0.0421
0.4	**0.49**	0.52	150	33%	0.0786	0.1048	0.0625	0.0575	0.0741	0.0370
			200	25%	0.0516	0.1147	0.0701	0.0606	0.0674	0.0381
			250	20%	0.0926	0.0990	0.0951	0.0528	0.0940	0.0346
0.6	**0.64**	0.84	150	33%	0.0522	0.0778	0.0547	0.0590	0.0453	0.0367
			200	25%	0.0500	0.0744	0.0516	0.0512	0.0515	0.0286
			250	20%	0.0543	0.0742	0.0585	0.0569	0.0596	0.0344
0.8	**0.81**	1.28	150	33%	0.0401	0.0702	0.0439	0.0645	0.0387	0.0432
			200	25%	0.0360	0.0568	0.0375	0.0513	0.0438	0.0364
			250	20%	0.0535	0.0641	0.0558	0.0585	0.0535	0.0390

† $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, 1)$, $X_{Z>\tau}\sim N(0,1)$. ‡The levels of J and CZ are achieved by $\mu_{Z\leq\tau}$= 0.51, 1.05, 1.68, 2.56, respectively.

Table 3.5 Bootstrap standard deviation, coverage probability and mean length of the 95% confidence interval estimation of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Normal scenario: different cases fractions and censoring levels[†].

					Youden Index			Concordance probability			Point closest-to-(0,1) corner		
$J(c_{opt})$ [‡]	$CZ(c_{opt})$ [‡]	N	Cases fraction	Censoring level	SD_B	Coverage	Mean Length	SD_B	Coverage	Mean Length	SD_B	Coverage	Mean Length
0.2	**0.36**	100	50%	12%	0.4606	0.968	1.6188	0.2844	0.964	1.0068	0.2409	0.960	0.8508
		100	50%	38%	0.4758	0.966	1.6586	0.3086	0.956	1.0881	0.2589	0.954	0.9158
		150	33%	25%	0.4330	0.961	1.5257	0.2609	0.930	0.9143	0.2188	0.923	0.7741
0.4	**0.49**	100	50%	12%	0.3504	0.961	1.2307	0.2677	0.956	0.9371	0.2158	0.943	0.7626
		100	50%	38%	0.3673	0.968	1.2888	0.2860	0.948	0.9917	0.2316	0.948	0.8178
		150	33%	25%	0.3235	0.954	1.1433	0.2448	0.938	0.8616	0.1961	0.927	0.6941
0.6	**0.64**	100	50%	12%	0.2876	0.958	0.9936	0.2552	0.954	0.8880	0.2064	0.938	0.7228
		100	50%	38%	0.3006	0.953	1.0473	0.2689	0.943	0.9386	0.2212	0.939	0.7712
		150	33%	25%	0.2699	0.939	0.9309	0.2371	0.928	0.8177	0.1876	0.917	0.6569
0.8	**0.81**	100	50%	12%	0.2565	0.920	0.8753	0.2469	0.918	0.8405	0.2110	0.905	0.7287
		100	50%	38%	0.2725	0.917	0.9324	0.2641	0.916	0.9024	0.2375	0.908	0.8307
		150	33%	25%	0.2461	0.878	0.8350	0.2362	0.880	0.8030	0.1967	0.864	0.6756

[†] $X_{Z\leq\tau}\sim N(\mu_{Z\leq\tau}, 1)$, $X_{Z>\tau}\sim N(0,1)$. [‡]The levels of J and CZ are achieved by $\mu_{Z\leq\tau}$= 0.51, 1.05, 1.68, 2.56, respectively.

Table 3.6 Relative Bias and Mean Square Error (MSE) of the Youden index, concordance probability and point closest-to-(0,1)-corner in the ROC plane estimators. Gamma scenario: cases and controls fraction of 50%, censoring level of 25%†.

Cases fraction=Controls fraction=50%, Censoring level=25%					Sample sizes	Youden Index		Concordance probability		Point closest-to-(0,1) corner	
$J(c_{opt})$ ‡	$CZ(c_{opt})$ ‡	c_J	c_{CZ}	c_{ER}	N	Relative Bias	MSE	Relative Bias	MSE	Relative Bias	MSE
0.2	**0.36**	1.12	1.35	1.38	100	0.1605	0.2987	0.0429	0.0975	0.0352	0.0712
					200	0.0840	0.1825	0.0138	0.0638	0.0087	0.0446
					400	0.0668	0.1156	0.0105	0.0348	0.0086	0.0261
0.4	**0.49**	1.79	1.81	1.82	100	0.0460	0.2585	0.0408	0.1529	0.0311	0.1001
					200	0.0267	0.1740	0.0136	0.0924	0.0059	0.0521
					400	0.0144	0.1183	0.0133	0.0571	0.0062	0.0346
0.6	**0.64**	2.45	2.41	2.36	100	0.0297	0.3288	0.0293	0.2380	0.0267	0.1536
					200	0.0186	0.2117	0.0170	0.1539	0.0138	0.0815
					400	0.0067	0.1433	0.0077	0.0941	0.0081	0.0445
0.8	**0.81**	3.42	3.38	3.24	100	0.0526	0.5678	0.0508	0.5156	0.0429	0.3593
					200	0.0210	0.3317	0.0241	0.3039	0.0166	0.1739
					400	0.0090	0.1929	0.0107	0.1693	0.0079	0.0885

† $X_{Z\leq\tau}\sim G(2.5, \beta_{Z\leq\tau})$, $X_{Z>\tau}\sim G(1.5,1)$. ‡The levels of J and CZ are achieved by $\beta_{Z\leq\tau}$= 0.79, 1.22, 1.97, 3.82, respectively.

3.4 THE CUT-POINT FINDING FOR THE CRLF2 IN ACUTE LYMPHOBLASTIC LEUKEMIA

Acute lymphoblastic leukemia (ALL) is the most common malignancy in children in developed countries and it represents a highly aggressive disease in all age groups [34]. B-cell precursor (BCP) ALL accounts for approximately 70% of childhood ALL. The cure rate of BCP-ALL is higher than 80% [35], but the probability of survival of patients who relapse is only 40%.

The identification of prognostic factors remains a formidable challenge in BCP-ALL. Recently, chromosomal translocations were identified as promising key factors in the pathogenesis of ALL. However, there is the real need to identify which of the recently discovered genetic alterations can improve patient's stratification for the development of targeted therapeutic approaches [35].

Recently, two studies [36, 37] showed that the overexpression of the *cytokine receptor-like factor 2* (CRLF2) gene was found to be correlated with poor prognosis in BCP-ALL patients.

Here, we discuss the cut-point finding task of the CRLF2 biomarker at diagnosis in 464 Italian BCP-ALL children enrolled from February 2003 to July 2005 and treated according to the protocol "AIEOP-BFM ALL2000" of the "*Associazione Italiana Ematologia Oncologia Pediatrica* (AIEOP)" [35]. We aimed to find a cut-point for the CRLF2 above which children can be considered at higher risk of relapse, and thus candidates for treatment intensification.

Briefly, within this cohort median age at diagnosis is 4.5 years (range, 1-17 years). Of the 464 enrolled children, 137 (28%) were at standard risk, 300 at intermediate risk (65%),

and the remaining 35 children (7%) were at higher risk according to the standard criteria defined in the protocol. Event free survival (EFS) was calculated from the date of diagnosis to the date of event, i.e. resistance, relapse, death or second malignant neoplasm, whichever occurred first. A total of 89 events were observed, of which 79 were relapses. Figure 3.2 shows the overall EFS Kaplan-Meier curve. Within this cohort [35], we can see that 5 years EFS is 81.6% (SE=1.8%).

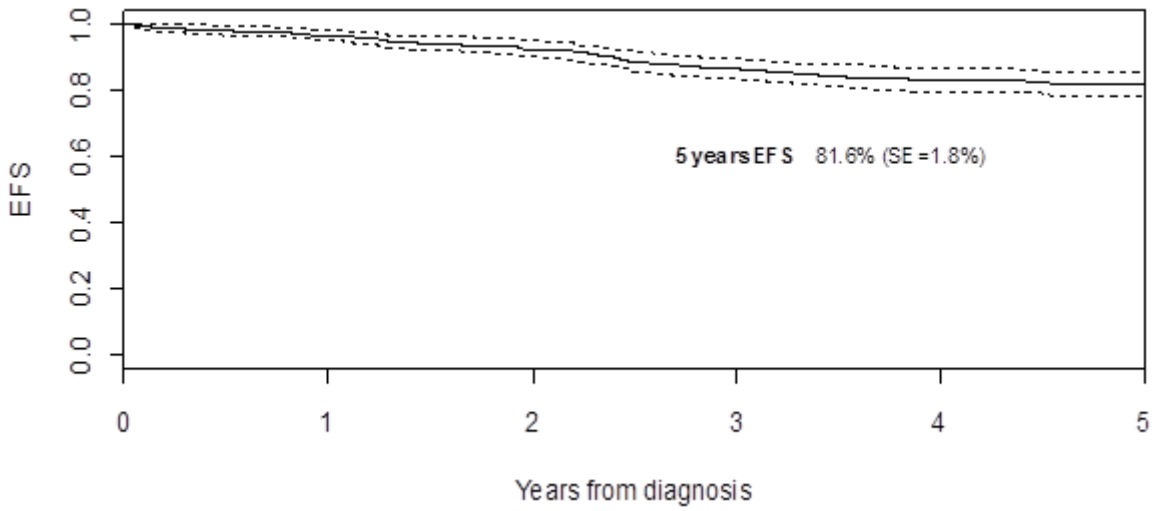

Figure 3.2 Kaplan-Meier EFS curve along with 95% confidence bands (dashed lines) of 464 BCP-ALL children [35].

CRLF2 expression, evaluated by real-time quantitative (RQ)-PCR, ranged from 0.006 to 810-fold change (Figure 3.3). The CRLF2 distribution is not Gaussian, and the Shapiro-Wilk non-parametric Normal distribution test returns a P-value less than 0.0001. The distribution of the CRLF2 is skewed to the right, and it can be approximated by a Gamma model.

A cut-point c is needed to identify children who are more likely to relapse and for whom treatment is thus not regarded as sufficient to provide a good disease control and

needs to be optimized or intensified in future protocols. To this aim, the three investigated methods, i.e. Youden index (3.11), concordance probability (3.12) and point closest-to-(0,1) corner in the ROC plane (3.13), have been applied within this failure time outcome scenario by defining the meaningful time point for the clinical problem τ equal to 5 years from diagnosis.

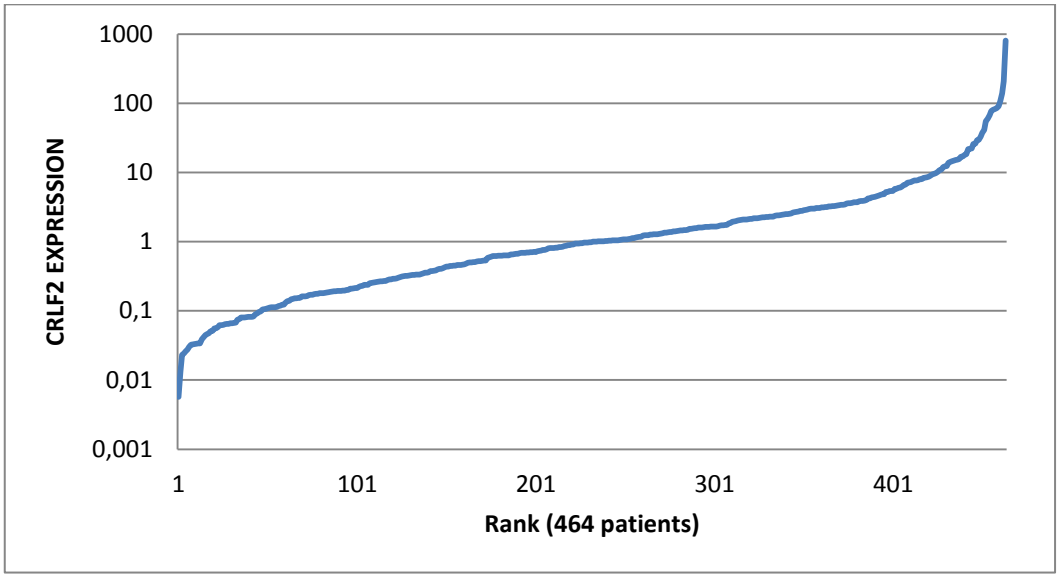

Figure 3.3 CRLF2 expression in 464 BCP-ALL patients [35].

The three investigated methods lead to the same estimated cut-point $\hat{c}$=1.46 (Table 3.7). We also estimate the 95% bootstrap CI estimate for the cut-point (Table 3.7), finding for the Youden index, concordance probability and point closest-to-(0,1) corner in the ROC plane methods the following 95% CI estimates: (0.12, 21.61), (0.70, 1.98) and (0.70, 1.86), respectively. The Youden index method achieves the largest 95% bootstrap CI for the cut-point estimate. However, this is consistent with the simulation results presented in paragraph 3.3.4 page 59, where we showed that the Youden index method achieves the largest confidence interval for the cut-point estimate.

Table 3.7 Estimated cut-point along with 95% bootstrap CI for the CRLF2 biomarker [35].

Method of cut-point finding	Estimated cut-point	95% Bootstrap confidence interval
Youden index [14]	1.46	0.12 - 21.61
Concordance probability [15]	1.46	0.70 - 1.98
Point closest-to-(0,1) corner in the ROC plane [18]	1.46	0.70 - 1.86

It should be noted that the TPF (3.5) and the FPF (3.6) computed at the estimated cut-point $\hat{c}$=1.46 are 0.47 and 0.37, respectively. Indeed, the classification accuracy of the CRLF2 biomarker is not so high. In fact, it can be easily derived that the Youden index (3.11) is $J(\hat{c}=1.46)=0.47-0.37=0.10$, while the concordance probability is equal to $CZ(\hat{c}=1.46)=0.47\cdot(1-0.37)=0.29$. However, this low discrimination potential emerges even better from the ROC curve of that lies just above the bisecting line (Figure 3.4).

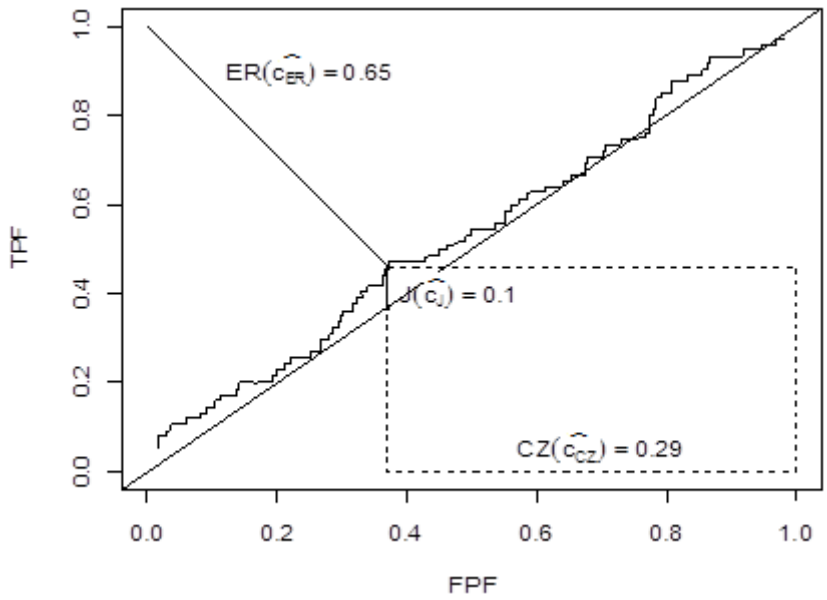

Figure 3.4 ROC curve for the CRLF2 [35] with the three objective functions J, CZ and ER.

The Kaplan-Meier EFS curves stratified according to the estimated CRLF2 cut-point $\hat{c}$=1.46 are shown in Figure 3.5. It can be noted that the 5 years EFS of 284 children

presenting with CRLF2 values greater than the estimated cut-point $\hat{c}$=1.46 is 79.2% (SE=3.1%), while the 5 years EFS of 180 children presenting with CRLF2 biomarker value lower than the estimated cut-point $\hat{c}$=1.46 is 83.2% (SE=2.2%). The two survival curves are not statistically significant different (Log-rank P=0.45).

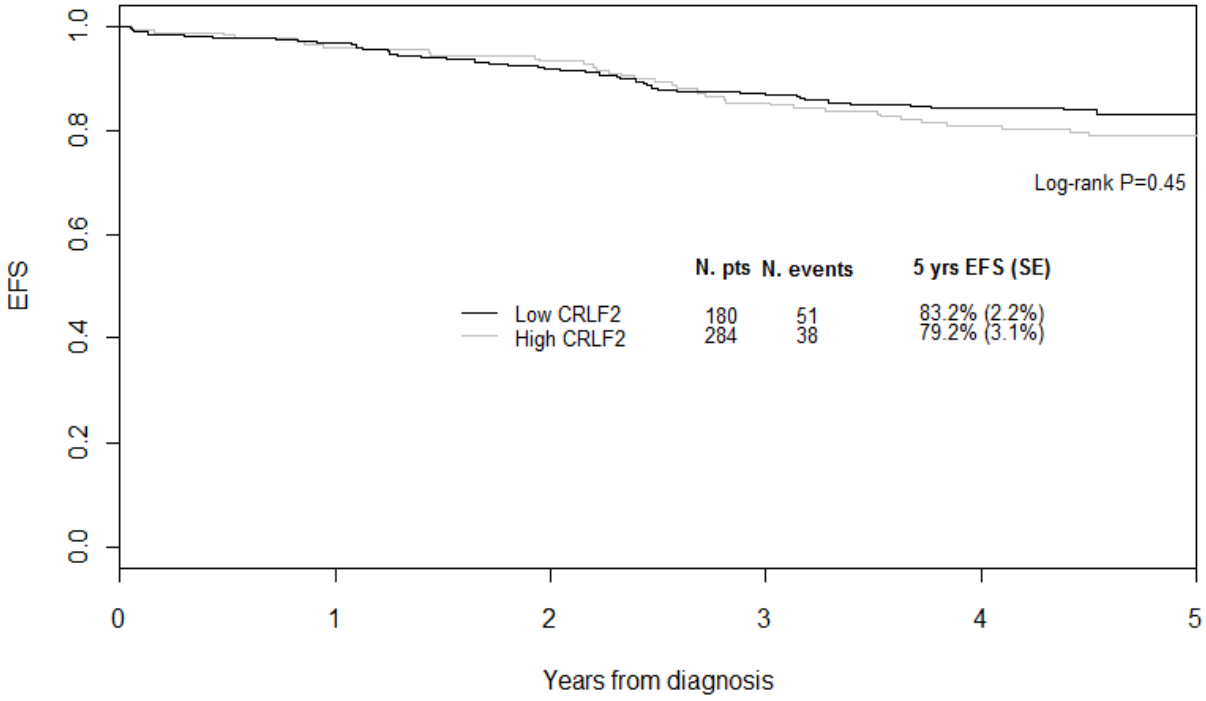

Figure 3.5 EFS curve stratified according to the estimated CRLF2 cut-point $\hat{c}$=1.46.

3.4.1 THE CRLF2 CUT-POINT FROM THE PREDICTIVE VALUES PERSPECTIVE

The cut-point for the CRLF2 biomarker identified above is not satisfactory since it does not discriminate the outcome as shown by the EFS curves in Figure 3.5. The investigated methods, i.e. Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18], are based on measures defined conditionally on the true disease status, i.e. sensitivity and specificity. These measures of test performance provide

the type of information typically needed for health policy purposes, but are not the most useful information for clinical decisions on treatments [38]. In fact, clinicians have to interpret test results on the basis of tested people [39], and so they typically need to know the predictive values of the test. The positive predictive value (PPV) represents the proportion of diseased subjects having the biomarker value greater than c, i.e. $\mathrm{PPV}(c) = \mathrm{P}(Z \leq \tau | X > c) = 1 - S_{X>c}(\tau)$, while the negative predictive value (NPV) represents the proportion of non-diseased subjects having the biomarker values less or equal than c, i.e. $\mathrm{NPV}(c) = \mathrm{P}(Z > \tau | X \leq c) = 1 - S_{X \leq c}(\tau)$.

For this reason, we explored the application of the Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18] methods to the CRLF2 cut-point finding task from the predictive values perspective. Briefly, within the three objective functions (3.11), (3.12) and (3.13), the two quantities $S_{Z \leq \tau}(c)$ and $S_{Z>\tau}(c)$ are replaced respectively by $1 - S_{X>c}(\tau)$ and $1 - S_{X \leq c}(\tau)$ (page 52). The estimated cut-points for the CRLF2 along with the 95% bootstrap CI computed by using the three investigated cut-point finding methods from the predictive values perspective are shown in Table 3.8. The bootstrap 95% CI estimate for the cut-point for the Youden index, concordance probability and point closest-to-(0,1) corner in the ROC plane methods are (0.08, 26.09), (0.18, 26.09) and (0.21, 26.09), respectively. As previously noted, the Youden index method achieves the largest 95% bootstrap CI for the cut-point estimate.

Moreover, the true positive fraction (3.5) and the false positive fraction (3.6) computed at the estimated cut-point $\hat{c}_{\mathrm{PRED}}$=17.89 are 0.10 and 0.04 (i.e. specificity equal to 0.96), while the positive predictive and the negative predictive values are 0.35 and 0.84, respectively.

Table 3.8 The CRLF2 [35] estimated cut-point and 95% bootstrap CI computed from the predictive values perspective.

Method of cut-point finding	Estimated cut-point	95% Bootstrap confidence interval
Youden index [14]	17.89	0.08 – 26.09
Concordance probability [15]	17.89	0.18 - 26.09
Point closest-to-(0,1) corner in the ROC plane [18]	17.89	0.21 – 26.09

Figure 3.6 shows the Kaplan-Meier EFS curves stratified according to the estimated CRLF2 cut-point $\hat{c}_{PRED}$=17.89. The 5 years EFS of 23 children presenting with CRLF2 values greater than the estimated cut-point $\hat{c}_{PRED}$=17.89 is 64.6% (SE=10.1%), while the 5 years EFS of 441 children presenting with CRLF2 biomarker value lower than the estimated cut-point $\hat{c}_{PRED}$=17.89 is 82.5% (SE=1.8%). The two survival curves are statistically significant different (Log-rank P=0.07).

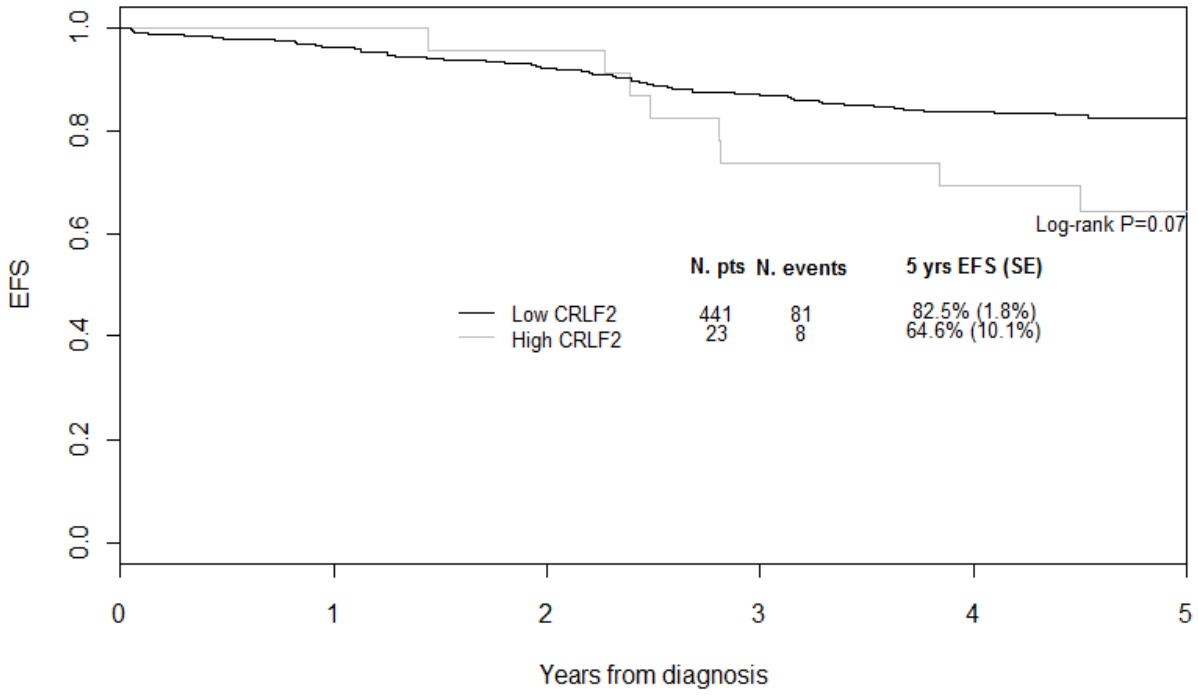

Figure 3.6 EFS curve stratified according to the estimated CRLF2 cut-point $\hat{c}_{PRED}$=17.89.

Summarizing, in this clinical application, the use of cut-point finding methods based on predictive values allows to identify a small group of patients (n=23) with higher expression of the CRLF2 biomarker who have a significantly lower long term EFS. However, it may be discussed whether such a rare subgroup is of clinical relevance.

CHAPTER 4

DISCUSSION AND CONCLUSIONS

4

4.1 DISCUSSION

In this PhD dissertation, we presented a theoretical investigation and a simulation study aimed to compare four methods commonly used to define cut-points of continuous biomarkers: the minimum P-value [13], the Youden index [14], the concordance probability [15] and the point closest-to-(0,1) corner in the ROC plane [18]. We addressed both the binary (Chapter 2) and possibly censored failure time outcome (Chapter 3) scenarios. We also presented some original applications to real datasets.

Here we are going to comment the theoretical results obtained. We proved that the considered methods are mathematically related, since the Youden function is the square root of the numerator of the Chi-square function of the minimum P-value approach. However, these methods do not necessarily identify the same optimal cut-point when considering the analysed scenarios of Gaussian and Gamma biomarker distributions. The equality between the underlying true cut-points is observed for the three ROC-based methods (i.e., Youden index, concordance probability and point closest-to-(0,1) corner) in the case of Gaussian homoscedastic distributions of the biomarker [15]. However, even

under these conditions, the equality of these common true underlying cut-points to that of the minimum P-value approach is not guaranteed unless we are in a balanced design.

We have shown that the point closest-to-(0,1) corner in the ROC plane [18] and concordance probability [15] approaches perform well and outperform both the minimum P-value [13] and Youden index [14] methods in the estimation of the cut-point. In fact, while the relative bias is essentially zero for the investigated methods in all the considered scenarios, the MSEs of the point closest-to-(0,1) corner in the ROC plane and concordance probability approaches are similar, but lower than the MSEs of the minimum P-value [13] and Youden index [14] methods, considering all levels of classification accuracy. As expected, MSE is inversely related to sample size. Moreover, we also provided the estimate of the standard deviation and the 95% confidence interval for the cut-point through the bootstrap resampling technique. From this result, we can conclude that the bootstrap 95% confidence interval estimate of the cut-point for the point closest-to-(0,1) corner in the ROC plane method is systematically narrower than the bootstrap confidence interval estimate of the other methods.

More in general, the observed difference in the MSE behaviour between the minimum P-value approach [13] and the three ROC-based methods [14] is due to the variance component included in the denominator of the Chi-square statistic of the minimum P-value approach. This variance component is computed under the null hypothesis of absence of association between the true disease status and the biomarker. However, the identification of a cut-point is based on the possible presence of a discrimination potential of the biomarker, i.e. the alternative hypothesis. For this reason, this variance component included

in the denominator of the Chi-square statistic does not provide an advantage for the performance of the minimum P-value approach.

We also noted, within the Gaussian non-balanced scenario, that the true cut-point underlying the minimum P-value approach varies with the disease prevalence in the sample. In particular, a reduction of the sample disease prevalence determines a systematic shift of c_{χ^2} towards the distribution of disease subjects, while c_{ROC} is unaffected. This is a key limitation of the minimum P-value approach [13] in cut-point finding.

It was noted in Böhning et al. [20] that the Youden function is invariant with respect to small changes in the cut-point. While this is a positive characteristic of the Youden index in a meta-analysis setting [40], this is not a good behaviour in a cut-point finding scenario. This additive invariance property does not hold for the concordance probability function, making this last criterion more suitable for cut-point finding. Perkins and Schisterman [18] advocated for the use of the Youden index method after comparison with the point closest-to-(0,1) corner in the ROC plane criterion for its sound intuitive clinical meaning, while Böhning et al. [20] recently showed that the Youden index method should be preferred to the diagnostic odds ratio approach [19]. Our investigation completes these series of evidences [15, 21] by addressing the minimum P-value approach [13].

In light of our results, in both scenarios (dichotomous and failure time censored outcome), the point closest-to-(0,1) corner in the ROC plane and concordance probability methods could be considered the best criteria to estimate the cut-point of a biomarker from the performance point of view. However, since the ROC-based methods identify the same underlying true cut-point in the Gaussian homoscedastic scenario, the confidence interval estimation of the cut-point through the point closest-to-(0,1) corner method in the ROC

plane could be followed by the calculation of the associated value of the Youden function and/or concordance probability. This could improve communicability in the applied context. In fact, the Youden index has a simple interpretation as the net gain of the true positive fraction accounting for the false positive fraction, while the classification accuracy as the probability of being below or beyond the cut-point for any random pair of non-diseased and diseased subjects. This motivates an ever-increasing body of literature [21, 41], comprising methods for constructing the confidence interval of the Youden index and its corresponding optimal cut-point [42]. However, in light of these works, we believe that more emphasis should be given on methods showing improvements in the performance of the cut-point finding estimator, such as the point closest-to-(0,1) corner in the ROC plane.

4.2 THE CUT-POINT FINDING FROM THE PREDICTIVE VALUES PERSPECTIVE

Within this PhD dissertation, we dealt with cut-point finding methods on the basis of measures, i.e. sensitivity and specificity, defined conditionally on the true disease status. The Chi square statistic of the minimum P-value approach [13] is also a function of sensitivity and specificity since it is based on the null hypothesis of equality between sensitivity and 1 minus specificity (i.e. false positive fraction). These measures of test performance provide the type of information typically needed for health policy purposes, but are not necessarily the most useful information for clinical decisions on treatments [38]. In fact, clinicians have to interpret test results on the basis of tested people [39], and so they typically need to know the predictive values of the test. The positive predictive value (PPV) represents the proportion of diseased subjects having the biomarker value greater than c, while the

negative predictive value (NPV) represents the proportion of non-diseased subjects having the biomarker values less or equal than c.

The three investigated ROC-based methods, i.e. Youden index [14], concordance probability [15] and point closest-to-(0,1) corner in the ROC plane [18], could be applied starting from predictive values. The minimum P-value approach [13] can also be applied since it is based on the null hypothesis of equality between the PPV and 1-NPV. However, while sensitivity and specificity are invariant to disease prevalence, predictive values vary across populations with different disease prevalence. So, the PPV and NPV are functions of both the cut-point c and the disease prevalence. In particular, when used in low prevalence settings, even excellent diagnostic tests have poor PPV, while within high prevalence settings, the NPV is nearly perfect [39]. In a such way, the Youden index [14], the concordance probability [15] and the point closest-to-(0,1) corner in the ROC plane [18] methods depend also on disease prevalence. Moreover, the use of cut-point finding methods based on predictive values might easily lead to the choice of threshold values on the boundary of the parameter range of the biomarker [20], as previously shown for the diagnostic odds ratio approach in Figure 2.3, page 10. In fact, when we performed the cut-point finding task of the CRLF2 in acute lymphoblastic leukemia by using the three investigated methods starting from the predictive values, we labelled as optimal a cut-point value corresponding to the 97th percentile of the biomarker distribution. Such cut-point identifies a very small group of patients who are at higher risk of relapse, and it may be discussed whether such a rare subgroup is of clinical relevance.

4.3 CONCLUSIONS

More in general, great care should be taken when establishing a biomarker cut-point for clinical use. Methods for categorizing new biomarkers are often essential in clinical decision-making even if categorization of a continuous biomarker is gained at a considerable loss of power and information [43]. In the future, new methods involving the study of the functional form between the biomarker value and the outcome through regression techniques such as fractional polynomials or spline functions should be considered to alternatively define cut-points for clinical use. Moreover, in spite of the two aforementioned drawbacks related to the use of predictive values in cut-point finding, we also think that additional new methods for cut-point finding should be developed starting from predictive values.

Papers specifically related to the work done within my PhD program are an applicative contribution in the electrocardiology field [23] and two methodological papers, one of which submitted to Computational Statistics and Data Analysis dealing with cut-point finding for a dichotomous outcome (“Finding the Optimal Cut-Point of a Continuous Biomarker: which method is better?”) and one currently under preparation dealing with the cut-point finding for censored failure time data.

This PhD dissertation was discussed on 28 January 2013 at Milano-Bicocca University.
The author detects the copyright of this work. No parts of this work can be copied or reproduced without his permission.

APPENDIX 1

R code for simulation: dichotomous outcome scenario

```
set.seed(5091985)
samples<-1000
nd<-50
nnd<-50
N<-nd+nnd
ud<-0.51          # ud values determine different classification accuracy values
und<-0
truec<-(ud+und)/2
CHIhat<-matrix(nrow=N-0.20*N,ncol=2)
Jhat<-matrix(nrow=N-0.20*N,ncol=2)
CZhat<-matrix(nrow=N-0.20*N,ncol=2)
ERhat<-matrix(nrow=N-0.20*N,ncol=2)
cutpoints<-matrix(nrow=samples,ncol=9)

for (ns in 1:samples) {

# Biomarker data generation
Xd<-rnorm(nd,mean=ud,sd=1)
Xnd<-rnorm(nnd,mean=und,sd=1)
data<-data.frame(rbind(cbind(Xd, rep(1,nd)),cbind(Xnd, rep(0,nnd))))
names(data)[names(data)=="Xd"] = "x"
```

```
names(data)[names(data)=="V2"] = "d"

data<-data[order(data[,1],data[,1]),]


# Cut-points finding process for the ns-th sample

index<-1

for(j in (N*0.10+1):(N-N*0.10)) {


c<-data[j,1]

n11<-sum(data$x<=c & data$d==0)

n12<-sum(data$x>c & data$d==0)

n21<-sum(data$x<=c & data$d==1)

n22<-sum(data$x>c & data$d==1)


# Estimation of the Chi-square objective  function (2.5) at cut-point c

CHIhat[index,1]<-c

CHIhat[index,2]<-(N*(n11*n22-
n12*n21)^2)/(sum(data$d==0)*sum(data$d==1)*(n11+n21)*(n12+n22))


# Estimation of the Youden objective function (2.8) at cut-point c

Jhat[index,1]<-c

Jhat[index,2]<-n22/sum(data$d==1)-n12/sum(data$d==0)


# Estimation of the concordance probability objective function (2.9) at cut-point c

CZhat[index,1]<-c

CZhat[index,2]<-(n22/sum(data$d==1))*(1-(n12/sum(data$d==0)))
```

```
# Estimation of the distance to the ideal marker objective function (2.10) at cut-point c
ERhat[index,1]<-c
ERhat[index,2]<-sqrt(((n12/sum(data$d==0))-0)^2+((n22/sum(data$d==1))-1)^2)

index<-index+1
}

CHIhat<-CHIhat[order(CHIhat[,2]),]
Jhat<-Jhat[order(Jhat[,2]),]
CZhat<-CZhat[order(CZhat[,2]),]
ERhat<-ERhat[order(ERhat[,2]),]

# Selection of the optimal cut-point for the ns-th sample
cutpoints[ns,1]<-ns
cutpoints[ns,2]<-CHIhat[nrow(CHIhat),1]
cutpoints[ns,3]<-CHIhat[nrow(CHIhat),2]
cutpoints[ns,4]<-Jhat[nrow(Jhat),1]
cutpoints[ns,5]<-Jhat[nrow(Jhat),2]
cutpoints[ns,6]<-CZhat[nrow(CZhat),1]
cutpoints[ns,7]<-CZhat[nrow(CZhat),2]
cutpoints[ns,8]<-ERhat[1,1]
cutpoints[ns,9]<-ERhat[1,2]
}

#Relative bias calculation
BiasCHI<-mean((cutpoints[,2]-truec)/truec)
```

```
BiasJ<-mean((cutpoints[,4]-truec)/truec)

BiasCZ<-mean((cutpoints[,6]-truec)/truec)

BiasER<-mean((cutpoints[,8]-truec)/truec)


#Mean Square Error calculation

MSECHI<-mean((cutpoints[,2]-truec)^2)

MSEJ<-mean((cutpoints[,4]-truec)^2)

MSECZ<-mean((cutpoints[,6]-truec)^2)

MSEER<-mean((cutpoints[,8]-truec)^2)
```

APPENDIX 2

R code for bootstrap resampling

```
set.seed(5091985)
samples<-1000
nd<-50
nnd<-50
N<-nd+nnd
B<-200
ud<-0.51          # ud values determine different classification accuracy values
und<-0
truec<-(ud+und)/2
CHIhat<-matrix(nrow=N-0.20*N,ncol=2)
Jhat<-matrix(nrow=N-0.20*N,ncol=2)
CZhat<-matrix(nrow=N-0.20*N,ncol=2)
ERhat<-matrix(nrow=N-0.20*N,ncol=2)
cutpointsboot<-matrix(nrow=B,ncol=4)
cutpoints<-matrix(nrow=samples,ncol=8)
BDevst<-matrix(nrow=samples,ncol=4)

for (ns in 1:samples) {

# Biomarker data generation
```

```
Xd<-rnorm(nd,mean=ud,sd=1)

Xnd<-rnorm(nnd,mean=und,sd=1)


# Bootstrap resampling technique for the ns-th sample

for (b in 1:B) {

resample_Xd <- sample(Xd, size=nd,replace=T)

resample_Xnd<- sample(Xnd, size=nnd,replace=T)

data<-data.frame(rbind(cbind(resample_Xd,rep(1,nd)),cbind(resample_Xnd,rep(0,nnd))))

names(data)[names(data)=="resample_Xd"] = "x"

names(data)[names(data)=="V2"] = "d"

data<-data[order(data[,1],data[,1]),]


# Cut-points finding process for the b-th bootstrap replication of the ns-th sample

index<-1

for(j in (N*0.10+1):(N-N*0.10)) {


c<-data[j,1]

n11<-sum(data$x<=c & data$d==0)

n12<-sum(data$x>c & data$d==0)

n21<-sum(data$x<=c & data$d==1)

n22<-sum(data$x>c & data$d==1)


# Estimation of the Chi-square function (2.5) at cut-point c

CHIhat[index,1]<-c

CHIhat[index,2]<-(N*(n11*n22-
n12*n21)^2)/(sum(data$d==0)*sum(data$d==1)*(n11+n21)*(n12+n22))
```

```
# Estimation of the Youden function (2.8) at cut-point c
Jhat[index,1]<-c
Jhat[index,2]<-n22/sum(data$d==1)-n12/sum(data$d==0)

# Estimation of the concordance probability function (2.9) at cut-point c
CZhat[index,1]<-c
CZhat[index,2]<-(n22/sum(data$d==1))*(1-(n12/sum(data$d==0)))

# Estimation of the distance to the ideal marker (2.10) at cut-point c
ERhat[index,1]<-c
ERhat[index,2]<-sqrt(((n12/sum(data$d==0))-0)^2+((n22/sum(data$d==1))-1)^2)

index<-index+1
}
CHIhat<-CHIhat[order(CHIhat[,2]),]
Jhat<-Jhat[order(Jhat[,2]),]
CZhat<-CZhat[order(CZhat[,2]),]
ERhat<-ERhat[order(ERhat[,2]),]

cutpointsboot[b,1]<-CHIhat[nrow(CHIhat),1]
cutpointsboot[b,2]<-Jhat[nrow(Jhat),1]
cutpointsboot[b,3]<-CZhat[nrow(CZhat),1]
cutpointsboot[b,4]<-ERhat[1,1]
}
```

```
# Selection of the inferior/superior limit of the bootstrap confidence interval for the ns-th sample
cutpoints[ns,1]<-quantile(cutpointsboot[,1],0.025)
cutpoints [ns,2]<-quantile(cutpointsboot[,1],0.975)
cutpoints [ns,3]<-quantile(cutpointsboot[,2],0.025)
cutpoints [ns,4]<-quantile(cutpointsboot[,2],0.975)
cutpoints [ns,5]<-quantile(cutpointsboot[,3],0.025)
cutpoints [ns,6]<-quantile(cutpointsboot[,3],0.975)
cutpoints [ns,7]<-quantile(cutpointsboot[,4],0.025)
cutpoints [ns,8]<-quantile(cutpointsboot[,4],0.975)

# Bootstrap standard deviation's calculation for the ns-th sample
BDevst[ns,1]<-sd(cutpointsboot[,1])
BDevst[ns,2]<-sd(cutpointsboot[,2])
BDevst[ns,3]<-sd(cutpointsboot[,3])
BDevst[ns,4]<-sd(cutpointsboot[,4])
}

# Bootstrap confidence interval's coverage calculation
CoverageCHI<-sum(ifelse(truec>cutpoints[,1] & truec< cutpoints [,2],1,0))/samples
CoverageJ<-sum(ifelse(truec> cutpoints [,3] & truec< cutpoints [,4],1,0))/samples
CoverageCZ<-sum(ifelse(truec> cutpoints [,5] & truec< cutpoints [,6],1,0))/samples
CoverageER<-sum(ifelse(truec> cutpoints [,7] & truec< cutpoints [,8],1,0))/samples

# Bootstrap confidence interval's length calculation
LengthCHI<-mean(cutpoints [,2]- cutpoints [,1])
LengthJ<-mean(cutpoints [,4]- cutpoints [,3])
```

```
LengthCZ<-mean(cutpoints [,6]- cutpoints [,5])

LengthER<-mean(cutpoints [,8]- cutpoints [,7])


# Bootstrap standard deviation's calculation

DevstCHI<-mean(BDevst[,1])

DevstJ<-mean(BDevst[,2])

DevstCZ<-mean(BDevst[,3])

DevstER<-mean(BDevst[,4])
```

APPENDIX 3

R code for simulation: censored failure time outcome scenario

```
library(survival)

# Kaplan Meier estimation of P(Z > τ|X > c) and P(Z > τ|X ≤ c)

SurvCalculation<-function(Xgreaterc,Xlessc){

    SZlesstau<- survfit(Surv(Xgreaterc[,2], Xgreaterc [,3]) ~1, error=c("greenwood"))

    position<-which(SZlesstau$time>=tau)[1]

    if (is.na(position)==TRUE) ShatZlesstau <- SZlesstau $surv[NROW(SZlesstau $surv)] else if
(position==1 & Xgreaterc [Xgreaterc $t==SZlesstau$time[position],]$delta ==1  ) ShatZlesstau <-1 else
if (Xgreaterc [Xgreaterc $t==SZlesstau$time[position],]$delta ==1 ) ShatZlesstau <- SZlesstau
$surv[position-1] else  ShatZlesstau <- SZlesstau $surv[position]

    SZgreatertau<- survfit(Surv(Xlessc[,2],Xlessc[,3]) ~1, error=c("greenwood"))

     position<-which(SZgreatertau$time>=tau)[1]

     if (is.na(position)==TRUE) ShatZgreatertau<- SZgreatertau $surv[NROW(SZgreatertau $surv)] else
if (position==1 & Xlessc [Xlessc$t==SZgreatertau$time[position],]$delta ==1) ShatZgreatertau<-1 else
if ( Xlessc [Xlessc$t==SZgreatertau$time[position],]$delta ==1) ShatZgreatertau<- SZgreatertau
$surv[position-1]  else ShatZgreatertau<-SZgreatertau$surv[position]

    output<-list(ShatZlesstau,ShatZgreatertau)

    return(output)

}


set.seed(5091985)

samples<-1000

N<-100

ud<-0.51          # ud values determine different classification accuracy values
```

```
und<-0
truec<-(ud+und)/2
lambda<-2
tau<-(-1/lambda)*log(0.50)
mincenslim<-0
maxcenslim<-1            # maxcenslim values determine different censoring percentages
X<-c(N,1)
TPF<-matrix(nrow=N-0.20*N,ncol=1)
FPF<-matrix(nrow=N-0.20*N,ncol=1)
S<-matrix(nrow=1,ncol=2)
Jhat<-matrix(nrow=N-0.20*N,ncol=2)
CZhat<-matrix(nrow=N-0.20*N,ncol=2)
ERhat<-matrix(nrow=N-0.20*N,ncol=2)
cutpoints<-matrix(nrow=samples,ncol=7)

# Data generation
for (ns in 1:samples) {
u<-runif(N, min=0, max=1)
z<-(-1/lambda)*log(1-u)

# Biomarker data generation
for (i in 1:N) {
if (z[i] <=tau)
  X[i]<-rnorm(n=1,mean=ud,sd=1)
else
  X[i]<-rnorm(n=1,mean=und,sd=1)
```

```
}
# Survival times data generation
c<-runif(N, min=mincenslim, max=maxcenslim)
t<-ifelse(c<=z, c, z)
delta<-ifelse(c<=z,0,1)
data<-data.frame(cbind(X,t,delta))
data<-data[order(data[,1],data[,1]),]

# Cut-points finding process for the ns-th sample
index<-1
for(j in (N*0.10+1):(N-N*0.10)) {
c<-data$X[j]
p<-sum(ifelse(data[,1]>=c,1,0))/N
S<- SurvCalculation(data[data$X>=c,],data[data$X<c,])
TPF[index]<-((1-S[[1]])*p)/((1-S[[2]])*(1-p) +(1- S[[1]])*p)
FPF[index]<- (S[[1]]*p)/(S[[2]]*(1-p)+ S[[1]]*p)
index<-index+1
}

# Estimation of the Youden objective function (3.11) for censored data
Jhat[,1]<-data$X[(N*0.10+1): (N-N*0.10)]
Jhat[,2]<-TPF[,1]-FPF[,1]

# Estimation of the concordance probability objective function (3.12) for censored data
CZhat[,1]<- data$X[(N*0.10+1): (N-N*0.10)]
CZhat[,2]<-TPF[,1]*(1-FPF[,1])
```

```
# Estimation of the point closest-to-(0,1) corner in the ROC plane objective function (3.12) for
censored data

ERhat[,1]<- data$X[(N*0.10+1): (N-N*0.10)]

ERhat[,2]<-sqrt((TPF[,1]-1)^2+(FPF[,1])^2)


Jhat<-Jhat[order(Jhat[,2]),]

CZhat<-CZhat[order(CZhat[,2]),]

ERhat<-ERhat[order(ERhat[,2]),]


# Selection of the optimal cut-point for the ns-th sample

cutpoints[ns,1]<-ns

cutpoints[ns,2]<-Jhat[nrow(Jhat),1]

cutpoints[ns,3]<-Jhat[nrow(Jhat),2]

cutpoints[ns,4]<-CZhat[nrow(CZhat),1]

cutpoints[ns,5]<-CZhat[nrow(CZhat),2]

cutpoints[ns,6]<-ERhat[1,1]

cutpoints[ns,7]<-ERhat[1,2]

}


#Relative Bias calculation

BiasJ<-mean((cutpoints[,2]-truec)/truec)

BiasCZ<-mean((cutpoints[,4]-truec)/truec)

BiasER<-mean((cutpoints[,6]-truec)/truec)


#Mean Square Error calculation

MSEJ<-mean((cutpoints[,2]-truec)^2)

MSECZ<-mean((cutpoints[,4]-truec)^2)
```

```
MSEER<-mean((cutpoints[,6]-truec)^2)
```

REFERENCES

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 2001; 69(3): 89-95.
2. Schisterman EF, Albert PS. The biomarker revolution. Stat. Med. 2012; 31(22): 2513-2515.
3. Gosho M, Nagashima K, Sato Y. Study Designs and Statistical Analyses for Biomarker research. Sensors 2012; 12(7): 8966-8986.
4. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M,Yasui Y. Phases of Biomarker Development for Early Detection of Cancer. JNCI. 2001; 93(14): 1054-1061.
5. Jenkins M, Flynn A, Smart T, Harbron C, Sabin T, Ratnayake J, Delmar P, Herath A, Jarvis P, Matchman J, on behalf of the PSI Biomarker Special Interest Group. A statistician's perspective on biomarkers in drug development. Pharm. Stat. 2011; 10(6): 494-507.
6. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann. Neurol. 2009; 65(4): 403-413.
7. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. Oncogene 2006; 25(11): 1679-91.
8. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009; 338: b375.

9. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. 1st ed. New York, NY, USA: Springer, 2008
10. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996; 312(7023): 71-72.
11. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009; 338: b604.
12. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. BMC. Med. 2010; 8: 21.
13. Miller R, Siegmund D. Maximally selected chi square statistics. Biometrics. 1982; 38(4): 1011-1016.
14. Youden WJ. Index for rating diagnostic tests. Cancer 1950; 3(1): 32-35.
15. Liu X. Classification accuracy and cut point selection. Stat. Med. 2012; 31(23): 2676-2686.
16. Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford University Press: New York, 2003.
17. Antolini L, Valsecchi MG. Performance of binary markers for censored failure time outcome: nonparametric approach based on proportions. Stat. Med. 2012; 31(11-12): 1113-1128.
18. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am. J. Epidemiol. 2006; 163(7): 670-675.
19. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J. Clin. Epidemiol. 2003; 56(11): 1129-1135.

20. Böhning D, Holling H, Patilea V. A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. Stat. Methods. Med. Res. 2011; 20(5): 541-550.
21. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology. 2005; 16(1):73-81.
22. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 2000; 19(9): 1141-1164.
23. Vincenti A, Rota M, Spinelli M, Corciulo M, De Ceglia S, Rovaris G, Antolini L, Genovesi S. A noninvasive index of atrial remodeling in patients with paroxysmal and persistent atrial fibrillation: a pilot study. J Electrocardiol. 2012; 45(2): 109-115.
24. Etzioni R, Pepe M, Longton G, Hu C, Goodman G. Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. Med. Decis. Making. 1999; 19(3): 242-251.
25. Smith DS, Catalona WJ. Rate of change in serum prostate specific antigen levels as a method for prostate cancer detection. J. Urol. 1994; 152(4): 1163-1167.
26. Connolly D, Black A, Murray LJ, Napolitano G, Gavin A, Keane PF. Methods of calculating prostate-specific antigen velocity. Eur. Urol. 2007; 52(4): 1044-1050.
27. Marubini E, Valsecchi MG. Analysing Survival Data from Clinical Trials and Observational Studies. John Wiley & Sons. 1995, Chichester, United Kingdom.
28. Kleinbaum DG. Survival Analysis: A Self-Learning Text. Springer-Verlag 1996, New York.
29. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. J. Am. Stat. Assoc. 1958; 53(282):457-481.

30. Greenwood M. A report on the natural duration of cancer. In Reports on Public Health and Medical Subjects. 1926; 33:1-26. Her Majesty's Stationery Office, London.
31. Bayes T. An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. Phil. Trans. 1763; 53:370-418.
32. Haegerty P, Lumley T, Pepe MS. Time dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56(2):337-344.
33. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. Statist. Sci. 1986; 1(1):54-75.
34. The United Kingdom Childhood Cancer Study: objectives, materials and methods. UK Childhood Cancer Study Investigators. Br. J. Cancer. 2000; 82(5):1073-1102.
35. Palmi C, Vendramini E, Silvestri D, Longinotti G, Frison D, Cario G, Shochat C, Stanulla M, Rossi V, Di Meglio AM, Villa T, Giarin E, Fazio G, Leszl A, Schrappe M, Basso G, Biondi A, Izraeli S, Conter V, Valsecchi MG, Cazzaniga G, Te Kronnie G. Poor prognosis for P2RY8-CRLF2 fusion but not for CRLF2 over-expression in children with intermediate risk B-cell precursor acute lymphoblastic leukemia. Leukemia. 2012; 26(10): 2245-2253.
36. Cario G, Zimmermann M, Romey R, Gesk S, Vater I, Harbott J, Schrauder A, Moericke A, Izraeli S, Akasaka T, Dyer MJ, Siebert R, Schrappe M, Stanulla M. Presence of the P2RY8-CRLF2 rearrangement is associated with a poor prognosis in non-high-risk precursor B-cell acute lymphoblastic leukemia in children treated according to the ALL-BFM 2000 protocol. Blood. 2010; 115(26): 5393-5397.
37. Harvey RC, Mullighan CG, Chen IM, Wharton W, Mikhail FM, Carroll AJ, Kang H, Liu W, Dobbin KK, Smith MA, Carroll WL, Devidas M, Bowman WP, Camitta BM, Reaman GH,

Hunger SP, Downing JR, Willman CL. Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. Blood. 2010; 115(26): 5312-5321.

38. Shiu SH, Gatsonis C. The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. Philos. Transact. A Math. Phys. Eng. Sci. 2008; 366(1874): 2313-2333.
39. Grimes DA, Schulz KF. Uses and abuses of screening tests. Lancet. 2002; 359(9309): 881-884.
40. Böhning D, Böhning W, Holling H. Revisiting youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. Stat. Methods. Med. Res. 2008; 17(6): 543-554.
41. Nakas CT, Alonzo TA, Yiannoutsos CT. Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. Stat. Med. 2010; 29(28): 2946-2955.
42. Lai CY, Tian L, Schisterman EF. Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. Computational Statistics and Data Analysis. 2012; 56(5): 1103-1114.
43. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006; 25(1): 127-141.

ACKNOWLEDGMENTS

The research presented within my PhD dissertation was performed from January 2010 to January 2013 at the Centre of Biostatistics for Clinical Epidemiology of the former Department of Clinical Medicine and Prevention of Milano-Bicocca University under the supervision of Prof. Maria Grazia Valsecchi and Dr. Laura Antolini.

My PhD was financially supported by “FONDAZIONE AMICI DELL’EPATOLOGIA PER LA RICERCA OSPEDALIERA SULLA PREVENZIONE, DIAGNOSI E TERAPIA DELLE MALATTIE EPATICHE” (FADE) O.N.L.U.S. through the “Value-Based Medicine in Hepatology (V.B.M.H) - Sviluppo ed Utilizzo di Indicatori di Outcome Clinico in Epatologia” project leaded by Prof. Mario Strazzabosco.

I also wish to acknowledge the European Network for Cancer Research in Children and Adolescents (ENCCA) FP7-HEALTH-F2-2011 Contract no. 261474 project for their partial support to this project.

A huge debt of gratitude must go to Dr. Laura Antolini, an extremely valuable mentor who provided the vision, encouragement and necessary advises for me to proceed through the PhD program and complete my dissertation. She followed me in the best and worst moments of my PhD program and she played a real part in this thesis accomplishment.

Thanks also to my supervisor Prof. Maria Grazia Valsecchi that give me the opportunity to work on this topic and for introducing me to didactic activities. She helped a lot me to address my scientific future.

On a personal level, I should say thanks for the good times spent in the office to several friends and Department colleagues I encountered during my PhD experience: Grazia, Stefania, Laura, Emanuela, Paola, Davide, Anita, Roberta.

At last, but not least, I want to thank my parents that gave me a scientific mind and my whole family for the lovely support who gives me over these years.

I really hope I did not forget anyone! In such case, I apologize with you!

January 2013