COMPARISON OF PROGNOSTIC MARKERS FOR CENSORED OUTCOMES: APPLICATION IN THE NSABP B-14 STUDY

by

Peng Yan

B.Med, Sichuan University, China, 1999

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Peng Yan

It was defended on

July 20, 2011

and approved by

Lan Kong, PhD
Assistant Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Dianxu Ren, MD, PhD
Assistant Professor, Center for Research and Evaluation,
School of Nursing, University of Pittsburgh

Thesis Advisor:
Gong Tang, PhD
Assistant Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

COMPARISON OF PROGNOSTIC MARKERS FOR CENSORED OUTCOMES: APPLICATION IN THE NSABP B-14 STUDY

Peng Yan, M.S.

University of Pittsburgh, 2011

Prognostic markers for risk of recurrence or mortality are becoming very popular and important in the decision making process of cancer patients and their physicians. Those with good prognostics can avoid unnecessary chemotherapies and the resulted agony. The receiver operating characteristic (ROC) curves are often used to assess and compare prognostic markers for binary outcomes. However, they cannot be directly used in assessing prognostic markers for time-to-event outcomes, which are usually subject to censoring. Recently several statistical methods such as the C–index, time-dependent ROC curve and the predictiveness curve have been developed for this purpose. In early stage estrogen receptor-positive (ER+) breast cancer, the 21-gene panel Onco*type* DX assay and the Adjuvant!, based on age, tumor size and grade and other clinical variables, are widely used tools for patient prognosis and provide guidance in decision making. The recurrence score (RS) from the Onco*type* DX assay and a risk index (RI) summarized from Adjuvant! both provide quantitative evaluation of recurrence risk. Here we applied those recently developed statistical methods to compare the prognostic utility of RS and RI in ER+, node-negative (N0), and tamoxifen-treated breast cancer patients enrolled on the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 trial. We showed that the RS was a stronger prognostic marker than RI, and combining RS with clinical variables also

improved the prognostic utility in the NSABP B-14 trial. The results will help to improve treatment decision for breast cancer patients in public health practice.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 TWO PROGNOSTIC MARKERS FOR BREAST CANCER 4

1.2 THE EXEMPLARY DATASET 7

2.0 METHODS 9

2.1 C-INDEX 9

2.2 TIME-DEPENDENT ROC 12

2.3 PREDICTIVENESS CURVE 15

3.0 RESULTS 18

3.1 KAPLAN-MEIER CURVES 18

3.2 COX PROPORTIONAL HAZARDS MODEL 21

3.3 C-INDEX 22

3.4 TIME-DEPENDENT ROC 23

3.5 PREDICTIVENESS CURVE 26

4.0 DISCUSSION 30

5.0 CONCLUTION 32

APPENDIX. R-CODE FOR FIGURES 33

BIBLIOGRAPHY 41

LIST OF TABLES

Table 1. Multivariate Cox proportional hazards models for RI-PCT/50, RS-PCT/50 and the traditional clinic pathologic predictors in B-14 trial 22

Table 2. Point estimates of the C-Index and their 95% confidence intervals (CI) under various Cox models with different sets of covariates 23

LIST OF FIGURES

Figure 1. Data input screen for Adjuvant! Online 6

Figure 2. Scatter plot of RI-PCT and RS-PCT 18

Figure 3. RS-PCT Kaplan-Meier curves by RS risk groups 19

Figure 4. RI-PCT Kaplan-Meier curves by RI risk groups 20

Figure 5. Time-dependent ROC curves at 10 years based on the nearest neighbor estimator (NNE) for B-14 trial. The thicker line represents the curve for RS-PCT and the thinner line represents the curve for RI-PCT. 24

Figure 6. Time-dependent ROC curves at 10 years based on the Kaplan-Meier estimator for B-14 trial. The thicker line presents the curve for RS-PCT and the thinner line presents the curve for RI-PCT 25

Figure 7. Predictiveness curve of RI-PCT and RS-PCT as markers for B-14 trial 27

Figure 8. Predictiveness curve of RI-PCT and RS-PCT combined separately with age, tumor size and tumor grade as markers for B-14 trial 28

1.0 INTRODUCTION

A clinical marker that recognizes or predicts health- and physiology-related event is an important subject in medicine. The event can be the diagnosis of certain disease, the recurrence of cancer, treatment response, or mortality, etc. Effective markers may lead to change in clinical practice in selecting the further proper diagnosis process or treatment.

Many clinical trials collect measurements of a disease marker and time to an interested clinical event. In datasets with time-to-event endpoints, there are usually two types of observations: one is that the event actually occurred during the study and the actual time when the event occurred in the study subject is recorded; the other type of observation is that the event did not occur during the study period and we only know the last time point when the subject was still event-free. The latter observations are often called censored observations in the literature. For example, in a breast cancer clinical trial, the event of interest is the time to recurrence after surgery. However, if a patient passed away because of the heart failure but had been recurrence-free, we only know that this patient was recurrence-free at the time when he or she died. Then this patient is censored at the time when he or she died while the risk of recurrence is the primary interest. Usually there are two primary causes of censored data: one is that a panel study often has limited length of study period and not all of observations will experience the event of interest by the end of the study; the other one is that some observations are lost to follow-up during the study. In statistical analyses of the time-to-event data, censored observations need to be taken

into account even though the actual event time was not observed. Cox proportional hazards models provide a popular and flexible analytical tool for studying the association between important predictors or covariate and the risk of the event of interest in analysis of censored data.

For clinical guidance, we often need to compare different markers or predictors to choose the better prognostic marker or identify a better treatment regimen. Some statistical methods were developed to evaluate the performance of prognostic markers for risk of an event where the event time is subject to censoring. They are C–index (Harrell et al., 1982; Pencina et al., 2004; Uno et al., 2009), time-dependent ROC (Heagerty et al., 2000 & 2005) and predictiveness curve (Pepe et al., 2007). In this thesis, we will describe these methods and illustrate them via comparing two prognostic markers for recurrence risk in early stage breast cancer using data from the NSABP B–14 trial.

C-index is often used in medical literatures to quantify the ability of estimating the risk score with censored or non-censored data. In the case of censored data, the C-index estimates the concordance between the actual survival time and the predicted risk estimated from a Cox proportional hazards model.

A receiver operating characteristic (ROC) is a standard tool to evaluate the predictive utility of a continuous diagnostic measure on a binary outcome, which is based on the notions of specificity and sensitivity. It is important to identify the diagnostic accuracy of quantitative diagnostic measures throughout the whole range of those measures, and helps to assess the optimal cutoff value of the diagnostic measure (Tripep et al., 2009). The result of the diagnosis model is dichotomous, which is either positive (has the disease or event) or negative (does not have the disease or event). Meanwhile, sensitivity and specificity are two types of statistical measures to assess the performance capability of a binary classification test: sensitivity is to

measure the proportion of the actual positive results that are correctly assessed as positive; specificity is the proportion of the actual negatives that are correctly assessed as negatives. The ROC is a graphical plot which y-axis shows the sensitivity or true positive rate, and x-axis shows false positive rate (1 − specificity or 1 − true negative rate). Initially, the ROC curve is designed for continuous (biological) marker X and a binary outcome D. However, for a time-to-event outcome that is subject to censoring, the time-dependent ROC curves should be considered because this method extends the concepts of sensitivity and specificity to time-varying binary variables that the traditional ROC does not include (Heagerty et al., 2000). So when the binary outcome is a time-dependent indicator function D(t), where D(t) = 1 when the event occurred prior to time t and zero otherwise, the ROC curves as a function of time is more appropriate. Two types of time-dependent ROC curves are considered in this thesis (Heagerty et al., 2000). A simple one is directly based on Bayes' theorem and the Kaplan-Meier estimator for each possible threshold value c. But this method may cause the problem of non-monotone sensitivity and specificity functions in X. An alternative method is based on a nearest neighbor estimator (NNE) and guarantees to yield monotone ROC curves. This estimator is provided by the bivariate distribution function of (X, T), where T is event time. For assessment of prognostic models of discrimination, the area under the receiver operating curve (AUC) is a wild using tool. In this thesis, we will use this tool to compare the performance of two clinical markers: the Recurrence Score (RS) derived from Onco*type* DX assay and Adjuvant! Online risk index (RI).

The predictiveness curve, which describes the distribution of disease risk prediction, is another approach for evaluating the marker performance in prognostic. It is a graphical plot where the y-axis is the risk percentile and the x-axis is the risk score from fitting a Cox proportional hazards model. Thus the predictiveness curve is considered to assess the efficiency

of risk model when the marker is applied to the population. It is useful to evaluate the risk associated with clinical population and the performance of marker. In the censored data, we will use the Cox proportional hazards model as the risk model.

1.1 TWO PROGNOSTIC MARKERS FOR BREAST CANCER

Early stage (I or II) breast cancer is the cancer has not spread beyond the breast which can be invasive or non-invasive. It has the highly heterogeneous molecular pathology that can be used to predict the prognosis and evaluate the therapy. For example, the molecular prognostic factor may identify patients with low rate of the cancer recurrence, and those people can gain benefit through shunning chemotherapy that all patients usually receive, because the chemotherapy regimens only have a little benefit for those people with low recurrence rate (Dowsett et al., 2010). Since there is about 15 percent of distant recurrence in women with primary breast cancer who were treated with tamoxifen after breast cancer surgery, about 85 percent of those patients with surgery may be over treated with the chemotherapy if it is given to everyone after surgery (Paik et al., 2004). Therefore identifying the risk of distant recurrence and potential benefit from chemotherapy is highly important for the patients; and the accurate prognostic can be used for guiding treatment after breast cancer surgery. Without such risk assessment as a reference for treatment decision, it is difficult for physicians to decide whether or what type of chemotherapy might be most suitable for those patients who have received breast surgery. There are many molecular factors which have been found for predicting the recurrence of primary breast cancer patients. But only few of them play the important roles to guide the treatment decision making. Recurrence Score (RS) derived from Onco*type* DX assay and

Adjuvant! Online risk index (RI) are two prognostic markers in ER+ early stage breast cancer (Ravdin et al., 2001; Dowsett et al., 2010; Tang et al., 2011).

The RS derived from Onco*type* DX assay is a diagnostic test for tamoxifen-treated women with node-negative and estrogen receptor-positive types of breast cancer. It can significantly help to indicate how likely the distant recurrence happens in breast cancer patients who undergo with or without hormonal therapy and/or chemotherapy. This assay use high-throughput real-time reverse-transcriptase-polymerase-chain-reaction (RT-PCR) method to identify gene expression with formalin-fixed paraffin embedded tissue sections. Onco*type* DX assay is a validated 21-gene assay within a tumor to determine a Recurrence Score, which is to investigate how they are expressed or how active they are. The RS result of the test is a number between 0 and 100 that correlates with the specific likelihood of breast cancer recurrence within 10 years after the initial diagnosis. It has been identified that the tamoxifen-treated patients who have high RS result will gain benefit from the additional chemotherapy and the low RS patients only have a little benefit from the additional chemotherapy (Dowsett et al., 2010).

Adjuvant! Online risk index is a prediction tool for risk of breast cancer related death at 10 years based on Surveillance, Epidemiology, and End Results (SEER) registry data (the web site is http://www.adjuvantonline.com) (**Figure 1**). Adjuvant! Online risk index is based on those variables which were input into the web database: tumor grade (1, 2 or 3), ER status (negative or positive), and tumor size (0.1 to 1, 1.1 to 2, 2.1 to 3, 3.1 to5, or > 5 cm). The risk score is often used to estimate the following status: risk of cancer-related mortality or recurrence without therapy; risk of reduction with therapy and risk of side effects from therapy. But RI also has some limitations, like that some important prognostic factors as HER2 status are not inclusive and small tumors are not characterized well (Paik et al., 2004).

Figure 1 is an example of the data input screen for the online version of Adjuvant! (http://www.adjuvantonline.com/index.jsp). This patient is a 55-year-old woman who had a 1.5 cm, grade 2, N0 and ER+ breast cancer and was treated with tamoxifen and CMF–like chemotherapy. Without receiving additional therapy, at 10 years, the chance for this woman to be alive and recurrence-free is 72%, the chance of having recurrence is 24.6% and the chance of dying from other causes is 3.4%. Had she received the hormonal therapy, her chance of being alive without recurrence at 10 years would increase by 8.8%. However, the corresponding benefit from a CMF-like chemotherapy instead of tamoxifen would be 3.6% at 10 years. After all, taking both tamoxifen and a CMF-like chemotherapy will increase her chance of being alive without recurrence at 10 years by 11.2%.

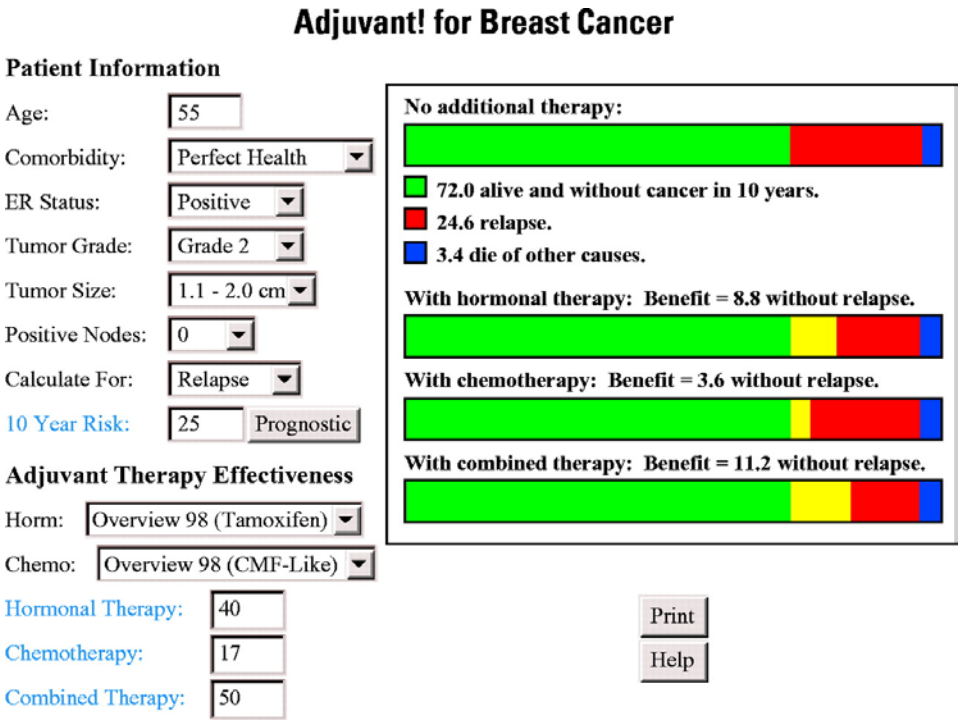

Figure 1. Data input screen for Adjuvant! Online

Both markers, RS and RI, are important for the prognostic and predictive in breast cancer patients treated with hormonal therapy and companied with or without adjuvant systemic therapy in node-negative and estrogen receptor-positive types. So it is clinically important to compare and evaluate the utility of prognostic and predictive among two markers and other factors like tumor size, tumor grade and patient age. For reaching this goal, we will calculate and evaluate the ten-year predicted risk of any distant recurrence or mortality from the dataset including RS and RI.

1.2 THE EXEMPLARY DATASET

National Surgical Adjuvant Breast and Bowel Project (NSABP) trials B-14 is a large clinical trial to assess whether tamoxifen and chemotherapy is more effective in the patients who have node-negative (N0) and estrogen receptor-positive (ER+) primary breast cancer.

In NSABP trial B-14 (entitled “A Clinical Trial to Assess Tamoxifen in Patients with Primary Breast Cancer and Negative Axillary Nodes Whose Tumors Are Positive for Estrogen Receptors”), 2892 breast cancer patients N0 and ER+ were randomly assigned to 5 years of placebo or 5 years of tamoxifen from January 4, 1982 to January 25, 1988, and an additional 1235 breast cancer patients were registered to receive 5 years of tamoxifen treatment from January 26, 1988 to October 17, 1988. In this whole study, 101 patients were ineligible or without follow-up, 2992 patients did not have tumor blocks or no sufficient tissues for the RT-PCR assay, 11 patients were no successful RT-PCR. At the end, the Onco*type* DX assay was successfully on 1023 patients with 355 of them enrolled on the placebo arm and 668 of them enrolled on the tamoxifen arms. In this thesis, our analyses were based on data from those 668

tamoxifen-treated patients. This study of the recurrence score was approved by the Essex Institutional Review Board (Lebanon, N.J.) and by the institutional review boards of Allegheny General Hospital and the University of Pittsburgh (both are in Pittsburgh) (Tang et al., 2011).

Distant recurrence-free interval (DRFI), the time between patient entry and the first distant recurrence occur, is considered as the primary endpoint in this thesis because the distant recurrence is the direct clinical consequence for early stage breast cancer. The censoring events were contralateral breast cancer, non-breast second primary cancer or deaths before distant recurrence. Loco-regional were not counted either as events or as censoring events.

Ligand-binding assay was used to determine the ER-positivity for eligibility in the B–14 trial. And Elston modification of Bloom-Richardson grading criteria was used to determine tumor grade of all patients with Onco*type* DX assay by a broad-certified pathologist from Stanford University (Elston et al., 1991).

In this thesis, the continuous percentiles of Recurrence Score and Adjuvant! Online risks index (RS-PCT and RI-PCT) were used for an appropriate comparison of those two markers. The range of RS-PCT and RI-PCT is from 0 to 100. So we divided them by 50 and re-scaled to a range from 0 to 2 in Cox proportional hazards model as in Paik et al. (2004). Thus the associated hazards ratio (HR) of two markers is comparable to those HRs that are associated with the traditional clinical variables such as patient age, tumor size and tumor grade (Tang et al., 2011).

2.0 METHODS

2.1 C-INDEX

For modern clinical medicine, we need to measure the model validation of discrimination. We hope that our approach is to meet this expectation for testing the probability that longer predicted survival time is actually survive longer without experiencing the event of interest. For the purpose, the risk score system has been established for the assessment in a period of development of cardiovascular disease (CVD), cancer or other conditions of personal risk by Framingham study (Anderson et al., 1991; D'Agostino et al., 2008; Shariat et al., 2008; Parikh et al., 2008). A key component of performance of a risk assessment algorithm is the ability to distinguish who will ("cases") and who will not ("controls") develop these events (Uno et al, 2011). In the research of AUC which is the area under the observed Receiver Operating Characteristics (ROC) curve, this discrimination concept has been well used and studied that is called as a "C-statistic" (Bamber, 1975). This statistics is the conditional probability estimation of the predicted risk of the "case" for any pairs of "case" and "control" (Hanley & McNeil, 1982). And a widespread used nonparametric C-index was suggested by Harrell et al. (1984).

We assume that $\{X_1, X_2 \ldots, X_n\}$ are the covariate vector in a study. Cox proportional hazards model (Cox, 1972) is one of the standard methods to analyze the censored data. The Cox proportional hazards model has the following form for the hazard function:

$$\lambda(t \mid X) = \lambda_0(t)\exp(\beta_1 X_1 + ... + \beta_n X_n) = \lambda_0(t)\exp(\beta' X)$$

Where $\lambda_0(t)$ is the baseline hazard function and β represents the regression parameters. In the following context, $\hat{\beta}$ is denoted as the maximum partial likelihood estimator for β'. Then the larger the value of $\hat{\beta}X$, the higher the risk of recurrence for a subject with the covariates *X*. Denote $\{Y_i, i = 1, \ldots, n\}$ as the event time for subjects and, for simplicity, assume that they are all observed in the following description. Now all the different pairs were considered as $\{(\hat{\beta}X_i, \hat{\beta}X_j, Y_i, Y_j), i, j = 1, \ldots, n; I < j\}$ (so there is no duplication). For a given pair, If $\hat{\beta}X_i < \hat{\beta}X_j$ and $Y_i < Y_j$ or $\hat{\beta}X_i > \hat{\beta}X_j$ and $Y_i > Y_j$ (the inequalities go in the same direction), we consider that this pair is concordant. And if $\hat{\beta}X_i < \hat{\beta}X_j$ and $Y_i > Y_j$ or $\hat{\beta}X_i > \hat{\beta}X_j$ and $Y_i < Y_j$, this pair will be called as discordant. However, not all of them are either concordant or discordant because in the construction of our index, we can use only those pairs which at least one subject had an event. For us, this match is usable. If at the end of the study, the two subjects do not develop any events, we cannot compare them in our predictions. Formed by the pairing of this subjects will be called as unusable. A commonly used concordance measure is:

$C = \Pr(\hat{\beta}X_i < \hat{\beta}X_j \text{ and } Y_i < Y_j \text{ or } \hat{\beta}X_i > \hat{\beta}X_j \text{ and } Y_i > Y_j)$

In reality, not all event times are observed. The discordance between observed event time or censored time and the estimated risk of event, from a Cox model, can only be performed on certain pairs of observations (Harrell et al., 1982). In general, the C-index is to measure the probability of concordance between predicted and observed observations with censored data. If C = 0.5, the prognostic model is random prediction like the coin-flip to determine which patients

will live longer, and if C = 0 or 1, the prognostic model has a perfectly discriminating capability (Harrell et al., 1982).

Pencina M.J and D'Agostino R.B suggested a method to calculate the confidence interval for C-index in 2004. Assuming that N = k + n (k is the events number and n is the non-events number) and C_h was the number of subjects concordant with the *h*th individual in the dataset (h = 1, …, N), this following equation was introduced as following:

$$c_h = \sum_{h \neq j} c_{hj}$$

(c_{hj} = 1 if $\hat{\beta}X_h < \hat{\beta}X_j$ and $Y_h < Y_j$ or $\hat{\beta}X_h > \hat{\beta}X_j$ and $Y_h > Y_j$; otherwise c_{hj} =0)

Then the unbiased estimates of probability of concordance (p_c) and discordance (p_d) can be calculated as:

$$p_c = \frac{1}{N(N-1)} \sum_h c_h$$

$$p_d = \frac{1}{N(N-1)} \sum_h d_h$$

Where d_h is the number of subjects discordant with the *h*th individual in the dataset. Then the predicted C can be calculated as:

$$\hat{C} = \frac{p_c}{p_c + p_d}$$

So confidence interval of the C can be calculated from the estimated probabilities of concordance and discordance. The statistics can be given by following equation (Pencina & D'Agostino, 2004):

$$\frac{w + 2\hat{c}}{2(1+w)} \pm \frac{\sqrt{w^2 + 4w(\hat{c}(1-\hat{c})}}{2(1+w)}$$

$$\text{Where } w = \frac{2z_{\alpha/2}^2}{N(p_c + p_d)}$$

2.2 TIME-DEPENDENT ROC

Sensitivity and specificity are two statistical measures to assess the performance capability of a dichotomous marker, in which the outcomes are labeled either as positive or negative. Sensitivity measures the proportion of the actual positive results that are correctly assessed as positive. Specificity measures the proportion of the actual negatives that are correctly assessed as negatives.

However, continuous markers are frequently used in clinic trial. The ROC curve often uses to present the sensitivity and specificity of a continuous marker. ROC curve is an effective and commonly used method for evaluating the performance of diagnostic tests by displaying sensitivity and specificity of continuous diagnostic markers (X) with various cut-offs in their ranges for predicting a binary disease outcome (D). General description and discussion of ROC can be found in Swets (1982), Hanley (1989), Begg (1991), Zweig (1993), and Pepe (1999). Assuming that X is the marker and D is the binary indicator of disease or event status, let c be a value in the range of X. Without loss of generality, a practical diagnosis using X will associate any subject with X greater than c as an event and any subject with X less than or equal to c as event-free. Therefore each c is associated with a value of sensitivity, that is, pr ($X > c \mid D = 1$), and a value of specificity, pr ($X \leq c \mid D = 0$). The ROC curve is a plot of the sensitivity of the marker X versus 1 – specificity for all possible threshold values c. So the ROC curve is a monotonically non-decreasing function. The larger the area under the ROC curve in the plot, the

better is the diagnostic capacity of the marker for distinguishing between two diagnostic groups (diseased versus normal).

In the classic ROC analysis, the disease status is fixed over time. However, in clinic research, many disease outcomes or interested event status can change over time. For example, a subject may be event-free for a certain time period and the event occurs later on. To analyze the time-to-event or censored data using the ROC method, Heagerty, Lumley and Pepe proposed a time-dependent ROC curve method (Heagerty et al, 2000). They extended the concepts of sensitivity and specificity to time-varying binary variables so that time-dependent ROC curve could be allowed to evaluate the performance of diagnostic tests for outcome variables that are subject to censoring or time-varying. Two types of time-dependent ROC curve are considered in this thesis and both of them can analyze the time-to-event data. The first method directly uses the Kaplan-Meier estimate and Bayes' theorem (Kaplan and Meier, 1958) to produce the time-dependent sensitivity and specificity, but it may lead to a non-monotone sensitivity or specificity function.

Let T_i be the event time and X_i be the marker value for subject i. Let C_i denote the censoring time, Z_i the min (T_i, C_i) during the follow-up time, and δ_i the censoring indicator as $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$. Let $D_i(t) = 1$ if $T_i \leq t$ and $D_i(t) = 0$ if $T_i > t$. Recall the classic sensitivity is $P(X > c \mid D = 1)$ and the classic specificity is $P(X \leq c \mid D = 0)$. Then the sensitivity and specificity of time-dependent function can be defined as:

$$\text{Sensitivity}(c, t) = P\{X > c \mid D(t) = 1\}$$

$$\text{Specificity}(c, t) = P\{X \leq c \mid D(t) = 0\}$$

And with simply using the Bayes' theorem, the sensitivity and specificity of time dependent function can be rewritten as:

$$P\{X > c \mid D(t) = 1\} = \frac{\{1 - S(t \mid X > c)\}P(X > c)}{\{1 - S(t)\}}$$

$$P\{X \le c \mid D(t) = 0\} = \frac{S(t \mid X \le c)P(X > c)}{S(t)}$$

Where S (t) is the survival function S (t) = P (T > t) and the S (t | X > c) is the conditional survival function when X > c.

With the Kaplan-Meier survival estimator and the empirical distribution function of the marker covariate X, a simple estimator of the time dependent sensitivity and specificity is given by as the following (Heagerty et al, 2000):

$$\widehat{P}_{KM}\{X > c \mid D(t) = 1\} = \frac{\{1 - \hat{S}_{KM}(t \mid X > c)\}\{1 - \widehat{F}_X(c)\}}{\{1 - \hat{S}_{KM}(t)\}}$$

$$\widehat{P}_{KM}\{X \le c \mid D(t) = 0\} = \frac{\hat{S}_{KM}(t \mid X \le c)\widehat{F}_X(c)}{\hat{S}_{KM}(t)}$$

Where

$$\hat{S}_{KM}(t) = \prod_{s \in T_n, s \le t} \{1 - \frac{\sum_j I(Z_j = s)\delta_j}{\sum_j I(Z_j \ge s)}\}$$

$$\widehat{F}_X(c) = \sum I(X_i \le c) / n$$

The second method can avoid the nonmonotonic problem which is based on the bivariate distribution function, F(c, t) = P (X ≤ c, T ≤ t), or equivalently S(c, t) = P (X > c, T > t), provided by Akritas (1994).

The estimates of sensitivity and specificity of nearest neighbor estimator (NNE) can be given as the followed (Heagerty et al, 2000):

$$\widehat{P}_{\lambda n}\{X > c \mid D(t) = 1\} = \frac{[\{1 - \widehat{F}_X(c)\} - \hat{S}_{\lambda n}(c, t)]}{\{1 - \hat{S}_{\lambda n}(t)\}}$$

$$\hat{P}_{\lambda n}\{X \le c \mid D(t)=0\} = 1 - \frac{\hat{S}_{\lambda n}(c,t)}{\hat{S}_{\lambda n}(t)}$$

Where

$$\hat{S}_{\lambda n}(c,t) = \frac{1}{n}\sum_i \hat{S}_{\lambda n}(t \mid X = X_i) I(X_i > c)$$

$$\hat{S}_{\lambda n}(t \mid X = X_i) = \prod_{s \in T_n, s \le t} \{1 - \frac{\sum_j K_{\lambda n}(X_j, X_i) I(Z_j = s)\delta_j}{\sum_j K_{\lambda n}(X_j, X_i) I(Z_j \ge s)}\}$$

To examine the performance of two time-dependent ROC curves, we can use the non-parametric method developed by DeLong (Delong et al., 1988). This method is to compare the AUC which can be interpreted as the probability that a randomly selected event or diseased case greater than a randomly selected non-event or no diseased case. From DeLong's method we can calculate the confidence interval for the AUC of time-dependent ROC which can compare the different marker performance under special time.

Another way to compare two markers with time-dependent ROC is the bootstrap method. Based on the functional central limit theorem, the bootstrap method can provide the asymptotically valid confidence limits of the AUCs of the ROC curves at time t.

In this thesis, we use the DeLong's method to develop the confidence interval to compare the two time-dependent ROC curves for examining the performance of RS and RI.

2.3 PREDICTIVENESS CURVE

The predictiveness curve proposed by Bura & Gastwirth (2001) and Copas (1999) shows the analysis combined concepts of risk prediction model and population distribution of risk and

marker. To evaluate how markers predict the cancer risk well, we need to quantify the capability of markers with identifying the cancer risk in population for the modern clinical research. It provides a useful approach for assessing the capacity of model and the clinical utility of the marker for stratifying the population according to risk. In this thesis, we present this plot of estimated risk versus the corresponding percentile of the marker for evaluation of the marker performance.

Using the fitted risk model, we can develop the individual estimated risk utilizing their risk factors from the given data. Assuming that D is the event status, X is the marker and F is the cumulative distribution function (CDF) of the maker, the predictive risk of marker with the *v*th percentile is as the following (Huang et al., 2007):

$$R(v) = P[D = 1 \mid X = F^{-1}(v)], \quad 0<v<1$$

In this thesis, we use Cox proportional hazards model to calculate the predicted risk at 10 years and D is the indicator of event status at 10 years.

Then we order the individual estimated risk from lowest to highest as percentile. To create the curve, we plot the marker percentile as the x-axis and the predictive risk as the y-axis.

In the predictiveness curve, the plot graph shows that the distribution of disease risk in the target population associated with the marker. As Huang (2007) suggested that from the plot graph, $R^{-1}(p)$, the proportion of the population with risks less than p, can be most easily seen in the plot itself. We can mathematically write the association in following equation (Huang et al.; 2007):

$$R^{-1}(p) = P[\text{risk}(X) < p]$$

So R^{-1} (p) is the cumulative distribution of disease risk (X). Correspondingly R (*v*) is the 100 × *v*th percentile of disease risk in the population. This description is simple and useful to display a marker capability for predicting the disease risk.

The disease prevalence in the population, ρ, is a great useful benchmark to assess the predictiveness curve because between the curve and the horizontal line at ρ, the area above and below the horizontal line is equal (Huang et al., 2007). In other words, ρ is the equal risk for all subjects in this population and the perfect marker predictiveness curve is a horizontal line at ρ. Therefore the steeper predictiveness curve indicates a better marker predictive capability.

We can compare the different risk models with their predictiveness curves. R-squared statistic is one approach to compare the different risk models (Pepe et al., 2007). The R-squared value can be interpreted as the percentage of variation in the disease explained by the predicted model. The R-squared value can be calculated as giving formula from the predictiveness curve (Pepe et al., 2007):

$$R^2 = \int_0^1 (pred(v) - \rho)^2 dv / \rho(1-\rho)$$

Where ρ is disease prevalence in the study population and pred (*v*) is the risk value at the *v*th percentile.

3.0 RESULTS

3.1 KAPLAN-MEIER CURVES

In the B–14 trial, RS-PCT was continuous from 2.69 to 100 and RI-PCT was discrete which has 15 subgroups for all 668 objects from 8.23 to 100. The Kendall's tau rank correlation between RS and RI was 0.273 ($P < 0.001$), and Spearman's rank correlation was 0.378 ($P < 0.001$). It shows that the correlation between RS and RI is at most moderate (**Figure 2**).

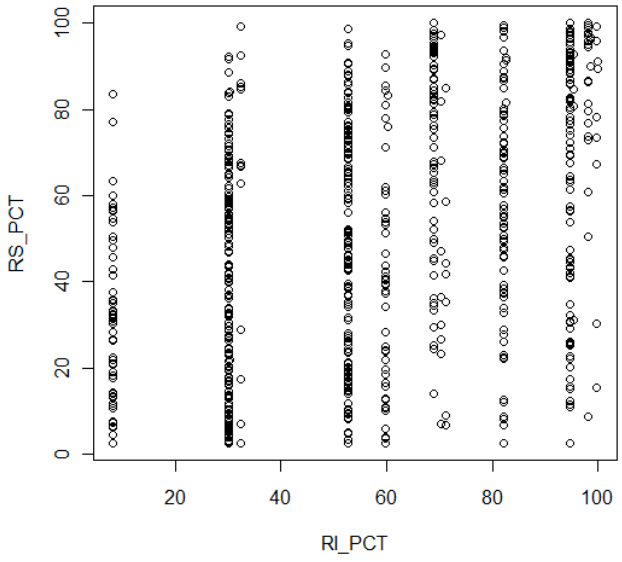

Figure 2. Scatter plot of RI-PCT and RS-PCT

According the published paper, in the B–14 trial the 668 estimated tamoxifen-treated patients were separated to three strata with the RS-PCT: 334 patients (50%) are in the low RS subgroup (RS-PCT < 50), 153 patients (22.9%) are in the intermediate RS subgroup (50≤ RS - PCT < 73) and the rest of 181 patients (27.1%) are in the high RS subgroup (Tang et al., 2011). With the similar distribution as RS-PCT, the RI-PCT was divided into three subgroups: 352 patients (52.7%) were in the low RI subgroup (RI-PCT < 53), 125 patients (18.7%) were in the intermediate RI subgroup (53 ⩽ RI-PCT < 72) and 191 patients (28.6%) were in the high RI subgroup (RI-PCT ⩾ 72).

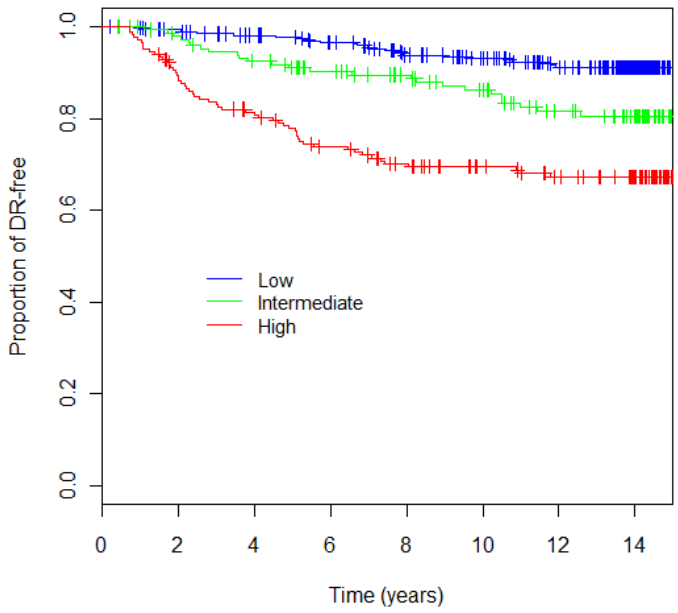

Figure 3. RS-PCT Kaplan-Meier curves by RS risk groups

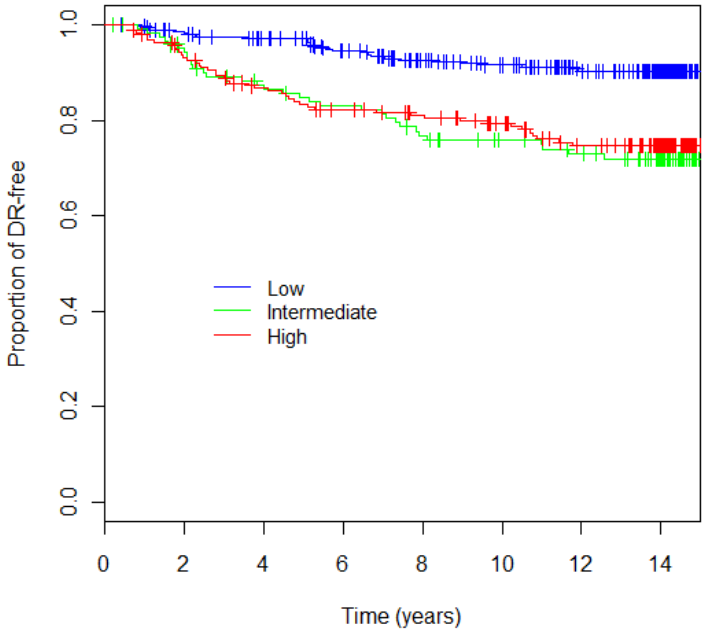

Figure 4. RI-PCT Kaplan-Meier curves by RI risk groups

From the Kaplan-Meier plot, the estimating survival time of low risk subgroup was obvious longer than other two subgroups in both markers (**Figure 3, 4**), and the p-value ($P < 0.0001$) also validated this result. However, in the RI, the two curves of intermediate risk and high risk subgroups were almost coincide and the p-value was 0.646 (**Figure 4**). In the RS, the estimating survival time of intermediate risk subgroup were still obviously longer than the estimating survival time of high risk subgroup ($P = 0.00194$) (**Figure 3**). These results indicate that the RS have better capability in the prognosis than RI especially between intermediate and high risk subgroups.

3.2 COX PROPORTIONAL HAZARDS MODEL

The **Table 1** shows the multivariate Cox proportional hazards model for six models with RS-PCT, RI-PCI and other traditional clinical predictors. Instead of using the original score, the RS-PCT and RI-PCT were divided by 50 respectively as described before. Thus the RS-PCT / 50 and RI-PCT / 50 ranged from 0 to 2. Furthermore, the age was recorded as binary variable (0 is ≤ 50, 1 is > 50), and the tumor grade was rearranged as dummy variable (moderate vs well, poor vs well). And the tumor size range was from 0 to 13 (cm). In this table, -2 LOG L is the -2 logarithm of the likelihood function.

From the Cox proportional hazards model with RS-PCT / 50 and RI-PCT / 50 respectively, the hazard ration (HR) of RS-PCT / 50 was 3.610 (95% CI = 2.486, 5.242), HR with RI-PCT / 50 was 2.874 (95% CI = 1.953, 4.229) and both of markers were statistically significant ($P < 0.0001$). Whereas, the -2 LogL of RS-PCT / 50 (1306.525) was smaller than RI-PCT / 50 (1326.616), which means Model 1 is better than Model 2. Model 5 and Model 6 were RS-PCT / 50 and RI-PCT / 50 combining the traditional clinic pathologic factors: patient age, tumor size and tumor grade. In Model 5, the hazard ration of RS-PCT / 50 was 2.324 (CI = 1.553, 3.477) and the p-value was statistic significant ($P < 0.0001$). However, in Model 6, the hazard ration of RI-PCT / 50 was 1.007 (CI = 0.540, 1.880) and the p-value was not statistic significant ($P = 0.9813$). So with the traditional clinic pathologic predictors, RS-PCT had much more power in the prediction of recurrence risk than RI-PCT.

Table 1. Multivariate Cox proportional hazards models for RI-PCT/50, RS-PCT/50 and the traditional clinic pathologic predictors in B-14 trial

model	variables	Hazard ration (95% CI)	p	-2 LOG L	Likelihood Ratio
1	RS-PCT/50	3.610 (2.486, 5.242)	<0.0001	1306.525	<0.0001
2	RI-PCT/50	2.874 (1.953, 4.229)	<0.0001	1326.616	<0.0001
3	Age (>50 vs<=50)	0.707 (0.479, 1.044)	0.0815	1299.237	<0.0001
	Tumor size	1.188 (1.075, 1.313)	0.0007		
	Grade (moderate vs well)	1.842 (1.043, 3.255)	0.0353		
	Grade (poor vs well)	4.850 (2.786, 8.444)	<0.0001		
4	RS-PCT/50	2.825 (1.910, 4.180)	<0.0001	1296.459	<0.0001
	RI-PCT/50	1.925 (1.273, 2.910)	0.0019		
5	RS-PCT/50	2.324 (1.553, 3.477)	<0.0001	1280.757	<0.0001
	Age (>50 vs<=50)	0.773 (0.522, 1.143)	0.1964		
	Tumor size	1.181 (1.061, 1.313)	0.0023		
	Grade (moderate vs well)	1.382 (0.768, 2.487)	0.28		
	Grade (poor vs well)	2.807 (1.521, 5.179)	0.001		
6	RI-PCT/50	1.007 (0.540, 1.880)	0.9813	1299.236	<0.0001
	Age (>50 vs<=50)	0.708 (0.479, 1.046)	0.0826		
	Tumor size	1.187 (1.039, 1.357)	0.0119		
	Grade (moderate vs well)	1.834 (0.920, 3.654)	0.0847		
	Grade (poor vs well)	4.819 (2.225, 10.437)	<0.0001		

3.3 C-INDEX

Table 2 shows that the C-index results of RS-PCT, RI-PCT and traditional clinic pathologic predictors for B-14 trial. The 95% confidence interval was calculated from conservative confidence interval method of C-index.

In **Table 2**, the C-index number of RS-PCT was 0.711 (CI = 0.615, 0.808) and the number of RI-PCT was 0.664 (CI = 0.563, 0.765). This result showed that the RS-PCT had the better capability in the prediction than the RI-PCT even the confidence interval measure was not

significant. Nevertheless, combination of two markers and adding the traditional clinic pathologic predictors to RS-PCT could not significantly increase the capability of prediction. Therefore, in the C-index method, the single RS model is the best choice for prediction of the clinic practice.

Table 2. Point estimates of the C-Index and their 95% confidence intervals (CI) under various Cox models with different sets of covariates

Included Covariates	C-Index	Lower 95%CL	Upper 95%CL
RS-PCT	0.711	0.615	0.808
RI-PCT	0.664	0.563	0.765
age, size & grade	0.613	0.504	0.721
RI-PCT + RS-PCT	0.719	0.624	0.815
RS-PCT + age, size & grade	0.714	0.618	0.81

3.4 TIME-DEPENDENT ROC

Ten years are usually used as a time point to evaluate the prognosis of an early stage breast cancer patient. In average, an N0 and ER+ breast cancer patient often has 10% - 20% chance to experience recurrence in ten years after treatment. **Figure 5** shows the time-dependent ROC curve at 10 years by using the nearest neighbor estimator (NNE) with a span λ_n = 6.8%. The relative small λ_n could develop moderate smoothing curve which might promote the comparison capability between the NNE and Kaplan-Meier estimator. The **Figure 6** may have monotonicity violations since it used the simple Kaplan-Meier method. **Figure 5** is very similar

to the **Figure 6,** but it may ensure that the estimates of the sensitivity and specificity are monotone.

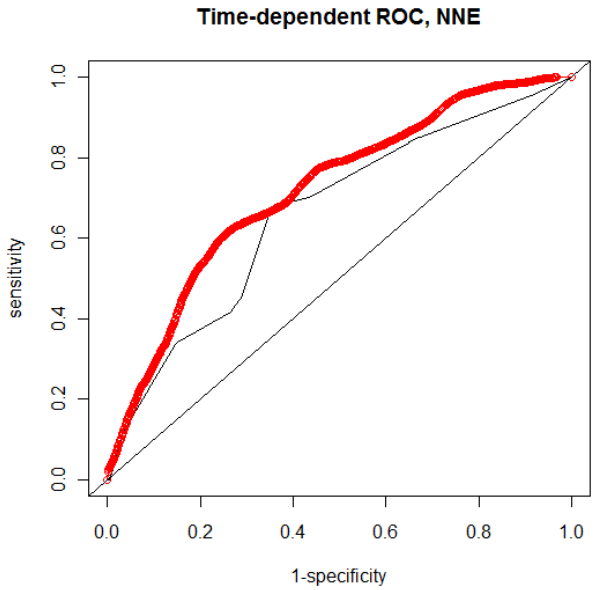

Figure 5. Time-dependent ROC curves at 10 years based on the nearest neighbor estimator (NNE) for B-14 trial. The thicker line represents the curve for RS-PCT and the thinner line represents the curve for RI-PCT.

In **Figure 5** and **Figure 6**, ROC curves show that, at ten years, the RS-PCT is better than RI-PCT, because for any fixed specify, the RS-PCT are always more sensitivity compared to RI-PCT in NEE and KM methods. For example, in the NNE estimator if we use 71.25 as the percentile threshold for distant recurrence at 10 years in RI-PCT (this is the high risk cutoff point for RI-PCT), 73.43% patients who did not have recurrence at ten years will be predicted to distance recurrence-free (specificity) while use the RI-PCT, $P\{ \text{RI-PCT} \leq 71.25 \mid D(10) = 0\} = 0.73$, and 41.66% patients who did have recurrence will be predicted to have distance recurrence (sensitivity), $P\{ \text{RI-PCT} > 71.25 \mid D(10) = 1\} = 0.42$. Moreover, With the RS-PCT, when we use

68.11 as the percentile threshold for assessing distant recurrence at 10 years, 73.43% patients who did not have recurrence at ten years will be predicted to distance recurrence-free (specificity) while use the RS-PCT, P {RS-PCT $\leqslant$ 68.11| D (10) = 0} = 0.73, and 62% patients who did have recurrence will be predicted to have distance recurrence (sensitivity), P {RS-PCT > 68.11| D (10) = 1} = 0.62 (**Figure 5)**.

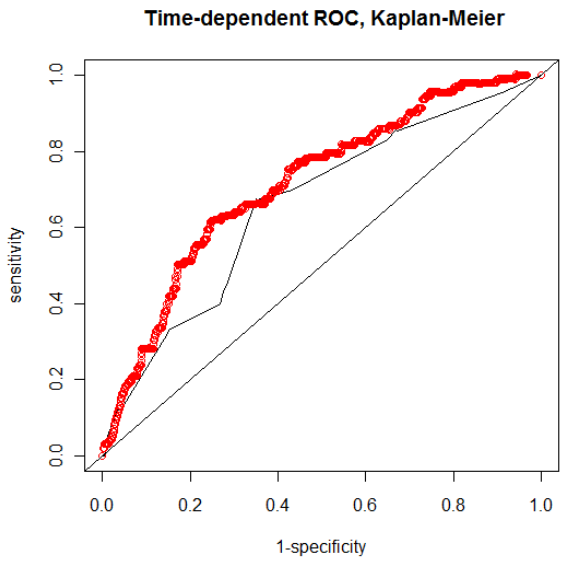

Figure 6. Time-dependent ROC curves at 10 years based on the Kaplan-Meier estimator for B-14 trial. The thicker line presents the curve for RS-PCT and the thinner line presents the curve for RI-PCT.

Nevertheless, in the Kaplan-Meier time-dependent ROC curve, if we use 71.25 as the percentile threshold for the distant recurrence at 10 years, 73.43% patients who did not have recurrence at ten years will be predicted to distance recurrence-free (specificity) when use the RI-PCT, P {RI-PCT $\leq$ 71.25 | D (10) = 0} = 0.73, and 39% patients who did have recurrence will be predicted to have distance recurrence (sensitivity), P {RI-PCT > 71.25 | D (10) = 1} = 0.39.

With the RS-PCT, when the percentile threshold of the distant recurrence at 10 years is set at 68.11, 73.43% patients who did not have recurrence at ten years will be predicted to distance recurrence-free (specificity) while use the RS-PCT, P {RS-PCT $\leqslant$ 68.11 | D (10) = 0} = 0.73, and 61.96% patients who did have recurrence will be predicted to have distance recurrence (sensitivity), P {RS-PCT > 68.11 | D (10) = 1} = 0.62 (**Figure 6)**. Above results suggest that these methods have the coincident results. It demonstrates that RS-PCT is a more efficient diagnosis marker for the recurrence of breast cancer patients.

Another common statistic method based on time–dependent ROC curve to compare two markers in diagnostic test is the AUC. The AUC of RI-PCT was 0.665 and RS-PCT was 0.717 (**Figure 5, 6**). This result also shows that RS-PCT is more efficient to identify the recurrence in the breast cancer patients.

3.5 PREDICTIVENESS CURVE

The **Figure 7** shows the distribution of risk with RS-PCT and RI-PCT. According to the risk model (Cox proportional hazards model), at 90 percentile of RS-PCT, the predicted risk of distant recurrence at 10 years based on the Cox model with RS-PCT is 0.308. That means that 90% those 668 B-14 patients would have a chance less than 30.8% to have distant recurrence at 10 years when RS-PCT is used for prediction and only 10% of those 668 patients' risk of recurrence at 10 years is larger than 30.8%. Based on the model with RI-PCT as the predictor, the 90^{th} percentile of the RI-PCT associated with 0.27 as a predicted risk of distant recurrence at 10 years. Thus a set of 90% breast cancer patients in the B-14 trial would have a less than 27% chance to experience distant recurrence at ten years, and the other 10% high risk B-14 trial

patients' risk of having recurrence at ten years is larger than 27%. Another approach to use this graph is that, we can see with the 0.2 predicted risks of distant recurrence at 10 years based on the Cox model, the associated percentile of RS-PCT is 70.3% and RI-PCT is 71.2%. So we estimate that, 29.7% of those 668 B-14 patients who have distant recurrence risk at ten years are larger than 20% when RI-PCT is used for prediction and 28.8% those patients' recurrence risk at ten years are larger than 20% with RI-PCT using as the predictor. And the recurrence rate at 10 years for B-14 trial is 16.3%.

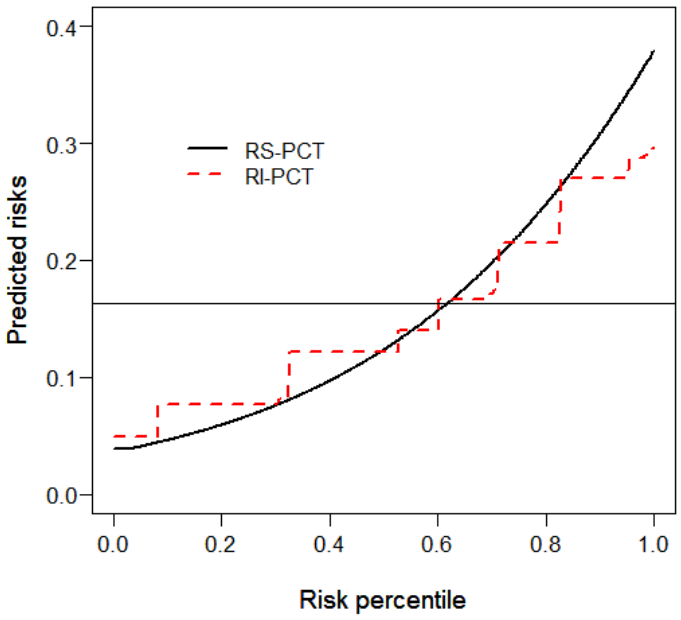

Figure 7. Predictiveness curve of RI-PCT and RS-PCT as markers for B-14 trial.

The **Figure 7** also can compare two predictiveness curves. To contrast two cure shapes, two scale points were picked on x-axis. At 90% of risk percentile, the predicted risk with RS-PCT was 30.8% which was higher than 27% of the predicted risk with RI-PCT. In the low end of

the scale, at 10% risk percentile of the predicted risk with RS-PCT was 4.6% which was lower than 7.7%, the predicted risk with RI-PCT. So in this plot, the RI-PCT curve is less steep than RS-PCT. It shows that the RI-PCT is the poor marker to predict the breast cancer recurrence than RS-PCT.

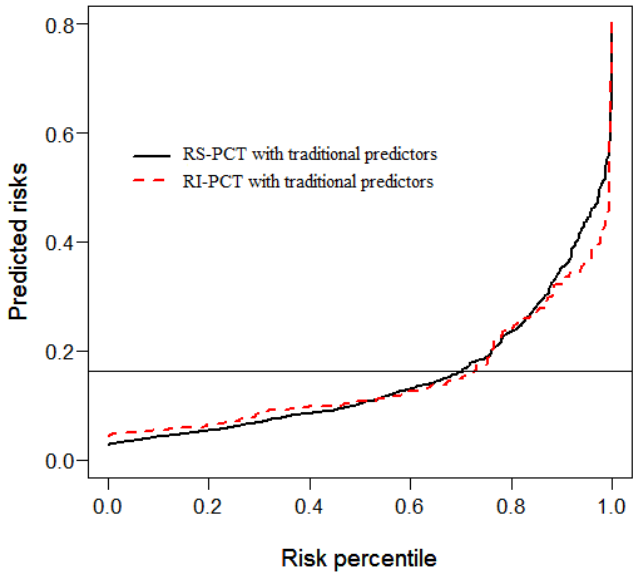

Figure 8. Predictiveness curve of RI-PCT and RS-PCT combined separately with age, tumor size and tumor grade as markers for B-14 trial.

The **Figure 8** shows the distribution of risk of RS-PCT and RI-PCT including the age, tumor size and tumor grade. At the 90% risk percentile, the predicted risk of RS-PCT with the traditional clinical predictors was 35.3% which was higher than 32.1% of the predicted risk of RI-PCT with the traditional clinical predictors. In the low end of the scale, at the 10% risk percentile, the predicted risk of RS-PCT with the traditional clinical predictors was 4.2% which was lower than 5.5% of the predicted risk of RI-PCT with the traditional clinical predictors. In

this plot, the RI-PCT curve with the traditional clinical predictors is still mild. That suggests that the efficient clinical marker for breast cancer is RS-PCT.

4.0 DISCUSSION

Previous research of breast cancer with node-negative and estrogen receptor-positive reports that RS derived from Onco*type* DX assay provides independent and useful prognostic information for the additional chemotherapy decision after surgery and it can be the complementary with traditional clinic pathologic factors like age, tumor size and tumor grade (Palk et al., 2004; Dowsett et al., 2010).

In this thesis, **Figure 2** demonstrated that the RS and RI were independent prediction factors (Spearman rank correlation = 0.378l, P < 0.001; **Figure 2**). According to Dowsett's paper in 2010, the two markers have only approximately 5% explaining by each other in the estimates of recurrence (Dowsett et al., 2010). **Table 1** shows that combination of RS and RI may not be able to improve the capability of prediction for recurrence. The -2 Log likelihood test of Model 4 is 1296.459 which is similar to the Model 1 (1306.525) and Model 2 (1326.616). So the model fit statistics demonstrates that, one marker has the enough competence for prediction of recurrence and combining two markers cannot increase this capability. The C-index also shows same result. In **Table 2**, the C-index of RS plus RI is 0.72 which is similar with the C-index of RS (0.71). Although the C-index of RS plus RI is larger than the C-index of RI (0.66), combination of the two markers still cannot show increased the prognostic capacity than one marker.

The Model 6 of **Table 1** also shows that the RI may be the surrogate of the traditional clinic pathologic factors. The reason is that in Model 6, the hazard ration of RI-PCT / 50 is 1.007

(CI = 0.540, 1.880) and its P–value is 0.9813 which is not statistical significant, and the hazard ratios associated with the traditional clinic pathologic factors (age, tumor size and tumor grade) in Model 6 are similar as in Model 3. But we can see that the RS play an important role in Model 5, because hazard ratio of RS-PCT / 50 is 2.825 ($p < 0.0001$) and the hazard ration of traditional clinic pathologic factors have changed compared with Model 3. Add the RS to the traditional clinic pathologic factors in the regression model, the model fit statistics shows that the new model (Model 5) is more capable (-2logL=1280.757) than the Model 1 (-2logL=1306.525), Model 2 (-2logL=1326.616) and Model 3 (-2logL=1299.237). Therefore combining RS with other clinic pathologic factors (Model 5) does improve the prognostic power over using individual prognostic factors alone.

The main goal of the thesis is to compare the prognostic utility of RS with RI in node-negative, estrogen receptor-positive breast cancer patients. Based on the analysis using Cox proportional hazards model, C-index, time-dependent ROC and predictiveness curve, we may consider that the RS have preferable predictive and prognostic ability for the additional chemotherapy decision after breast cancer surgery. From **Table 1**, the -2 Log likelihood test of RS (Model 1, -2logL=1306.525) performs better than RI (Model 2, -2logL=1326.616) with the Cox proportional hazards model. Meanwhile, the C-index of RS is 0.71 and RI is 0.66 (**Table 2**). Both time-dependent ROC of NNE and KM demonstrate that RS has the better prognostic power (**Figure 5, 6**). On the other side, AUC of RS is 0.717 and RI is 0.665. Thus the RS has better expected performance than RI to predict and prognose for recurrence. And the RS also has the better performance in the predictiveness curve than RI (**Figure 7**). In a nutshell, these four statistical methods all showed that the RS provided more precise prognosis in the 668 NSABP B-14 patients than RI.

5.0 CONCLUSION

In the recent years, more and more breast cancer patients have been diagnosed with node-negative and estrogen receptor-positive (Oetlivotto et al., 2011). So finding a clinic marker to improve the prognosis is an important study because it can help patients and their physicians to make appropriate treatment decision.

In this thesis, we show that the RS has a better performance of prognosis than RI in the NSABP B–14 trial with the N0 and ER+ breast cancer from the Likelihood-ratio test, C-index, time-dependent ROC and predictiveness curve statistic methods. And we also showed that combining RS with other clinic pathologic factors does improve the prognostic power over using individual prognostic factors alone.

APPENDIX

R-CODE FOR FIGURES

```
#### read dataset
setwd("F:\\thesis\\code")
D <- read.csv("thesis-new.csv")
head(D)

#### figure 2 ####

##########
# figure 2 #
#########
plot(D$AO_pctl,D$Recurrence_pctl,xlab="RI_PCT",ylab="RS_PCT",
main="Fig. 2 Scatter plot of RI_PCT and RS_PCT")

##### kendall correlation test####
cor.test(D$Recurrence_pctl, D$AO_pctl, method = "kendall", alternative = "g")
##### spearman correlation test####
```

```
cor.test(D$Recurrence_pctl, D$AO_pctl, method = "spearman", alternative = "g")


#### figure 3 ####

library(survival)

library(survcomp)


##########

# figure 3 #

#########

km.coxph.plot(formula.s=Surv(tdr,cdr) ~ rsc,

data.s=D, sub.s="all", x.label="Time (years)",

y.label="Proportion of DR-free", main.title="",

leg.text=paste(c("Low", "Intermediate", "High"), "    ", sep=""),

leg.pos="topright", leg.inset=0.5, .col=c("blue", "green", "red"),

.lty=c(1,1,1), xlim=c(0,15),verbose=TRUE)



#### figure 4 ####


##########

# figure 4 #

#########


km.coxph.plot(formula.s=Surv(tdr,cdr) ~ ric,

data.s=D, sub.s="all", x.label="Time (years)",
```

```
y.label="Probability of survival", main.title="",
leg.text=paste(c("Low", "Intermediate", "High"), "    ", sep=""),
leg.pos="topright", leg.inset=0.5, .col=c("blue", "green", "red"),
.lty=c(1,1,1), xlim=c(0,15),verbose=TRUE)



##### figure 5############

library (pROC)
library (survivalROC)

nobs <- NROW(D)
cutoff <- 10                                                    # 10 years as cotoff time
Mayo4.1= survivalROC(Stime=D$tdr,
status=D$cdr,marker = D$AO_pctl,
predict.time = cutoff,span = 0.25*nobs^(-0.20) )                # RI marker
Mayo5.1= survivalROC(Stime=D$tdr,
status=D$cdr,marker = D$Recurrence_pctl,
predict.time = cutoff,span = 0.25*nobs^(-0.20) )                # RS marker

##########
# figure 5 #
#########

plot(Mayo4.1$FP, Mayo4.1$TP, type="l", xlim=c(0,1), ylim=c(0,1),
```

```
xlab=paste( "1-specificity"),
ylab="sensitivity",main=paste("Time-dependent ROC, NNE"))
points(Mayo5.1$FP, Mayo5.1$TP, type='o', col="red", xlab='x', ylab='y')
abline(0,1)


##### figure 6############

Mayo4.2= survivalROC(Stime=D$tdr,
status=D$cdr,marker = D$AO_pctl,
predict.time = cutoff, method="KM")                    # RI marker
Mayo5.2= survivalROC(Stime=D$tdr,
status=D$cdr,marker = D$Recurrence_pctl,
predict.time = cutoff, method="KM")                   # RS marker

##########
# figure 6 #
#########

plot(Mayo4.2$FP, Mayo4.2$TP, type="l", xlim=c(0,1), ylim=c(0,1),
xlab=paste( "1-specificity"),
ylab="sensitivity",main=paste(" Time-dependent ROC, Kaplan-Meier"))
points(Mayo5.2$FP, Mayo5.2$TP, type='p', col="red", xlab='x', ylab='y')
abline(0,1)
```

```
for (i in 1:668){
 if (D$tdr[i]<cutoff & D$cdr[i]==1) {D$y[i]<-1} else {D$y[i]<-0}
}
roc1 <- roc(D$y, D$AO_pctl)
roc2 <- roc(D$y, D$Recurrence_pctl)
roc.test(roc1, roc2)                                    #AUC calculate


##########figure 7########
library (PredictABEL)

coxph.fit <- coxph(Surv(D$tdr,D$cdr)~ D$Recurrence_pctl)                  # RS marker
eta <- coxph.fit$linear.predictor
etasort<-sort(eta)
expetasort<-exp(etasort)
Haz0<-basehaz(coxph.fit)[,1]
matrix.Haz<-exp(etasort)%o%Haz0
matrix.risk2<-1-exp(-matrix.Haz)

ind_pr_risk_cox<-max(which(coxph.detail(coxph.fit)$time<=10))        #10 years cutoff time
Cox_med<-matrix.risk2[,ind_pr_risk_cox]

coxph.fit <- coxph(Surv(D$tdr,D$cdr)~ D$AO_pctl)                  # RI marker
eta <- coxph.fit$linear.predictor
```

```
etasort<-sort(eta)
expetasort<-exp(etasort)
Haz0<-basehaz(coxph.fit)[,1]
matrix.Haz<-exp(etasort)%o%Haz0
matrix.risk2<-1-exp(-matrix.Haz)

ind_pr_risk_cox<-max(which(coxph.detail(coxph.fit)$time<=10))        #10 years cutoff time
Cox_med1<-matrix.risk2[,ind_pr_risk_cox]

##########
# figure 7 #
#########

rangeyaxis<-c(0,.4)
labels <- c("RS-PCT", "RI-PCT")
plotPredictivenessCurve(predrisk=cbind(Cox_med,Cox_med1),rangeyaxis=rangeyaxis,
labels=labels,plottitle='',xlabel='Risk percentile')

a<-table(D$cdr)
b<-1-a[1]/(a[1]+a[2])
abline(h=b)

##########figure 8########
```

```
coxph.fit <- coxph(Surv(D$tdr,D$cdr)~
D$Recurrence_pctl+D$tumorsize+D$agec+D$grade2+D$grade3)                # RS marker
eta <- coxph.fit$linear.predictor
etasort<-sort(eta)
expetasort<-exp(etasort)
Haz0<-basehaz(coxph.fit)[,1]
matrix.Haz<-exp(etasort)%o%Haz0
matrix.risk2<-1-exp(-matrix.Haz)

ind_pr_risk_cox<-max(which(coxph.detail(coxph.fit)$time<=10))        #10 years cutoff time
Cox_med<-matrix.risk2[,ind_pr_risk_cox]

coxph.fit <- coxph(Surv(D$tdr,D$cdr)~
D$AO_pctl+D$tumorsize+D$agec+D$grade2+D$grade3)                # RI marker
eta <- coxph.fit$linear.predictor
etasort<-sort(eta)
expetasort<-exp(etasort)
Haz0<-basehaz(coxph.fit)[,1]
matrix.Haz<-exp(etasort)%o%Haz0
matrix.risk2<-1-exp(-matrix.Haz)
```

```
ind_pr_risk_cox<-max(which(coxph.detail(coxph.fit)$time<=10))        #10 years cutoff time
Cox_med1<-matrix.risk2[,ind_pr_risk_cox]

##########
# figure 8 #
#########

rangeyaxis<-c(0,0.8)
labels <- c("RS-PCT",
"RI-PCT")
plotPredictivenessCurve(predrisk=cbind(Cox_med,Cox_med1),rangeyaxis=rangeyaxis,
labels=labels,
plottitle="",xlabel='Risk percentile')

a<-table(D$cdr)
b<-1-a[1]/(a[1]+a[2])
abline(h=b)
```

BIBLIOGRAPHY

Dowsett, M., J. Cuzick, et al. (2010). "Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study." J Clin Oncol 28(11): 1829-1834.

Harrell, F. E., Jr., R. M. Califf, et al. (1982). "Evaluating the yield of medical tests." JAMA 247(18): 2543-2546.

Heagerty, P. J., T. Lumley, et al. (2000). "Time-dependent ROC curves for censored survival data and a diagnostic marker." Biometrics 56(2): 337-344.

Heagerty, P. J. and Y. Zheng (2005). "Survival model predictive accuracy and ROC curves." Biometrics 61(1): 92-105.

Huang, Y., M. Sullivan Pepe, et al. (2007). "Evaluating the predictiveness of a continuous marker." Biometrics 63(4): 1181-1188.

Liu, X., V. Minin, et al. (2004). "Statistical methods for analyzing tissue microarray data." J Biopharm Stat 14(3): 671-685.

Olivotto, I. A., C. D. Bajdik, et al. (2005). "Population-based validation of the prognostic model ADJUVANT! for early breast cancer." J Clin Oncol 23(12): 2716-2725.

Paik, S., S. Shak, et al. (2004). "A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer." N Engl J Med 351(27): 2817-2826.

Pencina, M. J. and R. B. D'Agostino (2004). "Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation." Stat Med 23(13): 2109-2123.

Pencina, M. J., R. B. D'Agostino, Sr., et al. (2008). "Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond." Stat Med 27(2): 157-172; discussion 207-112.

Pepe, M. S., Z. Feng, et al. (2008). "Integrating the predictiveness of a marker with its performance as a classifier." Am J Epidemiol 167(3): 362-368.

Ravdin, P. M., L. A. Siminoff, et al. (2001). "Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer." J Clin Oncol 19(4): 980-991.

Tang, G., S. Shak, et al. (2011). "Comparison of the prognostic and predictive utilities of the 21-gene Recurrence Score assay and Adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20." Breast Cancer Res Treat 127(1): 133-142.

Uno, H., T. Cai, et al. (2009). "On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data." Harvard University Biostatistics Working Paper Series (2009); Paper 101.

Viallon, V. and A. Latouche (2011). "Discrimination measures for survival outcomes: connection between the AUC and the predictiveness curve." Biom J 53(2): 217-236.

Whelan, T. J. and C. Loprinzi (2005). "Physician/patient decision aids for adjuvant therapy." J Clin Oncol **23**(8): 1627-1630.