國立臺灣大學生物資源暨農學院農藝研究所

生物統計學組博士論文

Division of Biometry, Graduate Institute of Agronomy

College of Bioresources and Agriculture

National Taiwan University

Doctoral Dissertation

強化設計標的臨床試驗下設限資料統計推論之研究

A Study on Statistical Inference Based on Censored Data for Targeted Clinical Trials under Enrichment Design

陳瑱芳

Chen-Fang Chen

指導教授：劉仁沛 博士

Advisor: Jen-Pei Liu, Ph.D.

中華民國 102 年 6 月

June, 2013

誌謝

能完成這篇論文，首先要感謝指導教授劉仁沛老師的悉心指導。在論文寫作和期刊文章的發表上，劉老師均不厭其煩地給予莫大的指導與鼓勵，老師的教誨之恩，永誌難忘。同時要感謝口試委員周賢忠老師、季瑋珠老師、蕭金福老師、蔡政安老師、與林志榮老師在口試期間不吝指正，給予許多寶貴地意見，讓我獲益良多。並要感謝研究室學長、同窗同學與學弟妹的陪伴與鼓勵，讓我能順利拿到學位。

最後，謹以此論文獻給我最摯愛的家人，感謝你們在我求學的過程中無限的支持與鼓勵，謝謝你們。

瑱芳　謹誌於

國立台灣大學農藝學研究所生物統計組

中華民國一百零二年七月

中文摘要

在傳統的臨床試驗中，納入和排除標準通常是基於一些臨床指標而未考量受試者的基因或基因的變異。在完成人類基因體計畫後，因可鑑別疾病的分子標的，進而發展出分子標的治療方法。但是分子標的鑑定的診斷試劑通常並非百分之百準確，所以納入標的臨床試驗的陽性診斷病人實際上有些可能並沒有此分子標的。因此，標的臨床試驗下之標的療法對於真正擁有分子標的之病人族群而言的療效估計值會有偏差。因此，我們提出對於真正擁有分子標的之病人配合標的療法之不偏推論的統計方法。在強化設計的臨床試驗及指數分佈及比例化風險迴歸模式下，我們提出利用 EM 演算法配合拔靴技術並考慮鑑定分子標的之診斷試劑的準確度，針對設限資料來進行處理效應之推論。並運用模擬研究加以評估所提出估計式與檢定方式的表現，及提出實例數據以說明方法的應用。

關鍵字：標的臨床試驗; 強化設計; 設限資料; EM 演算法;

ABSTRACT

For the traditional clinical trials, inclusion and exclusion criteria are usually based on some clinical endpoints, the genetic or genomic variability of the trial participants are not totally utilized in the criteria. After completion of the human genome project, the disease targets at the molecular level can be identified and can be utilized for the treatment of diseases. However, the accuracy of diagnostic devices for identification of such molecular targets is usually not perfect. Some of the patients enrolled in targeted clinical trials with a positive result for molecular target might not have the specific molecular targets. As a result, the treatment effect may be underestimated in the patient population truly with the molecular target. To resolve this issue, under the exponential distribution and the Cox-Proportional hazard model, we develop inferential procedures for the treatment effects of the targeted drug based on the censored endpoints in the patients truly with the molecular targets. Under an enrichment design, we propose using the EM algorithm in conjunction with the bootstrap technique to incorporate the inaccuracy of the diagnostic device for detection of the molecular targets on the inference of the treatment effects. A simulation study was conducted to empirically investigate the performance of the proposed methods. The impact of the simulation of the assumption for the proportional hazard model was also examined in the simulation study. Numerical examples illustrate the proposed procedures.

Keywords: Targeted clinical trials, Enrichment design, Censored data, EM algorithm

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Over the years, the most common medical treatment of cancer is the chemotherapy. These drugs rapidly kill all cells indiscriminately, including tumor cells and certain normal tissues. As a result, many patients experience hair loss, gastrointestinal symptoms, and myelosuppression. In the past decade, however, a dramatic shift in cancer therapy has occurred. The Human Genome Project (HGP) started in 1990 to sequence an estimated 3 billion base pairs and identify all human genes and was published in 2003. The genetic information provided by the project combined with the advanced technologies and bioinformatic systems has changed the field of medical research. New breakthrough technologies such as microarrays, mRNA transcript profiling, single nucleotide polymorphisms (SNP), and genome-wide association studies (GWAS) have emerged in a rapid speed since the completion of the human genome project. Therefore, the treatments could be developed to be specific for the patients with the identified molecular targets. As researchers have learned more about the gene changes in cells that cause cancer, they have been able to develop drugs that target these changes. Hence, personalized medicine can finally become a reality. Treatments with these drugs are often called targeted therapies. Targeted therapy is a newer type of cancer treatment that uses drugs or other substances, such as monoclonal antibodies, for the identified molecular targets involved in the pathways of the disease pathogenesis, to more precisely identify and attack cancer cells, usually while doing little damage to normal cells. It is hoped that the patients will benefit from the treatment without toxicity. Targeted clinical trials are the trials that are employed to evaluate the efficacy and safety of the targeted therapies. Current paradigm to develop and evaluate a drug or a

treatment uses a shot-gun approach that may not be beneficial for most of the patients. On the other hand, the targeted therapy employs a guided-missile approach to reach the molecular targets. Targeted therapy is a growing part of many cancer treatment regimens. To address the issues of development of the targeted drugs, the United States Food and Drug Administration (U.S. FDA) issued *Draft Drug-Diagnostic Co-Development Concept Paper* and *Draft Guidance In Vitro Diagnostic Multivariate Index Assays*, respectively, in April 2005 and in July, 2007.

For traditional clinical trials, the intended patient population inclusion and exclusion criteria included clinical endpoints and clinical pathological signs or symptoms. However, despite efforts to reduce the heterogeneity of the patients, there is a big difference in the reaction, even if it is to meet the same inclusion and exclusion criteria with a new treatment. The current paradigm for the development of a drug or a treatment uses a shot-gun approach that may not be beneficial for most patients as these endpoints and clinical signs or symptoms are not well correlated with the clinical benefits of the treatments in the patient population defined by clinical-based inclusion and exclusion criteria. One reason is that the inclusion/exclusion criteria of the traditional trials fail to account for potentially important genes or genomic variation.

The development of targeted modalities requires: (a) the knowledge of the molecular targets involved in the disease pathogenesis; (b) a device for detecting the molecular targets; and (c) a treatment aimed at the molecular targets. Hence, the development of targeted therapies involves evaluation of the translational ability from molecular disease targets to the treatment modalities for the patient population with the targets. A key component of this new paradigm is development of biomarkers that can guide

application of new and existing treatments. This requires a thorough understanding of the relationship between the biomarker and the treatment effect. To address these new challenges in clinical research and development, the U.S. Food and Drug Administration (FDA) recently issued the draft "Drug-Diagnostic Co-development Concept Paper" (FDA 2005). There three designs were introduced in the FDA draft concept paper. One of the three designs is the enrichment design (Chow and Liu 2004). Similar to the traditional trials, the targeted trials using an enrichment design consist of two phases. The first phase is the enrichment phase in which each patient is screened by a diagnostic device for detection of the pre-defined molecular targets, in addition to the inclusion/exclusion criteria based on some clinical endpoints, signs or symptoms. Then only those patients with a positive result for the disease molecular target by a validated diagnostic device are randomized to receive either the targeted treatment or the concurrent control in the second phase. In general the primary analysis of a trial should be based on the overall arm comparison because it is the best estimate of the efficacy of the treatment in the real world.

Statistical concepts and methods for analyzing continuous and categorical endpoints under the enrichment clinical trials were discussed by Liu et al. (2008), and Liu and Lin (2008). They proposed to apply the EM algorithm in conjunction with the bootstrap technique to incorporate the uncertainty on the inaccuracy of the diagnostic device in detection of the molecular targets for the inference of the treatment effects for the targeted therapy for the binary and continuous endpoints under the enrichment design. Clinical endpoints for assessment of efficacy and safety of a promising therapy usually include occurrence of some predefined events such as death, the response to a new chemotherapy in treatment of some advanced cancers, the eradication of an infection

caused by a certain microorganism (e.g., *Helicobacter pylori* for gastric ulcers), serious adverse events (e.g., neutropenia), or the elevation of asparate transaminase three times over the upper limit of the normal range. For these events, the primary parameter of interest is usually time to the occurrence of such an event. Subjects are recruited into the trial at different calendar time points. Note that the predefined event may not be observed on the subjects who complete the scheduled duration of treatment and follow-up. On the other hand, some subjects may withdraw prematurely without observing any occurrences of the event before the end of the study. These individuals are said to be lost to follow-up. As a result we do not have any information on these subjects with respect to the event. The only information we have is that the predefined event did not occur at these subjects in their last visit (either at the end of study or at the time they dropped out from the study). The time to the occurrence of the event therefore is not known for these subjects. We refer an endpoint of this kind to as a censored endpoint. Analysis of censored data has become common practice for clinical trials.

1.1 Accuracy of Diagnostic Devices

In practice, no diagnostic test is perfect with 100% positive predicted value (PPV). For example, MammaPrint is a Class II device approved by the FDA to assess a patient's risk of distant metastasis based on a 70 gene signatures. The PPV is computed based on the data of the TRANSBIG study (Buyse et al. 2006). In decision summary of MammaPrint, the PPV is the probability that metastatic disease occurs within a given time frame given the device output for that patient that is high risk (FDA 2007a). For the metastatic disease at 10 years, the TRANSBIG trial provides an estimate of 0.29 for the PPV with a 95% confidence interval from 0.22 to 0.35. In other words, the patients

testing positive for high risk using the MammaPrint in fact have a 71% probability that the metastatic disease will not occur within 10 years and may receive unnecessary chemotherapy from which these patients will not benefit. Therefore, a futile result from the component of chemotherapy randomization of the MINDACT trial does not mean that the chemotherapy is not effective for the patients truly with a high risk of metastasis. This is because 71% of the patients that tested positive for high risk of distant metastasis by the MammaPrint in fact are not at high risk at all, and the treatment effect of chemotherapy may be underestimated for patients truly at a high risk of distant metastasis.

For another example, the human epidermal growth factor receptor (*HER2*) is a growth factor receptor gene that encodes the *HER2* protein found on the surface of some normal cells that play an important role in the regulation of cell growth. Tumors with over-expressed *HER2* are more likely to recur and the patients have a statistically significantly shorter progression-free survival (PFS) and overall survival (OS) (Seshadri et al. 1993; Ravdin and Chamness 1995). Because the over-expression of the *HER2* gene is a prognostic and predictive marker for clinical outcomes, it provides a target to search for an inhibitor of the *HER2* protein as a treatment for patients with metastatic breast cancer.

Herceptin® (trastuzumab) is a recombinant DNA-derived humanized monoclonal antibody that selectively binds with high affinity in a cell-based assay to the extracellular domain of the *HER2* protein. Several large-scaled, randomized Phase III trials were conducted in the patients with metastatic breast cancer with over-expressed *HER2* protein to confirm the effectiveness and safety of Herceptin®. (Slamon, et al,

2001).

The safety and efficacy of HERCEPTIN were studied in a randomized, controlled clinical trial in combination with chemotherapy (469 patients) and an open-label single agent clinical trial (222 patients). Both the studies employed the enrichment design, which restricted enrolment of women whose breast cancer demonstrated 2+ or 3+ over-expression of *HER2* protein, observed either by IHC(immunohistochemica) assay or gene amplification by FISH (Fluorescence in situ hybridization). All the patients received the standard adjuvant chemotherapy that consists of four 21-day cycles of doxorubicin and AC (anthracyclines and cyclophosphamide), followed by paclitaxel administrated weekly or every 3 weeks for a total of 12 weeks. Both the studies employed a randomized, two-parallel group design, which compared the standard adjuvant chemotherapy plus Herceptin® with the standard adjuvant chemotherapy alone (no treatment control). The treatment of Herceptin® included Herceptin® at 4mg/kg on the day of paclitaxel initiation and subsequently at 2mg/kg for a total of 52 weeks.

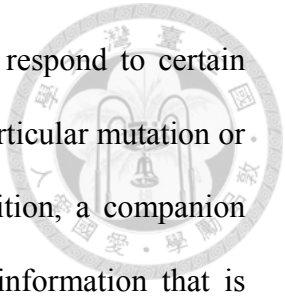
Compared with patients randomized to chemotherapy alone, the patients randomized to HERCEPTIN and chemotherapy experienced a significantly longer median time to disease progression, a higher overall response rate (ORR), a longer median duration of response, and a longer median survival (see Table 1.1). These treatment effects were observed both in patients who received HERCEPTIN plus paclitaxel and in those who received HERCEPTIN plus AC, however the magnitude of the effects was greater in the paclitaxel subgroup.

The commercial assays, HercepTest™ (IHC assay) and PathVysion™ (FISH assay), are

appropriate assays to aid in the selection of patients for HERCEPTIN therapy. The comparability of either assay with regard to the ability to predict clinical benefit from HERCEPTIN therapy has not been prospectively studied. In addition, the utility of either assay in patients whose tumors would score as 0 or 1+ the Clinical Trial Assay (CTA) has not been established because patients with tumors that scored as 0 or 1+ were excluded from the clinical studies described.

*HER*2 protein overexpression can be established by measuring expressed *HER*2 protein using IHC methodology. In the clinical trial studies described above, specimens were tested with the CTA and scored as 0, 1+, 2+, or 3+ with 3+ indicating the strongest positivity. Only patients with 2+ or 3+ positive tumors were eligible (about 33% of those screened). Data from the randomized trial suggest that the beneficial treatment effects were largely limited to patients with the highest level of *HER*2 protein overexpression (3+). In an exploratory analysis (see Table 1.2), the relative risk (rr) for time to progression was lower in the patients whose tumors tested as CTA 3+ (rr = 0.42 with 95% CI: 0.33, 0. 54) than in those tested as CTA 2+ (rr = 0.76 with 95% CI: 0.50, 1.15). The relative risk represents the risk of progression in the HERCEPTIN plus chemotherapy arm versus the chemotherapy arm. Therefore, a lower ratio represents longer time to progression in the HERCEPTIN arm.

This understanding that different people metabolize certain drugs differently, or that their cells bind or process them differently has led to the realization that it is often possible to identify patients who will respond well—or badly—to certain drugs, before they are treated. The molecular tests used to determine the molecular targets are called “companion diagnostics.” Another area in which companion diagnostics are changing



standards of care is the field of cancer therapy. Patients will only respond to certain drugs (such as a monoclonal antibody) if their cancer cells carry a particular mutation or express a particular protein. According to the FDA's formal definition, a companion diagnostic is an in vitro diagnostic test or device that "provides information that is essential for the safe and effective use of a corresponding therapeutic product." (US FDA 2005) But because companion diagnostics are intended to be used in concert with specific drugs, it has not been entirely clear whether the path to FDA approval of these tests is different than that for other in vitro diagnostic tests. To begin to address this problem, in 2011 the US FDA released a document called "Draft Guidance--In Vitro Companion Diagnostic Devices", describing the agency's policies for reviewing a companion diagnostic and the corresponding therapy. The guidance is also intended to clarify when such tests will be required for regulatory approval of a drug, and to outline the regulatory process.

1.2 Statistical Designs

Traditionally, most randomized clinical trials (RCTs) focus on obtaining a reliable estimate of the average treatment effect in a broad patient population. Evaluation of targeted therapies (and biomarkers) often requires larger trials with more complex designs to provide a comprehensive assessment of the relationship between the biomarker and the treatment effect. However, in practice, clinical studies involve a delicate balance between the need for reliable evidence, the need to provide this evidence quickly, and feasibility. Further details can be found in "Advanced in Targeted Therapies Tutorial" in the Website of the US National Cancer Institute. Establishing clinical relevance of a biomarker test for guiding therapy decisions requires

demonstrating that it can classify patients into distinct subgroups with different recommended managements. Conventional RCTs (with no biomarker evaluation) only allow for estimation of the average treatment effect in the overall study population, and therefore, alternative designs must be considered to evaluate biomarker-guided therapy. In general, there are three classes of targeted clinical trials: unselected design, stratified design, and enrichment design. These three designs are given from Figures 1.1 to Figures 1.3.

For the unselected design in Figure 1.1, the information of the test results for the molecular targets is primarily used as covariates and is not involved with randomization. Sometimes, only a part of the patients are tested for the molecular targets. It is useful when the association of the treatment effect of the drug with the results of the diagnostic test needs to be further explored. However, for the unselected design, one of the primary objectives of allocating additional resources for measuring biomarkers is to account for variability of the estimated treatment effects due to the biomarkers. Therefore, it is very important to pre-specify the analysis plan for the inference on the treatment effect with incorporation of the biomarkers in the protocol.

The design in Figure 1.2 is a stratified randomized design and stratification factor is the results of the test for the molecular targets. In other words, the patients are stratified into two groups depending upon whether the diagnostic test is either positive or negative. Then a separate randomization is independently performed within each group to receive the test drug or concurrent control. Analysis of covariance can be employed to explore the patterns of the correlation of the treatment effects with the changes in the magnitudes of biomarkers. For example, if the estimated regression lines of the

responses on the biomarkers are parallel between the treatment and control groups, then there is no treatment-by-biomarker interaction and the estimated treatment effect is relatively the same across the entire range of the biomarker.

The last design is the enrichment design (Chow and Liu, 2004) in which the only patients tested positively for identification of molecular targets are randomized either to receive the test drug or the concurrent control. The enrichment design is usually employed when there is a high degree of certainty that the drug response occurs only in the patients tested positively for the molecular targets and the mechanism of pathological pathways is clearly understood. Most of the Herceptin® phase III clinical trials used the enrichment design. However, as pointed out in the U.S. FDA Concept Paper, the description of test sensitivity and specificity will not be possible using the type of this design without drug and placebo data in the patients tested negative for the molecular targets.

1.3 Aims

From the above example, some of the patients enrolled in targeted clinical trialsunder the enrichment design might not have the specific targets and hence the treatment effects of the drug for the molecular targets could be under-estimated (Liu and Chow, 2008). Liu, et al. (2008) and Liu and Lin (2008) proposed to apply the EM algorithm in conjunction with the bootstrap technique to incorporate the uncertainty on the inaccuracy of the diagnostic device in detection of the molecular targets for the inference of the treatment effects for the targeted therapy for the binary and continuous endpoints under the enrichment design. On the other hand, most of the current targeted

drugs are for the treatment of cancers such as breast cancer, lung cancer, or colorectal cancer. The efficacy endpoints for evaluation of targeted therapies in cancer trials are censored endpoints such as overall survival (OS) or progression free survival (PFS). Currently, literature for the statistical methods taking into account the variability and accuracy of the diagnostic device for molecular targets for the inference based on censored endpoints is scarce. Under the enrichment design, we propose using the EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997) in conjunction with the bootstrap technique (Efron and Tibshirani, 1993) to incorporate the uncertainty on the accuracy of the diagnostic device in detection of the molecular targets for the inference of the treatment effects.

The rest of this dissertation is organized as follows. In the next chapter, the theory of EM algorithm, and the estimation of the standard errors will be reviewed. In Chapter 2, under the assumption that the probability of the patients with a positive diagnostic result having the desired molecular targets is 100%, the traditional procedures for inference of the treatment effects are reviewed. On the other hand, we apply the EM algorithm in conjunction with the bootstrap technique for inference of the treatment effects based on the censored endpoints with parametric hazard model in Chapter 3. The proposed procedure not only incorporates the information of PPV for estimation of the treatment effects and its estimated standard errors, but also to construct the confidence intervals of the treatment effects. In addition, procedures for hypothesis testing and practical example illustrate the utility of the proposed method are also presented in this chapter. In Chapter 4, the EM algorithm is extended to estimation of the treatment effects based on the Cox proportional hazard model. In Chapter 5, the simulation studies, under various combinations of differences in hazard ratio, variability, sample sizes, and

positive predicted values were conducted to empirically investigate the performance of the proposed procedure in terms of the bias and variability of the proposed estimation procedure as well as the size and power of the proposed test procedure. Discussion and remarks on future research are given in Chapter 6.

Table 1.1 Phase III Clinical Efficacy in First-Line Treatment

	Combined Results		Paclitaxel Subgroup		AC Subgroup	
	HERCEPTIN + All Chemotherapy (n = 235)	All Chemotherapy (n = 234)	HERCEPTIN + Paclitaxel (n = 92)	Paclitaxel (n = 96)	HERCEPTIN +AC (n = 143)	AC (n = 138)
Primary Endpoint						
Time to Progression						
Median (months)	7.2	4.5	6.7	2.5	7.6	5.7
95% confidence interval	6.9, 8.2	4.3, 4.9	5.2, 9.9	2.0, 4.3	7.2, 9.1	4.6, 7.1
p-value (log rank)	< 0.0001		< 0.0001		0.002	
Secondary Endpoints						
Overall Response Rate						
Rate (percent)	45	29	38	15	50	38
95% confidence interval	39, 51	23, 35	28, 48	8, 22	42, 58	30, 46
p-value (χ^2-test)	< 0.001		< 0.001		0.1	
Duration of Response						
Median (months)	8.3	5.8	8.3	4.3	8.4	6.4
25%, 75% quartile	5.5, 14.8	3.9, 8.5	5.1, 11.0	3.7, 7.4	5.8, 14.8	4.5 , 8.5
Survival Time						
Median Survival (months)	25.1	20.3	22.1	18.4	26.8	21.4
95% confidence interval	22.2, 29.5	16.8, 24.2	16.9, 28.6	12.7, 24.4	23.3, 32.9	18.3 26.6
p-value (log rank)	0.05		0.17		0.16	

a AC = anthracycline (doxorubicin or epirubicin) and cyclophosphamide.

b Assessed by an independent Response Evaluation Committee.

c Kaplan-Meier Estimate.

Source: FDA (2006).

Table 1.2 Treatment Effect versus Level of HER2 Expression Phase III Randomized Trial

HER2 Assay Result	Number of Patients (N)	Relative Risk for Time to Disease Progression (95% CI)	Relative Risk for Mortality (95% CI)
CTA 2+ or 3+	469	0.49 (0.40, 0.61)	0.80 (0.64, 1.00)
FISH (+)	325	0.44 (0.34, 0.57)	0.70 (0.53, 0.91)
FISH (-)	126	0.62 (0.42, 0.94)	1.06 (0.70, 1.63)
CTA 2+	120	0.76 (0.50, 1.15)	1.26 (0.82, 1.94)
FISH (+)	32	0.54 (0.21, 1.35)	1.31 (0.53, 3.27)
FISH (-)	83	0.77 (0.48, 1.25)	1.11 (0.68, 1.82)
CTA 3+	349	0.42 (0.33, 0.54)	0.70 (0.51, 0.90)
FISH (+)	293	0.42 (0.32, 0.55)	0.67 (0.51, 0.89)
FISH (-)	43	0.43 (0.20, 0.94)	0.88 (0.39, 1.98)

Source: FDA (2006).

FISH, fluorescence *in situ* hybridization.

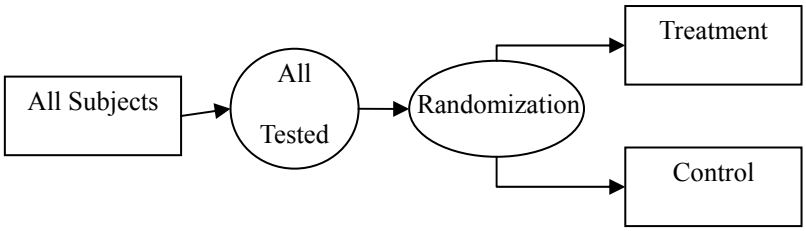


Figure 1.1 Unselected Design for Targeted Clinical Trials

Source: FDA (2005).

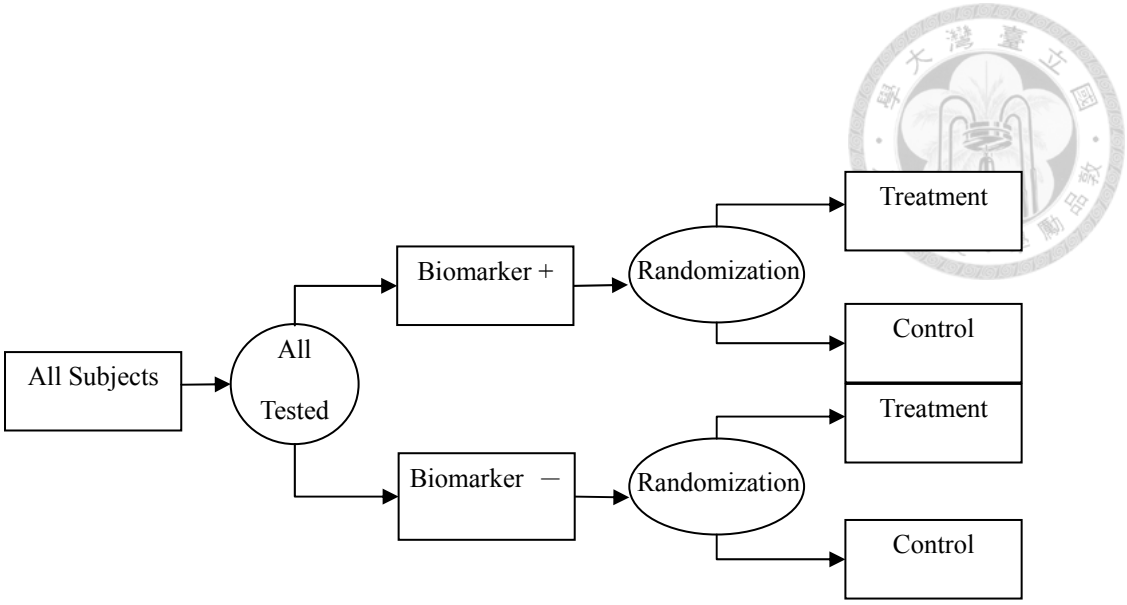


Figure 1.2 Stratified Design for Targeted Clinical Trials

Source: FDA (2005).

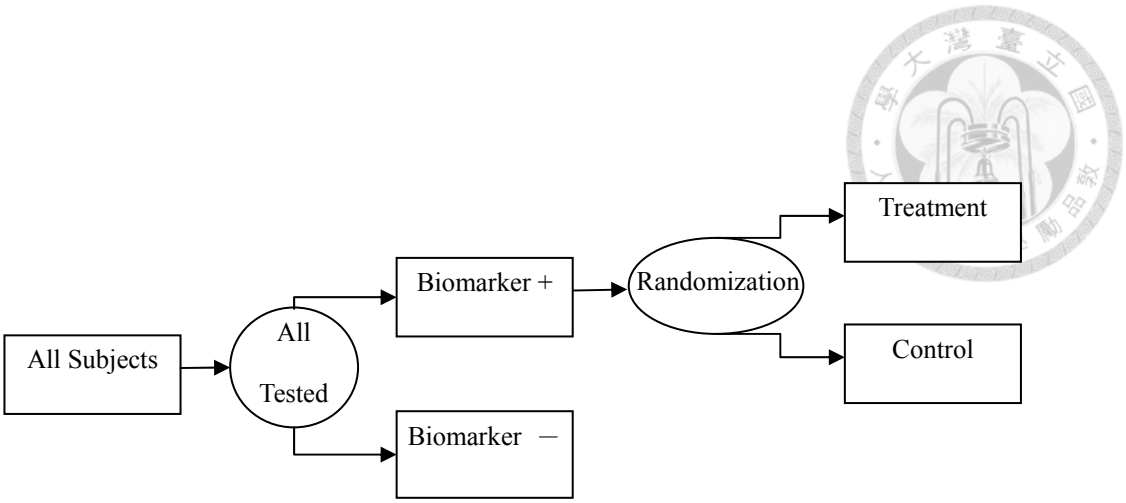


Figure 1.3 Enrichment Design for Targeted Clinical Trials

Source: FDA (2005).

Chapter 2 Literature Review

In the following, we consider the situation where a particular molecular target involved with the pathway in pathogenesis of the disease has been identified and there is a validated diagnostic device available for detection of the identified molecular target. Furthermore, this device is only for detection of the molecular target and is not for prognosis of clinical outcomes of patients. However, this device has been evaluated in the diagnostic effectiveness trial and met the regulatory requirements for diagnostic accuracy. Also suppose that, a test drug for the particular molecular target is available and is currently being developed. The targeted clinical trials consist of two phases under the enrichment design. The first phase is the enrichment phase in which each patient is screened by a diagnostic device for detection of the pre-defined molecular targets. Then the patients with a positive result by the diagnostic device are randomized to receive either the targeted treatment or the untargeted concurrent control.

However, in practice, no diagnostic test is perfect with 100% PPV. As a result, some of the patients enrolled in targeted clinical trials under the enrichment design might not have the specific targets and hence the treatment effects of the drug for the molecular targets could be under-estimated (Liu and Chow, 2008). Under the enrichment design, one of the objectives of targeted clinical trials is to evaluate the treatment effects of the molecular targeted test treatment in the patient population truly with the molecular target. The diagram in the FDA concept paper (U.S. FDA, 2005) for demonstration of this design is reproduced in the Figure 1.3. We consider a two-group parallel design in which patients with a positive result by the diagnostic device are randomized in a 1:1

ratio to receive the molecular targeted test treatment (T) or a control treatment (C). Following Liu and Chow (2008), we propose using the EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997) in conjunction with the bootstrap technique (Efron and Tibshirani, 1993) to incorporate the uncertainty on the accuracy of the diagnostic device in detection of the molecular targets for the inference of the treatment effects. In this Chapter, the theory of EM algorithm, the estimation of the standard errors will be briefly reviewed.

2.1 Efficiency of Enrichment Design

The analysis of an enrichment study design by Friedlin and Simon (2005) addresses the critical drug development issue of efficiency. The randomized discontinuation design is not as efficient as upfront randomization if treatment has a fixed effect on tumor growth rate or if treatment benefit is restricted to slower-growing tumors. On the other hand, the randomized discontinuation design can be advantageous under a model where only a subset of patients, those expressing the molecular target, is sensitive to the agent. To achieve efficiency, the design parameters must be carefully structured to provide adequate enrichment of the randomly assigned patients. Simon and Maitournam (2004) and Maitournam and Simon (2005) found that the efficiency of the enrichment design depended on the prevalence of test-positive patients and on the effectiveness of the new treatment in test-negative patients. For binary end point trials, they showed that the ratio of number of patients to be randomized for the standard trial (n_S) compared with the number randomized in the enrichment trial (n_E) is approximately

$$\frac{n_S}{n_E} \approx \left(\frac{\delta_+}{\text{prev} \times \delta_+ + (1 - \text{prev}) \times \delta_-} \right)^2 f$$

where prev is the proportion of patients who are test positive; δ_+ is the treatment effect for test-positive patients; and δ_- is the treatment effectiveness for test-negative patients. The variable f is a constant that does not depend on the prevalence or treatment effects; it is generally close in value to 1 unless the control response rate is very low. When fewer than half of the patients are test positive and the new treatment is relatively ineffective in test negative patients, the number of randomized patients required for an enrichment design is often dramatically smaller than the number of randomized patients required for a standard design (Simon, 2008). However, they failed to account for the variability associated with the estimates of the diagnostic accuracy. Some of the patients enrolled in targeted clinical trials under the enrichment design might not have the specific targets.

2.2 EM Algorithm

The Expectation-Maximization (EM) algorithm is used to find the maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. Typically these models involve latent variables in addition to unknown parameters and known data observations. That is, either there are missing values among the data, or the model can be formulated more simply by assuming the existence of additional unobserved data points. (For example, a mixture model can be described more simply by assuming that each observed data point has a corresponding unobserved data point, or latent variable, specifying the mixture component that each data point belongs to.) (Dempster et al., 1977). Although the diagnostic effectiveness trials of the diagnostic device can provide independent estimates of the positive predictive value and all patients randomized under the enrichment design have a positive diagnosis, the true

status of the molecular target for individual patients in the target clinical trial is in fact unknown. We assume that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are a set of independent random variables each having a mixture proportion r of survival distributions with a vector of unknown parameter(s) $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. Both for the patient population are truly with and without the molecular target. Denote Ψ contain all of the unknown parameters. Then, the incomplete-data log-likelihood function for Ψ is given by

$$\log L_I(\Psi) = \sum_{j=1}^{n} \log\left\{ r\varphi(\mathbf{y}_j \mid \boldsymbol{\theta}_1) + (1-r)\varphi(\mathbf{y}_j \mid \boldsymbol{\theta}_2) \right\},$$

where $\varphi(\mathbf{y}_j \mid \Psi)$ denotes the survival distribution. The observed-data $\mathbf{y}_{obs} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ are regarded as being incomplete, the latent variables x_j are introduced, where x_j is defined to be one or zero according to whether x_j did or did not arise from the *ith* component of mixture model. For this specification, the complete-data log likelihood is

$$\log L_c(\Psi) = \sum_{j=1}^{n} \left\{ x_j \log\left[r\varphi(\mathbf{y}_j \mid \boldsymbol{\theta}_1) \right] + (1-x_j) \log\left[(1-r)\varphi(\mathbf{y}_j \mid \boldsymbol{\theta}_2) \right] \right\} \quad .$$

The EM algorithm is easy to program and proceeds iteratively in two steps, the E-step (expectation) and the M-step (maximization) (McLachlan and Krishnan, 1997). The algorithm converges to a local maximum of the likelihood of the observed data on the $(k+1)st$ iteration. The current fit for the mixing proportions, the component means, and the variances is given explicitly by

$$\hat{r}^{(k+1)} = \sum_{j=1}^{n} x_j^{(k)} \Big/ n \ ;$$

$$\hat{\boldsymbol{\theta}}_1^{(k+1)} = \sum_{j=1}^{n} x_j^{(k)} \mathbf{y}_j \Big/ \sum_{j=1}^{n} x_j^{(k)} \quad ; \quad \hat{\boldsymbol{\theta}}_2^{(k+1)} = \sum_{j=1}^{n} (1-x_j^{(k)}) \mathbf{y}_j \Big/ \sum_{j=1}^{n} (1-x_j^{(k)}) \ ;$$

2.3 Convergence of EM Algorithm

Wu (1983) discuss the convergence of an EM sequence of iterate. Two convergence aspects of the EM algorithm are studied:

(1) Dose the EM algorithm find a local maximum or a stationary value of the (incomplete-data) likelihood function?

(2) Dose the sequence of parameter estimates generated by EM converge? Several convergence results are obtained under conditions that are applicable to many practical situations.

Therefore, two useful special cases are:

(a) If the unobserved complete-data specification can be described by a curved exponential family with compact parameter space, all the limit points of any EM sequence are stationary points of the likelihood function.

(b) If the likelihood function is unimodal and a certain differentiability condition is satisfied, then any EM sequence converges to the unique maximum likelihood estimate.

2.4 Estimator of the Standard Error

The EM algorithm is a popular method for computing maximum likelihood estimates. However, the EM algorithm fails to automatically provide an estimate of the standard errors of the MLE. Basford et al. (1997) compared two methods of estimation of the standard errors: the standard information-based method and the computationally -intensive bootstrap method. The first method is based on the information in the sample. Asymptotic variances of the estimated parameters in the mixture model are obtained from the diagonal elements of the inverse of the Fisher information matrix. While the information-based method is asymptotically applicable, it may not provide reliable

estimates of the standard errors of the component means unless the sample size is very large or the component means are well separated. The second method involves the calculation of bootstrap estimates of the covariance matrix. This method will provide accurate standard error estimates provided that a sufficient number of bootstrap samples are generated. Under a normal mixture model with g components, the parametric bootstrap method used to obtain estimates of the standard errors of the elements of the component-mean vectors $\hat{\mu}_i$, has been developed by (Efron and Tibshirani, 1993). The parametric estimate of the distribution F of the observation vector is

$$\hat{F} = \sum_{i=1}^{g} \hat{r}_i \hat{F}_i .$$

Where $\hat{F}_i$ is the distribution function for the p-dimensional normal distribution with mean $\hat{\mu}_i$ and covariance matrix $\hat{\Sigma}_i$. The parameter $\hat{\mu}_i$, $\hat{\Sigma}_i$ and $\hat{r}_i$ are estimates obtained from fitting a normal mixture model to the original data. A Monte Carlo approximation to the sample covariance matrix of the bootstrap replicates of the fitted means is then calculated as

$$S_i^{(B)} = \sum_{b=1}^{B} (\hat{\mu}_i^{*(b)} - \bar{\hat{\mu}}_i^{*})(\hat{\mu}_i^{*(b)} - \bar{\hat{\mu}}_i^{*})' \Big/ (B-1),$$

where

$$\bar{\hat{\mu}}_i^{*} = \sum_{b=1}^{B} \hat{\mu}_i^{*(b)} \Big/ B .$$

Thus $S_i^{(B)}$ is an approximation to the bootstrap covariance matrix of $\hat{\mu}_i^{*}$, and Hence, to the covariance matrix of $\hat{\mu}_i$. So the standard error of the $\hat{\mu}_i$ can be estimated by the positive square root of $S_i^{(B)}$.

Chapter 3 Statistical Inference under the Exponential Distribution Model

3.1 Current Methods

In the following, based on the assumption that the probability of the patients with a positive diagnostic result having the desired molecular targets is 100%, the traditional procedures for inference of the treatment effects are reviewed. We argue in this section that the traditional estimator for the treatment effects of the target drugs is biased if the PPV is not 100%. If a predefined clinical event is observed in some subjects before the completion of the study, then their exact failure times are known. On the other hand, some subjects may withdraw prematurely without observing any occurrences of the event of interest due to some known or unknown reasons. Sometimes, the event does not occur for some subjects who completed the study. As a result, the time to the occurrence of the event is censored at the last known contact, and it is at least as long as the time from randomization to the time of the last contact. Let C' denote the censoring time associated with the failure time Y. If C' is greater than or equal to Y, then the survival time is actually observed. On the other hand, if the survival time is greater than the censoring time, then the survival time is not observed and is censored. As a result, the censored data for a subject consist of a pair of responses. The first response is the observed time and the second is an indicator identifying whether the observed time is the survival time or was censored at the last contact. In other words, the data for the time to the occurrence of a predefined event obtained from n subjects of a clinical trial can be arranged as (y_1, δ_1), ..., (y_n, δ_n), where y, is the observed time for subject i and

$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is the survival time,} \\ 0 & \text{if } y_i \text{ is censored.} \end{cases} \qquad i = 1, ..., n$$

Suppose that the probability density function of a random variable Y follows an exponential distribution with a mean $\mu = \lambda^{-1}$. For comparing two groups of survival time, suppose that the observations from n_1 individuals in treatment group (T) are expressed as $(y_{i\text{T}}, \delta_{i\text{T}})$, $i = 1, 2, \ldots, n_{\text{T}}$. Let $(y_{i'\text{C}}, \delta_{i'\text{C}})$, $i' = 1, 2, \ldots, n_{\text{C}}$, be the observations from the n_{C} individuals in control group (C). For individuals in treatment group, the hazard function is taken to be λ, and the probability density function and survivor function are given by

$$f(y_{iT}) = \lambda e^{-\lambda y_{iT}}, \quad S(y_{iT}) = e^{-\lambda y_{iT}}, \quad i = 1, ..., n_T .$$

For those in control group, the hazard function is $\psi\lambda$, and the probability density function and survivor function are given by

$$f(y_{i'C}) = \psi\lambda e^{-\psi\lambda y_{i'C}}, \quad S(y_{i'C}) = e^{-\psi\lambda y_{i'C}}, \quad i' = 1, ..., n_T .$$

The joint likelihood of the $n_{\text{T}} + n_{\text{C}}$ observations is

$$\begin{aligned} L(\psi, \lambda) &= \prod_{i=1}^{n_T} \{f(y_{iT})\}^{\delta_{iT}} \{S(y_{iT})\}^{1-\delta_{iT}} \prod_{i'=1}^{n_C} \{f(y_{i'C})\}^{\delta_{i'C}} \{S(y_{i'C})\}^{1-\delta_{i'C}} \\ &= \prod_{i=1}^{n_T} \{\lambda e^{-\lambda y_{iT}}\}^{\delta_{iT}} \{e^{-\lambda y_{iT}}\}^{1-\delta_{iT}} \prod_{i'=1}^{n_C} \{\psi\lambda e^{-\psi\lambda y_{i'C}}\}^{\delta_{i'C}} \{e^{-\psi\lambda y_{i'C}}\}^{1-\delta_{i'C}} . \\ &= \prod_{i=1}^{n_T} \lambda^{\delta_{iT}} e^{-\lambda y_{iT}} \prod_{i'=1}^{n_C} (\psi\lambda)^{\delta_{i'C}} e^{-\psi\lambda y_{i'C}} \end{aligned}$$

The log-likelihood function is

$$\begin{aligned} \log L(\psi, \lambda) &= \log\lambda \sum_{i=1}^{n_T} \delta_{iT} - \lambda \sum_{i=1}^{n_T} y_{iT} + \log(\psi\lambda) \sum_{i'=1}^{n_C} \delta_{i'C} - \psi\lambda \sum_{i'=1}^{n_C} y_{i'C} \\ &= r_T \log\lambda - \lambda T_T^* + r_C \log(\psi\lambda) - (\psi\lambda) T_C^* \\ &= (r_T + r_C)\log\lambda + r_C \log\psi - \lambda(T_T^* + \psi T_C^*) \end{aligned} ,$$

where $r_T = \sum_{i=1}^{n_T} \delta_{iT}$ and $r_C = \sum_{i'=1}^{n_C} \delta_{i'C}$ are the numbers of actual death time in the two groups. $T_1^* = \sum_{i=1}^{n_T} y_{iT}$ and $T_2^* = \sum_{i'=1}^{n_C} y_{i'C}$ are the totals of uncensored and censored survival times in each group. The maximum likelihood estimators for ψ and λ are given respectively

$$\hat{\lambda} = \frac{r_T}{T_T^*} = \frac{r_C}{\hat{\psi} T_C^*}, \text{ and } \hat{\psi} = \frac{r_C T_T^*}{r_T T_C^*}.$$

The estimated value of λ is the reciprocal of the average time survived by individuals in the treatment group, which the estimated relative hazard, $\hat{\psi}$, is the ratio of the average times survived by the individuals in the two groups. The asymptotic variance-covariance matrix of the parameter estimates is the inverse of the information matrix, whose elements are found from the second derivatives of the log likelihood function.

The observed information matrix is given as

$$I(\psi, \lambda) = \begin{pmatrix} r_C / \psi^2 & T_C^* \\ T_T^* & (r_T + r_C) / \lambda^2 \end{pmatrix},$$

with its inverse

$$\frac{1}{(r_T + r_C) r_C - T_C^{*2} \psi^2 \lambda^2} \begin{pmatrix} (r_T + r_C)\psi^2 & -T_C^* \psi^2 \lambda^2 \\ -T_C^* \psi^2 \lambda^2 & r_C \lambda^2 \end{pmatrix}.$$

The standard error of $\hat{\psi}$ and $\hat{\lambda}$ are given by

$$se(\hat{\psi}) = \sqrt{\frac{(r_T + r_C)\hat{\psi}^2}{(r_T + r_C) r_C - T_C^{*2} \hat{\psi}^2 \hat{\lambda}^2}} = \hat{\psi} \sqrt{\frac{r_T + r_C}{r_T r_C}} = \hat{\psi} \sqrt{\frac{1}{r_T} + \frac{1}{r_C}}, \text{ and}$$

$$se(\hat{\lambda}) = \hat{\lambda} / \sqrt{r_T}.$$

The standard errors of these estimates cannot be used directly in the construction of confidence intervals for ψ and λ. The values of both parameters must be positive and their estimated values will tend to have skewed distribution. The distribution of the logarithm of an estimate of rather ψ or λ is much more likely to be symmetric, and confidence limits for the logarithm of the parameter are found using the standard error of the logarithm of the parameter estimate. The approximate variance of $\log\hat{\psi}$ is

$$\text{var}(\log\hat{\psi}) \approx \hat{\psi}^{-2} \text{ var}(\hat{\psi}).$$

Therefore the standard error of $\log\hat{\psi}$ is given by

$$se(\log\hat{\psi}) \approx \hat{\psi}^{-1} se(\hat{\psi}) = \sqrt{\frac{r_T + r_C}{r_T r_C}} = \sqrt{\frac{1}{r_T} + \frac{1}{r_C}}.$$

A 100(1 − α)% confidence interval for the logarithm of the relative hazard ratio has limits $\log\hat{\psi} \pm z_{\alpha/2} se(\log\hat{\psi})$, and confidence limits for hazard ratio ψ are found by exponentiating these limits for logψ.

Under enrichment design, to evaluate the treatment effects of the targeted test treatment in the patient population truly with the molecular target is one of the objectives of targeted clinical trials. Following the enrichment design, a two-group parallel design is considered. The patients with a positive result by the diagnostic device are randomized to receive either the molecular targeted test treatment (T) or an untargeted concurrent control treatment (C). The primary endpoint considered here is the censored data. The data set that we are interested in analyzing consists of observations of the random vecters (Y_{ij}, δ_{ij}), j = 1,…, n_i; i=T, C, where Y_{ij} is the observed survival time for event or censored, and δ_{ij} is an indicator variable that takes 1=event and 0=censored. Therefore, under the assumption of the one-parameter exponential distribution, the probability density functions of the two exponential distributions are $f_i(y_{ij}) = \lambda_i e^{-\lambda_i y_{ij}}$, i=T, C with a mean survival time $\mu_i = \lambda_i^{-1}$, where λ_i is the hazard rate in the *i*th treatment group. The corresponding hazard functions are $h_i(y_{ij}) = \lambda_i$, and the survival functions are $S_i(y_{ij}) = e^{-\lambda_i y_{ij}}$, for $0 \le y_{ij} < \infty$. The maximum likelihood estimators for the mean survival time μ_i are $\hat{\mu}_i = T_i^* / r_i$, wrere $T_i^* = \sum_{j=1}^{n_i} y_{ij}$ and $r_i = \sum_{j=1}^{n_i} \delta_{ij}$, i=T, C are the totals of uncensored and censored survival times and the numbers of actual death time in each group.

Table 3.1 gives the expected values of Y_{ij} whether or not including censored observations by treatment and diagnostic result of the molecular target. In Table 3.1, μ_{T+}, μ_{C+} (μ_{T-}, μ_{C-}) are the mean survival times of test and control groups for the patients with (without) the molecular target. The hypothesis for detection of treatment difference in the patient population truly with the molecular target is the hypothesis of interest:

$$H_0 : \mu_{T+} - \mu_{C+} = 0 \quad \text{vs.} \quad H_a : \mu_{T+} - \mu_{C+} \neq 0 \tag{3.1}$$

Let $\bar{y}_T$ and $\bar{y}_C$ be the sample mean survival time of test and control treatments, respectively. Since no diagnostic test is perfect for diagnosis of the molecular target of interest without error, therefore, some patients with a positive diagnostic result may in

fact do not have the molecular target. It follows that

$$\begin{aligned} E(\bar{y}_T - \bar{y}_C) &= \gamma(\mu_{T+} - \mu_{C+}) + (1-\gamma)(\mu_{T-} - \mu_{C-}) \\ &= \gamma(\lambda_{T+}^{-1} - \lambda_{C+}^{-1}) + (1-\gamma)(\lambda_{T-}^{-1} - \lambda_{C-}^{-1}) \end{aligned}, \quad (3.2)$$

where γ is the positive predicted value.

The expected value of the difference in sample mean survival time consists of two parts. The first part is the treatment effects of the molecular target drug in patients with a positive diagnosis who truly have the molecular target of interest. The second part is the treatment effects of the patients with a positive diagnosis but in fact they do not have the molecular target. Note that the molecular target test drug is assumed efficacious only in the patient population truly with the molecular target. It is ineffective in those patients without the target. Since $\mu_{T+} - \mu_{C+} > \mu_{T-} - \mu_{C-}$, the difference in sample mean survival time obtained under the enrichment design for targeted clinical trials actually under-estimates the true treatment effects of the molecular target test drug in the patient population truly with the molecular target of interest. As it can be seemed from (3.2), the bias of the difference in sample mean survival time decreases as the positive predicted value increases. On the other hand, the positive predicted value of a diagnostic test increases as the prevalence of the disease increases (Fleiss, et al., 2003). For a disease which is highly prevalent, say greater than 10%, and the diagnostic accuracy is quite high, say both sensitivity and specificity reach 95%, the positive predicted value is only about 67.86%. It follows that the downward bias of the traditional difference in sample mean survival time could be substantial for estimation of treatment effects of the molecular target drug in patients truly with the target of interest.

The traditional hypotheses without identification of the molecular targets is given as

$$H_0 : \mu_T = \mu_C \quad \text{vs.} \quad H_a : \mu_T \neq \mu_C .$$

In addition, the null hypothesis can be rewritten in terms of the log of the hazard ratio, $\theta = \log(\lambda_T / \lambda_C) = \log(\mu_C / \mu_T)$, or equivalently,

$$H_0: \theta = 0 \text{ vs. } H_a: \theta \neq 0.$$

The maximum likelihood estimate $\hat{\theta}$ has an approximate normal distribution with estimated variance of $\frac{1}{r_T} + \frac{1}{r_C}$. The traditional z-test approach is to reject the null hypothesis at the α significance level if

$$z = \left| \frac{\hat{\theta}}{se(\hat{\theta})} \right| \geq z_{\alpha/2}$$

Based on the above z-statistic, the corresponding $100(1-\alpha)\%$ confidence interval can be obtained as follows

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{1}{r_T} + \frac{1}{r_C}}$$

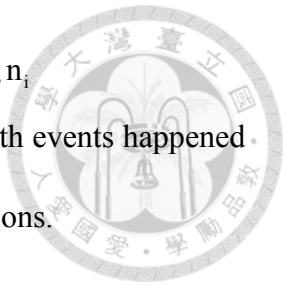
Since $\bar{y}_T - \bar{y}_C$ under-estimates $\mu_{T+} - \mu_{C+}$, the planned sample size may not be sufficient for achieving the desired power for detecting the true treatment effects in the patients truly with molecular target of interest.

3.2 The Proposed Procedure

Although the diagnostic effectiveness trials of the diagnostic device can provide independent estimates of the positive predictive value and all patients randomized under the enrichment design have a positive diagnosis, the true status of the molecular target for individual patients in the target clinical trial is in fact unknown. It follows that Y_{ij} are independently distributed as a mixture of two exponential distributions with hazard λ_{i+} and λ_{i-} respectively

$$\varphi(y_{ij}, \delta_{ij} \mid \lambda_{i+})^{\gamma} \varphi(y_{ij}, \delta_{ij} \mid \lambda_{i-})^{1-\gamma} \qquad i = T, C \;\; ; \; j = 1, \ldots, n_i$$

where $\varphi(. \mid .)$ denotes the density function of a exponential variable with events happened or survival function of a exponential variable with censored observations.



However, γ is an unknown positive predictive value which must be estimated from the data. Therefore, the data obtained from the targeted clinical trials are incomplete because the true status of the molecular target of the patients is unknown.

We apply the EM algorithm to estimate the treatment effects for the population of the patients truly with the molecular target by incorporating the estimates of the positive predictive value of the device obtained from the diagnostic effectiveness trials as the initial values.

For each patient, we have a set of variables (Y_{ij}, δ_{ij}, X_{ij}), where Y_{ij} is observed survival time for event or censored, and δ_{ij} is an indicator variable that takes 1=event and 0=censored of patient j in treatment i; X_{ij} is the latent variable indicating the true status of the molecular target of patient j in treatment i; j=1,…,n_i, i=T,C. Therefore X_{ij} is an indicator variable with value of 1 for the patients truly with the molecular target and with a value of 0 for the patients truly without the target. In addition, X_{ij} are assumed i.i.d. Bernoulli random variables with probability with the molecular target being γ. Let $\boldsymbol{\Psi}$ be the vector containing all unknown parameters and (y_{obs}, δ_{obs}) denote the vectors of the observed primary efficacy endpoints from the targeted clinical trial, where

$$\boldsymbol{\Psi} = (\gamma, \lambda_{T+}, \lambda_{T-}, \lambda_{C+}, \lambda_{C-})'$$

and

$$\mathbf{y}_{obs} = (y_{T1}, \ldots, y_{Tn_T}, y_{C1}, \ldots, y_{Cn_C})'$$

$$\boldsymbol{\delta}_{obs} = (\delta_{T1}, \ldots, \delta_{Tn_T}, \delta_{C1}, \ldots, \delta_{Cn_C})'$$

It follows that the complete-data log-likelihood function for Ψ is given by

$$\log L_c(\Psi) = \sum_{j=1}^{n_T} x_{Tj}\left[\log\gamma + \log\varphi(y_{Tj}, \delta_{Tj} \mid \lambda_{T+})\right] + \sum_{j=1}^{n_T}(1-x_{Tj})\left[\log(1-\gamma) + \log\varphi(y_{Tj}, \delta_{Tj} \mid \lambda_{T-})\right] + \sum_{j=1}^{n_C} x_{Cj}\left[\log\gamma + \log\varphi(y_{Cj}, \delta_{Cj} \mid \lambda_{C+})\right] + \sum_{j=1}^{n_C}(1-x_{Cj})\left[\log(1-\gamma) + \log\varphi(y_{Cj}, \delta_{Cj} \mid \lambda_{C-})\right]$$

Furthermore, from the previous diagnostic effectiveness trials, an estimate of the PPV of the device is known. Therefore, at the initial step of the EM algorithm for estimation the treatment effects in the patients truly with the molecular target, the observed latent variable X_{ij} are generated as i.i.d. Bernoulli random variables with γ estimated by that obtained from the diagnostic effectiveness trial. What follow are the procedures for implementation of the EM algorithm to estimate θ in the patient population truly with the molecular target.

At the initial step of the EM algorithm, the observed latent variable X_{ij} are generated as i.i.d. Bernoulli random variables with the positive predictive value γ estimated by that obtained from the diagnostic effectiveness trial. At the $(k+1)$st iteration, the E-step requires the calculation of the conditional expectation of the complete-data log-likelihood $L_c(\boldsymbol{\Psi})$, given the observed data $(\mathbf{y}_{obs}, \delta_{obs})$, using currently fitting $\widehat{\boldsymbol{\Psi}}^{(k)}$ for $\boldsymbol{\Psi}$.

$$Q(\boldsymbol{\Psi}; \widehat{\boldsymbol{\Psi}}^{(k)}) = E_{\boldsymbol{\Psi}(k)}\left\{\log L_c(\boldsymbol{\Psi}) \mid \mathbf{y}_{obs}, \delta_{obs}\right\}$$

Since $\log L_c(\boldsymbol{\Psi})$ is a linear function of the unobservable component labeled variables x_{ij}, the E-step is calculated by replacing x_{ij}, by its conditional expectation given y_{ij}, using $\widehat{\boldsymbol{\Psi}}^{(k)}$ for $\boldsymbol{\Psi}$. In other words, x_{ij} is replaced by

$$\hat{x}_{ij}^{(k)} = E_{\Psi(k)}\{x_{ij} \mid y_{ij}, \delta_{ij}\} = \frac{\hat{\gamma}_i^{(k)} \varphi(y_{ij}, \delta_{ij} \mid \hat{\lambda}_{i+}^{(k)})}{\hat{\gamma}_i^{(k)} \varphi(y_{ij}, \delta_{ij} \mid \hat{\lambda}_{i+}^{(k)}) + (1-\hat{\gamma}_i^{(k)}) \varphi(y_{ij}, \delta_{ij} \mid \hat{\lambda}_{i-}^{(k)})} \quad i = T, C$$

which is the estimate of the posterior probability of the observation y_{ij}, δ_{ij} with molecular targets after the k^{th} iteration.

The M-step requires the computation of $\hat{\gamma}_i^{(k+1)}$, $\hat{\lambda}_{i+}^{(k+1)}$, and $\hat{\lambda}_{i-}^{(k+1)}$; i = T, C, by maximizing $\log L_c(\Psi)$. Since $\log L_c(\Psi)$ is linear in the x_{ij}, it follows that x_{ij} are replaced by their conditional expectations $\hat{x}_{ij}^{(k)}$. On the $(k+1)$th iteration, the intent is to choose the value of Ψ, say $\widehat{\Psi}^{(k+1)}$, that maximizes $Q(\Psi; \widehat{\Psi}^{(k)})$. It follows that on the M-step of the $(k+1)$st iteration, the current fit for the positive predictive value of treatment group and control group is given by

$$\hat{\gamma}_i^{(k+1)} = \frac{\sum_{j=1}^{n_i} \hat{x}_{ij}^{(k)}}{n_i}, \; i = T, C.$$

Under the assumption of $n_T = n_C$, it follows that the overall positive predictive value is estimated by

$$\hat{\gamma}^{(k+1)} = (\hat{\gamma}_T^{(k+1)} + \hat{\gamma}_C^{(k+1)}) / 2 .$$

The hazard rate of the molecularly target test drug and control can then be estimated respectively as

$$\hat{\lambda}_{T+}^{(k+1)} = \frac{\sum_{j=1}^{n_T} \hat{x}_{Tj}^{(k)} \delta_{Tj}}{\sum_{j=1}^{n_T} \hat{x}_{Tj}^{(k)} y_{Tj}}, \quad \hat{\lambda}_{T-}^{(k+1)} = \frac{\sum_{j=1}^{n_T} (1-\hat{x}_{Tj}^{(k)}) \delta_{Tj}}{\sum_{j=1}^{n_T} (1-\hat{x}_{Tj}^{(k)}) y_{Tj}}$$

$$\hat{\lambda}_{C+}^{(k+1)} = \frac{\sum_{j=1}^{n_C} \hat{x}_{Cj}^{(k)} \delta_{Cj}}{\sum_{j=1}^{n_C} \hat{x}_{Cj}^{(k)} y_{Cj}}, \text{ and } \hat{\lambda}_{C-}^{(k+1)} = \frac{\sum_{j=1}^{n_C} (1-\hat{x}_{Cj}^{(k)}) \delta_{Cj}}{\sum_{j=1}^{n_C} (1-\hat{x}_{Cj}^{(k)}) y_{Cj}}$$

Therefore, the estimator for the treatment effects in the patients truly with the molecular target θ obtained from the EM algorithm is given as

$$\hat{\theta} = \log(\hat{\lambda}_{T+} / \hat{\lambda}_{C+}) = \log(\hat{\mu}_{C+} / \hat{\mu}_{T+}) .$$

Following Basford, et al. (1997), we propose to apply the parametric bootstrap method to estimate the standard error of $\hat{\theta}$.

Step 1: Choose a large bootstrap sample size, say B = 1000 or above. For $1 \le b \le B$, generate the bootstrap sample $\mathbf{y}_{\mathbf{obs}}^{\mathbf{b}}, \delta_{\mathbf{obs}}^{\mathbf{b}}$ according to the probability model. The parameters for generating bootstrap samples $\mathbf{y}_{\mathbf{obs}}^{\mathbf{b}}, \delta_{\mathbf{obs}}^{\mathbf{b}}$ are substituted by the estimators obtained from the EM algorithm based on the original observations of primary efficacy endpoints from the targeted clinical trial.

Step 2: The EM algorithm is applied to the bootstrap sample $\mathbf{y}_{\mathbf{obs}}^{\mathbf{b}}, \delta_{\mathbf{obs}}^{\mathbf{b}}$ to obtain estimates $\hat{\theta}_b^*$, b=1,..,B.

Step 3: An estimator for the variance of $\hat{\theta}$ by the parametric bootstrap procedure is given as

$$S_B^2 = \frac{\sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2}{B-1},$$

where

$$\bar{\hat{\theta}}^* = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B}.$$

Let $\hat{\theta}$ be the estimator for the treatment effects in the patients truly with the molecular target obtained from the EM algorithm. Let S_B^2 denote the estimator of the variance of $\hat{\theta}$ obtained by the bootstrap procedure. The null hypothesis is rejected and the efficacy of the molecular targeted test drug is different from that of the control in the patient population truly with the molecular target at the α significance level if

$$z = \left| \frac{\hat{\theta}}{\sqrt{S_B^2}} \right| \ge z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper percentile of a standard normal distribution.

The corresponding $100(1-\alpha)\%$ asymptotic confidence interval for $\theta = \log(\lambda_{T+} / \lambda_{C+})$ can be constructed as $\hat{\theta} \pm z_{\alpha/2}\sqrt{S_B^2}$.

It should be noted that although the assumption that $\mu_{T+} - \mu_{C+} > \mu_{T-} - \mu_{C-}$ is one of the reasons for developing the targeted treatment, this assumption is not used in the EM algorithm for estimation of θ. Hence, the inference for θ by the proposed procedure is not biased in favor of the targeted treatment.

3.3 Numerical Example

A targeted drug is developed for the treatment of the patient with a certain cancer whose specific biomarker is over-expressed as measured by an immunohistochemical assay. Suppose that the immunohistochemical assay has a PPV of 0.75. From previous studies, the hazard ratios for the patients truly with and without the biomarker are 0.7 and 1.26, respectively, and are given in Table 3.2. Under the enrichment design, 480 patients with positive test results were randomized in 1:1 ratio to receive either the targeted drug plus the standard chemotherapy or the standard chemotherapy. The censored rate is assumed to be 30%. Table 3.3 provides the point estimates of hazard ratio formortality between the two groups, and their standard error and 95% CIs for the risk when PPV is 0.75.

When PPV is 0.75, the traditional approach without consideration of inaccuracy of diagnostic device yields the estimate of hazard ratio for mortality of 0.8097 with a 95% CI from 0.6550 to 1.0009. Because the 95% CI contains 1, the observed hazard ratio of death is not statistically significant and the targeted drug does not prove its superior efficacy over chemotherapy alone at the 5% level. The reason for the failure of the targeted drug is that 25% positive patients randomized do not have the molecular targets. On the other hand, our proposed EM method provides the estimated hazard ratio of mortality is 0.7108. The 95% CI for the hazard ratio of mortality is (0.5168, 0.9777), which does not contain 1. As a result, the efficacy of the targeted drug can be concluded superior to the control group based on the hazard ratio.

Table 3.1 Population mean survival time by treatment and diagnosis

Positive diagnosis	True target condition	Indicator of diagnostic	Test group	Control group	Difference
+	+	γ	μ_{T+}	μ_{C+}	$\mu_{T+}-\mu_{C+}$
	–	$1-\gamma$	μ_{T-}	μ_{C-}	$\mu_{T-}-\mu_{C-}$

γ is the positive predicted value.

Table 3.2 Treatment effects as a function of a specific biomarker overexpression.

IHC assay result	No. of patients	Hazard ratio for mortality (95% CI)
Test result +	469	0.80 (0.64, 1.00)
True status −	120	1.26 (0.82, 1.94)
True status +	349	0.70 (0.51, 0.90)

Table 3.3 Point and interval estimator of hazard ratio for mortality

	PPV = 0.75	
Results	Traditional	EM
Hazard ratio for mortality	0.8097	0.7108
S.E.	0.1082	0.1626
95% L.C.I.	0.6550	0.5168
95% U.C.I.	1.0009	0.9777

Chapter 4 Statistical Inference under the Parametric Proportional Hazard Regression Model

4.1 Current Methods

In Chapter 3, we only consider the parametric mixture models based on the standard exponential distribution. However, Two-component survival mixture models, in both proportional hazards and accelerated failure time settings, are presented as a flexible method of analyzing such data. Following the diagram of the enrichment design given in Figure 1.3, a two-group parallel design is considered where the patients with a positive result by the diagnostic device are randomized in a 1:1 ratio to receive either the molecular targeted test treatment (T) or an untargeted concurrent control treatment (C). The primary endpoint considered here is the censored data. The data set that we have interest in analyzing consists of observations of the random vectors (Y_{ij}, δ_{ij}), j = 1,…, n_i; i=T, C, where Y_{ij} is observed survival time for event or censored, and δ_{ij} is an indicator variable that takes 1=event and 0=censored.

Under the proportional hazards model, the hazard of death at time *t* for *j*th individuals is given by

$$h(y_j) = e^{\lambda Z} h_0(y_j),\ j = 1,\ldots,n, \tag{4.1}$$

where the covariate Z is a scalar which is the indicator of treatment group (Z = 1 if molecular targeted test treatment group; Z = 0 if an untargeted concurrent control treatment). Consequently, the hazard at time t for an individual in control group is $h_0(y_j)$. Next, we make the additional assumption that the survival times for the individual in control group have a Weibull distribution with scale parameter κ and shape parameter α. The hazard function for the individual in control group is

$$h_0(y_j) = \kappa\alpha y_j^{\alpha-1}, \tag{4.2}$$

The hazard function for the individual in targeted group is $e^{\lambda}\kappa\alpha y_j^{\alpha-1}$ for a Weibull distribution with scale parameter $e^{\lambda}\kappa$ and shape parameter α. The hazard of death at time y for an individual in the targeted test group is proportional to that of an individual in the untargeted control group. The hazard of death for an individual in the targeted test group compared to an individual in the untargeted control group is exp(λ). Therefore, the probability density, survival and hazard function of a Weibull($e^{\lambda Z}\kappa$, α) distribution are given respectively

$$\begin{aligned} f(y) &= e^{\lambda Z}\kappa\alpha y_j^{\alpha-1}\exp(-e^{\lambda Z}\kappa y^{\alpha}), \\ S(y) &= \exp(-e^{\lambda Z}\kappa y^{\alpha}), \\ h(y) &= e^{\lambda Z}\kappa\alpha y_j^{\alpha-1}. \end{aligned} \tag{4.3}$$

The likelihood function consisting of data (y_j, δ_j, z_j), j = 1,…, n is given by

$$\begin{aligned} L(\alpha,\kappa,\lambda) &= \prod_{j=1}^{n}\left(f(y_j)\right)^{\delta_j}\left(S(y_j)\right)^{1-\delta_i} = \prod_{j=1}^{n}\left(h(y_j)\right)^{\delta_j}\left(S(y_j)\right) \\ &= \prod_{j=1}^{n}\left(e^{\lambda Z_j}\kappa\alpha y_j^{\alpha-1}\right)^{\delta_j}\left(\exp(-e^{\lambda Z_j}\kappa y_j^{\alpha})\right). \end{aligned} \tag{4.4}$$

The corresponding log-likelihood is

$$\begin{aligned} L &= \log L(\alpha,\kappa,\lambda) \\ &= (\log\kappa\alpha)\sum_{j=1}^{n}\delta_j + (\alpha-1)\sum_{j=1}^{n}\delta_j\log y_j + \lambda\sum_{j=1}^{n}\delta_j Z_j - \kappa\sum_{j=1}^{n}y_j^{\alpha}e^{\lambda Z_j} \\ &= r(\log\kappa\alpha) + (\alpha-1)\sum_{j=1}^{n}\delta_j\log y_j + \lambda\sum_{j=1}^{n}\delta_j Z_j - \kappa\sum_{j=1}^{n}y_j^{\alpha}e^{\lambda Z_j}. \end{aligned} \tag{4.5}$$

where $r=\sum_{j=1}^{n}\delta_j$. The maximum likelihood estimates for $\Theta=(\kappa,\alpha,\lambda)'$ are obtained by Newton-Raphson method as the solution of the score function $\mathrm{U}(\Theta)=0$, where $\mathrm{U}(\Theta)=\frac{\partial L}{\partial\Theta}$. This solution is then

$$\hat{\Theta}_{k+1} = \Theta_k + \mathrm{I}^{-1}(\Theta_k)\mathrm{U}(\Theta_k), \tag{4.6}$$

where $\hat{\Theta}_k$ at $k = 0$ is an initial guess and the $(l, m)^{\text{th}}$ element of the observed information matrix $\mathrm{I}(\hat{\Theta}) = \dfrac{-\partial^2 L(\Theta)}{\partial\Theta_l\Theta_m}$, $l = 1, 2, 3$ and $m = 1, 2, 3$.

Therefore, the score equation for scale parameter κ is

$$\mathbf{U}(\Theta)_\kappa = \frac{\partial \mathrm{L}(\Theta)}{\partial\kappa} = \sum_{j=1}^{\mathrm{n}}\delta_j \Big/ \kappa - \sum_{j=1}^{\mathrm{n}} \mathrm{y}_j^{\alpha} \mathrm{e}^{\lambda Z_j} \overset{set}{=} 0 \Rightarrow \hat{\kappa} = \sum_{j=1}^{\mathrm{n}}\delta_j \Big/ \sum_{j=1}^{n} \mathrm{y}_j^{\alpha} \mathrm{e}^{\lambda Z_j}, \tag{4.7}$$

the score equation for shape parameter α is

$$\mathbf{U}(\Theta)_\alpha = \frac{\partial L(\Theta)}{\partial\alpha} = \sum_{j=1}^{\mathrm{n}}\delta_j \Big/ \alpha - \sum_{j=1}^{\mathrm{n}}\delta_j \log y_j - \kappa\sum_{j=1}^{\mathrm{n}}(\mathrm{y}_j^{\alpha}\cdot\log y\cdot \mathrm{e}^{\lambda Z_j}), \tag{4.8}$$

the score equation for the common regression coefficient λ is

$$\mathbf{U}(\Theta)_\lambda = \frac{\partial \mathrm{L}(\Theta)}{\partial\lambda} = \sum_{j=1}^{\mathrm{n}}\delta_j \mathrm{Z_j} - \kappa\sum_{j=1}^{\mathrm{n}}(\mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\mathrm{Z}_j). \tag{4.9}$$

Then we establish the estimated covariance matrix of coefficients as

$$\mathrm{I}(\hat{\Theta}_{(k)})^{-1} = \begin{bmatrix} \hat{\mathrm{V}}(\hat{\kappa}) & \hat{\mathrm{cov}}(\hat{\kappa},\hat{\alpha}) & \hat{\mathrm{cov}}(\hat{\kappa},\hat{\lambda}) \\ \hat{\mathrm{cov}}(\hat{\alpha},\hat{\kappa}) & \hat{\mathrm{V}}(\hat{\alpha}) & \hat{\mathrm{cov}}(\hat{\alpha},\hat{\lambda}) \\ \hat{\mathrm{cov}}(\hat{\lambda},\hat{\kappa}) & \hat{\mathrm{cov}}(\hat{\lambda},\hat{\alpha}) & \hat{\mathrm{V}}(\hat{\lambda}) \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial^2\kappa} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\kappa\partial\alpha} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\kappa\partial\lambda} \\ -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\alpha\partial\kappa} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial^2\alpha} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\alpha\partial\lambda} \\ -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\lambda\partial\kappa} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial\lambda\partial\alpha} & -\dfrac{\partial^2 \mathrm{L}(\hat{\Theta})}{\partial^2\lambda} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sum_{j=1}^{\mathrm{n}}\delta_j \Big/ \kappa^2 & 0 & \sum_{j=1}^{\mathrm{n}} \mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\mathrm{Z}_j \\ 0 & \sum_{j=1}^{\mathrm{n}}\delta_j \Big/ \alpha^2 + \kappa\sum_{j=1}^{\mathrm{n}}[\mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\cdot(\ln y_j)^2] & \kappa\sum_{j=1}^{\mathrm{n}}(\mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\cdot\ln y_j\cdot \mathrm{Z}_j) \\ \sum_{j=1}^{\mathrm{n}} \mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\mathrm{Z}_j & \kappa\sum_{j=1}^{\mathrm{n}}(\mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}\ln y_j\cdot \mathrm{Z}_j) & \kappa\sum_{j=1}^{\mathrm{n}}[\mathrm{y}_j^{\alpha}\mathrm{e}^{\lambda Z_j}(\mathrm{Z}_j)^2] \end{bmatrix}^{-1} \tag{4.10}$$

The observed information matrix $I(\hat{\Theta})$ is obtained by replacing $\Theta = (\kappa, \alpha, \lambda)'$ by their MLEs $\hat{\Theta} = (\hat{\kappa}, \hat{\alpha}, \hat{\lambda})'$, respectively. Wald tests and large sample confidence limits for the individual parameters, such as λ, are readily computed using the large sample variance $\hat{\mathbf{V}}\left(\hat{\lambda}\right) = \left[I(\hat{\Theta})^{-1} \right]_{\lambda}$ obtained as the corresponding diagonal element of the estimated expected information, $I(\hat{\Theta})$.

The hypotheses is analogous to $H_0 : S_T = S_C$, equivalently, H_0: λ = 0 vs. H_a: λ ≠ 0. Test and interval estimates for regression parameter λ can be obtained by using the approximate normality of the MLE. The traditional z-test approach is to reject the null hypothesis at the α significance level if

$$z = \left| \frac{\hat{\lambda}}{se(\hat{\lambda})} \right| \geq z_{\alpha/2} \tag{4.11}$$

where $se(\hat{\lambda})$ is the square root of $\hat{\mathbf{V}}\left(\hat{\lambda}\right) = \left[I(\hat{\Theta})^{-1} \right]_{\lambda}$. Based on the above z-statistic, the corresponding $100(1-\alpha)\%$ confidence interval can be obtained as follows

$$\hat{\lambda} \pm z_{\alpha/2} se(\hat{\lambda}) \tag{4.12}$$

4.2 The Proposed Procedure

All patients randomized under the enrichment design have a positive diagnosis, but the true status of the molecular target for individual patients in the target clinical trial is in fact unknown. Let Y_{ij} denote the observable failure or censoring time for the jth individual (j=1,…, n_i; i=T, C), $\mathbf{Z}_{ij}$ is a vector of covariates associated with the jth individual. The survival function of Y is modeled by a two-component mixture model as

$$S(y_j; \mathbf{z}_j) = \gamma S_+(y_j; \mathbf{z}_j) + (1-\gamma) S_-(y_j; \mathbf{z}_j) \qquad j = 1, \ldots, n \tag{4.13}$$

and the corresponding probability density function of Y is

$$f(y_j;\mathbf{z}_j)=\gamma f_+(y_j;\mathbf{z}_j)+(1-\gamma)f_-(y_j;\mathbf{z}_j) \qquad j=1,\ldots,n \tag{4.14}$$

where γ is a positive predicted value (PPV), the proportion of patients truly with the molecular targets, and $S_g(y_j;\mathbf{z}_j)$ and $f_g(y_j;\mathbf{z}_j)$ is the conditional survival function and conditional density function of the *g*th component (g=+, −). Under the proportional hazards assumption, the conditional hazard function for the *g*th component is given by

$$h_g(y_{ij})=h_{g0}(y_{ij})\exp(\eta(\mathbf{z}_{ij})) \tag{4.15}$$

where $h_{g0}(y_{ij})$ is the baseline hazard function and $\eta(\mathbf{z}_{ij})$ is the linear predictor relating to the covariate $\mathbf{Z}_{ij}$. The commonly used Weibull distribution maybe assumed for $h_{g0}(y_{ij})$ because it is flexible as either a monotonic increasing, constant, or monotonic decreasing baseline hazard. That is,

$$h_{g0}(y_j)=\kappa_g\alpha_g y_j^{\alpha_g-1} \tag{4.16}$$

where $\kappa_g,\alpha_g>0$ are unknown parameters

Under proportional hazards model and the Weibull distribution assumption,

$$h_g(y_j)=h_{g0}(y_j)\exp(\lambda_g\mathbf{z}_j)=\kappa_g\alpha_g y_j^{\alpha_g-1}\exp(\lambda_g\mathbf{z}_j) \quad g=+,-\ ;\ j=1,\ldots,n \tag{4.17}$$

where λ_g is the vector of regression coefficients. Under the formulation above, the vector of unknown parameters becomes

$$\boldsymbol{\psi}=(\gamma,\lambda_+,\lambda_-,\kappa_+,\kappa_-,\alpha_+,\alpha_-)'$$

On the basis of the observed data, the log-likelihood function for $\boldsymbol{\psi}$ under the mixture model is given by

$$\begin{aligned}\log L_c(\boldsymbol{\Psi})=\sum_{j=1}^{n}\Big\{ & x_{j+}\left[\log\gamma+\delta_j\log f_+(y_j;\mathbf{z}_j)+(1-\delta_j)\log S_+(y_j;\mathbf{z}_j)\right]\\ & +x_{j-}\left[\log(1-\gamma)+\delta_j\log f_-(y_j;\mathbf{z}_j)+(1-\delta_j)\log S_-(y_j;\mathbf{z}_j)\right]\end{aligned} \tag{4.18}$$

$$x_j=(x_{j+},x_{j-})^T \text{ and } x_{j+}+x_{j-}=1;\quad j=1,\ldots,n$$

where $\delta_j=1$ and $\delta_j=0$ indicate a failure and a censored observation respectively. The best linear unbiased predictor (BLUP) estimate is obtained as a solution of the equation, which can be solved via the EM algorithm as presented below.

In order to pose the estimation procedure as an incomplete-data problem, an unobservable random vector $X_{ij} = (x_{ij+}, x_{ij-})^T$ is introduced, indicating whether the observation y_{ij} belongs to the positive or negative component. On the $(k+1)$th iteration, the E-step of the EM algorithm involves the calculation of the Q-function, which is the expectation of the complete-data log-likelihood conditional on the current estimate of the parameter and the observed data. In particular, the Q-function can be decomposed as

$$Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(k)}) = Q_{\gamma}^{(k)} + Q_{\xi_+}^{(k)} + Q_{\xi_-}^{(k)} \tag{4.19}$$

with respect to the parameters γ, $\boldsymbol{\xi}_+ = (\lambda_+, \kappa_+, \alpha_+)'$, and $\boldsymbol{\xi}_- = (\lambda_-, \kappa_-, \alpha_-)'$ respectively.

The E-step involves the calculation of $Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(k)}) = Q_{\gamma}^{(k)} + Q_{\xi_+}^{(k)} + Q_{\xi_-}^{(k)}$, where

$$\begin{aligned}
Q_{\gamma}^{(k)} &= \sum_{j=1}^{n} \left[x_j^{(k)} \log\left(\frac{\hat{\gamma}}{1-\hat{\gamma}}\right) + \log(1-\hat{\gamma}) \right] \\
Q_{\xi_+}^{(k)} &= \sum_{j=1}^{n} \left\{ x_j^{(k)} \left[\delta_j \log f_+(y_j; \mathbf{z}_j) + (1-\delta_j) \log S_+(y_j; \mathbf{z}_j) \right] \right\} \\
Q_{\xi_-}^{(k)} &= \sum_{j=1}^{n} \left\{ \left(1 - x_j^{(k)}\right) \left[\delta_j \log f_-(y_j; \mathbf{z}_j) + (1-\delta_j) \log S_-(y_j; \mathbf{z}_j) \right] \right\}
\end{aligned} \tag{4.20}$$

where

$$x_j^{(k)} = E_{\psi^{(k)}}(x_j \mid y_j, \mathbf{z}_j) = \frac{\gamma^{(k)} (f_+^{(k)})^{\delta_j} (S_+^{(k)})^{(1-\delta_j)}}{\gamma^{(k)} (f_+^{(k)})^{\delta_j} (S_+^{(k)})^{(1-\delta_j)} + (1-\gamma^{(k)}) (f_-^{(k)})^{\delta_j} (S_-^{(k)})^{(1-\delta_j)}} \tag{4.21}$$

The M-step provides the updated estimate $\boldsymbol{\psi}^{(k+1)}$ that maximizes $Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(k)})$ with respect to $\boldsymbol{\psi}$ and thus involves solving the non-linear equations

$$\begin{aligned}
&\lambda_g (+: g=1;\ -: g=2): \sum_{j=1}^{n} (x_j^{(k)})^{(2-g)} (1-x_j^{(k)})^{(g-1)} \left[\delta_j + \log S_g(y_j; z_j) \right] z_j = 0 \\
&\kappa_g (+: g=1;\ -: g=2): \sum_{j=1}^{n} (x_j^{(k)})^{(2-g)} (1-x_j^{(k)})^{(g-1)} \left[\frac{\delta_j}{\lambda_g} - \exp(\beta z_j) y_j^{\alpha_g} \right] = 0 \\
&\alpha_g (+: g=1;\ -: g=2): \sum_{j=1}^{n} (x_j^{(k)})^{(2-g)} (1-x_j^{(k)})^{(g-1)} \left[\frac{\delta_j + (\delta_j \alpha_g - h_g(y_j; z_j) y_j) \log y_j}{\alpha_g} \right] = 0
\end{aligned} \tag{4.22}$$

and the following closed-form equation for γ:

$$\gamma^{(k+1)} = \sum_{j=1}^{n} x_j^{(k)} \Big/ n \qquad (4.23)$$

With respect to $\boldsymbol{\Psi}$, which involves solving a set of non-linear equations and the MINPACK routine HYBRD1 (More, et al., 1980) is used for this purpose.

The estimation procedure of the EM-based approach is summarized as follows

1. Set initial values for $\gamma^{(0)}$, $\boldsymbol{\xi}_+^{(0)}$, $\boldsymbol{\xi}_-^{(0)}$,

2. Calculate x_j using (4.21) and update $\boldsymbol{\xi}_g (+ : g = 1;\ - : g = 2)$ by (4.22), and update γ by (4.23).

3. Repeat Step 2 until convergence.

4. The standard errors of the maximum likelihood estimator $\hat{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$ is assessed using the bootstrap methodology of Efron (1979, 1982). A number *K* of independent bootstrap samples are obtained with each being randomly drawn with replacement from the observed data (Y_j, δ_j, Z_j), j = 1,…, *n*.

The null hypothesis is rejected and the efficacy of the molecular targeted test drug is different from that of the control in the patient population truly with the molecular target at the α significance level if

$$z = \left| \frac{\hat{\lambda}}{\sqrt{S_B^2}} \right| \geq z_{\alpha/2},$$

where $z_{\alpha/2}$ is the α/2 upper percentile of a standard normal distribution, and S_B^2 denote the estimator of the variance of $\hat{\lambda}$ obtained by the bootstrap procedure.

The corresponding 100(1 – α)% asymptotic confidence interval for λ can be constructed as $\hat{\lambda} \pm z_{\alpha/2}\sqrt{S_B^2}$.

Chapter 5 Simulation Studies

In this section, the simulation studies were conducted to empirically investigate and compare performance of the proposed methods with the current methods for the inference of the treatment effects of the targeted treatment. FORTRAN 95 and IMSL's STAT/LIBRARY FORTRAN subroutines were used in the simulation study.

5.1 The Exponential Distribution Model

5.1.1 Simulation Procedure

The random samples of patient units with or without the molecular target were generated from the Bernoulli distribution with probability γ. Then the units are randomized in a 1:1 ratio to the test group or control group. Exponential random deviates were generated with the specified parameters λ_{i+} and λ_{i-} according the status of molecular target, i = T, C. For the purpose of illustration, we assume that the placebo control is employed in the targeted clinical trial. It is presumed that it is not efficacious in the patients with and without the molecular target. In addition the molecularly targeted test drug is not effective in the patients truly without the target either. Therefore, for simplicity, in the simulation, λ_{T-}, λ_{C+}, and λ_{C-} are assumed equal and set to be a generic value of 1. To investigate the impact of the PPV, sample size, hazard ratio, and variability, we consider the following specifications of parameters in the simulation.

The PPV is set to be 0.5, 0.6, 0.7, 0.8, and 0.8 which reflect a range of low, median, and high positive predicted value. We use random right censoring. Each unit has a potential censoring time C_i' and a potential lifetime T_i', which are assumed to be independent

random variables. Consider $Y_i = \min\{C_i', T_i'\}$ and an indicator δ_i for the type of event (censored or death). If T' and C' are independent exponential random variables with parameters λ_1 and λ_2, respectively, then $P\{T' < C'\} = \lambda_2 / (\lambda_1 + \lambda_2)$. By matching survival and censoring times in n pairs in sequence, we generate the observed times, y = min(t', c'). The censoring proportions considered in the simulation study are 0, 0.1, 0.2, 0.3 and 0.4. To investigate the finite sample properties, the sample sizes are set as 300, 600, and 900 per group. The power of the proposed testing procedure was investigate at λ_{T+} = 0.70, 0.75, 0.80 and 0.85. For each combination, 5000 random samples were generated and the number of the bootstrap samples was set to be 1000. Furthermore, in the simulation, we employed the traditional sample mean survival time of test and control treatments, $\overline{y}_T$ and $\overline{y}_C$, the inverse of hazard, as the initial values for $\hat{\lambda}_{T+}$ and $\hat{\lambda}_{C+}$, respectively, i=T, C.

For estimation, we investigate the bias of the estimators and the coverage probability of the 95% confidence interval. For hypothesis testing, the performance measures include empirical size and power. The bias is estimated as the average of the differences between the estimates and the true value of θ over 5000 simulated samples. The coverage probability is calculated as the proportion of the 5000 95% confidence intervals that contain θ. The size and power were computed as the proportion of the 5000 samples that the null hypothesis is rejected for the two-sided test at the 5% significance level. For a 95% confidence level, with 5000 simulation random samples implies that 95% of the empirical coverage probabilities will be within 0.94396 and 0.95604 if the proposed methods provide sufficient coverage probability. The limit is computed as

$$0.95 \pm Z_{0.05/2}\sqrt{\frac{(0.95)(0.05)}{5000}} = (0.94396, 0.95604).$$

In addition, for a 5 % nominal significance level, a simulation study with 5000 random samples implies that 95% of empirical sizes will be within 0.04396 and 0.05604 if the proposed methods can adequately control the size at the nominal level of 0.05. The limit is computed as

$$0.05 \pm Z_{0.05/2}\sqrt{\frac{(0.95)(0.05)}{5000}} = (0.04396, 0.05604).$$

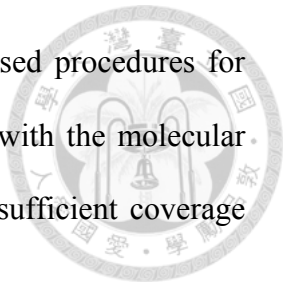
The flow chart for simulation study is given in Figure 5.1.

5.1.2 Simulation Results

Relative Bias and Coverage Probability

The simulation results on estimation are provided in Table 5.1. The results in Table 5.1 demonstrate that the absolute relative bias of the estimator for traditional hazard ratio for the patients truly with the molecular target by the current method ranges from 3.1% to more than 17%. It increases as the PPV decreases. On the other hand, the absolute relative bias of the estimator for hazard ratio for the patients truly with the molecular target obtained by the EM algorithm does not exceed 5.0% while most of them are smaller than 3.0%. The variability has little impact on the bias of both methods. Consequently, the empirical coverage probabilities of the corresponding 95% confidence interval constructed by the current method can be as low as 56.1% when the PPV is 50%, censored rate is 20%, hazard ratio is 0.70 and n is 300. The coverage probability of the 95% confidence interval by the current method is an increasing function of the PPV. None of the coverage probabilities of the 95% confidence intervals by the current method in Table 5.1 exceed 0.95. On the contrary, no coverage

probability of the EM method is below 0.95. Therefore, the proposed procedures for estimation of the treatment effects in the patients population truly with the molecular target by the EM algorithm is not only unbiased but also provide sufficient coverage probability.



Size and Power

The simulation results on the empirical sizes are provided in Table 5.2. The simulation results on sizes reveal that all empirical sizes of both the current method and the proposed EM procedure for testing the hypothesis are within 0.0452 and 0.0610. These results demonstrate that both methods can adequately control the size at its nominal level of 5% under the null hypothesis. The empirical powers of the simulation for the hazard ratio are given in Table 5.3. In addition, Figure 5.2 presents the power curves when n = 300, censored rate = 10% and PPV is 0.6. From Table 5.3, we observe that the power of the current method is an increasing function of the PPV. For both the methods, the power increases as the sample size increases. However, the simulation results clearly demonstrate that the proposed testing procedure for the treatment effects based on the EM algorithm in the patient population truly with the molecular target is uniformly more powerful than the current method as depicted in Figure 5.2.

5.2 The Parametric Proportional Hazard Regression Model

5.2.1 Simulation Procedure

The random samples of patient units with or without the molecular target were generated from the Bernoulli distribution with probability γ. Then the units are randomized in a 1:1 ratio to the test group or control group. Weibull random deviates

are generated with the specified parameters $\lambda_{i+}, \lambda_{i-}, \kappa_{i+}, \kappa_{i-}, \alpha_{i+}$ and α_{i-} according the status of molecular target, i = T, C. We also assume that the placebo control is employed in the targeted clinical trial. It presumed that it is not efficacious in the patients with and without the molecular target. In addition the molecularly targeted test drug is not effective in the patients truly without the target either. Therefore, for simplicity, in the simulation, κ_+, κ_-, α_+, α_-, λ_- are assumed equal and set to be a generic value of 1.

To investigate the impact of the PPV and censored rate, we consider the following specifications of parameters in the simulation. The PPV is set to be 0.5, 0.6, 0.7 and 0.8 which reflect a range of low and high positive predicted value. We use random right censoring. Each unit has a potential censoring time C_i' and a potential lifetime T_i', which are assumed to be independent random variables. Consider $Y_i = \min\{C_i', T_i'\}$ and an indicator δ_i for the type of event (censored or death). If T' and C' are two independent Weibull distributions with different scale parameters but having the same shape parameter, T' and C' follow Weibull(θ^*, λ_1) and Weibull(θ^*, λ_2) respectively, then $P\{T' < C'\} = \lambda_2 / (\lambda_1 + \lambda_2)$. By matching survival and censoring times in n pairs in sequence, we generate the observed times, y = min(t', c'). The censoring proportions considered in the simulation study are 0 and 0.2. The hazard ratio is set as 0.75 for bias, coverage probability and power. To investigate the finite sample properties, the sample sizes are set as 600 per group. The 5000 random samples were generated. For estimation, we investigate the bias of the estimators and the coverage probability of the 95% confidence interval. For hypothesis testing, the performance measures include empirical size and power. The bias is estimated as the average of the differences between the estimates and the true value of θ over 5000 simulated samples. The coverage probability is calculated as the proportion of the 5000 95% confidence intervals that contains θ. The

size and power were computed as the proportion of the 5000 samples that the null hypothesis is rejected for the two-sided test at the 5% significance level. For a 95% confidence level, with 5000 simulation random samples implies that 95% of the empirical coverage probabilities will be within 0.94396 and 0.95604 if the proposed methods provide sufficient coverage probability.

5.2.2 Simulation Results

Relative Bias and Coverage Probability

The simulation results on relative bias and coverage probabilities are provided in Table 5.4 show that the relative bias of the current approach ranges from 17.45% to 4.49%. On the other hand, the EM procedure is nearly unbiased with the absolute relative bias smaller than 5%. The relative bias is a decreasing function of PPV and censoring rate. The empirical coverage probability of the 95% confidence interval for the hazard ratio by the current method can be as low as 45.12%. It reaches to 99.88% when PPV is 0.8. On the other hand, the EM method provides an empirical coverage probability of the 95% confidence interval for the hazard ratio above 95%. In summary, the proposed EM procedure not only is nearly unbiased but also its corresponding 95% confidence interval for hazard ratio provides sufficient coverage probability.

Size and Power

Table 5.5 presents the empirical sizes for the parametric proportional hazard model. The results demonstrate that both the current and EM method can adequately control the size at the nominal level of 5% under the null hypothesis. The results of the empirical powers for the parametric proportional hazard model is given in Table 5.6. The results given in Table 5.6 reveal that the empirical power is an increasing function of PPV and

a decreasing function of censoring rate. However, the proposed EM procedure is more powerful than the current method. For PPV=0.5, the empirical power of the EM procedure is 20% more than that of the current method for censoring rate being either 0 or 0.2. In summary, the proposed EM procedure not only can control the size at its nominal level but also is more powerful than the current method under the parametric proportional hazard model.

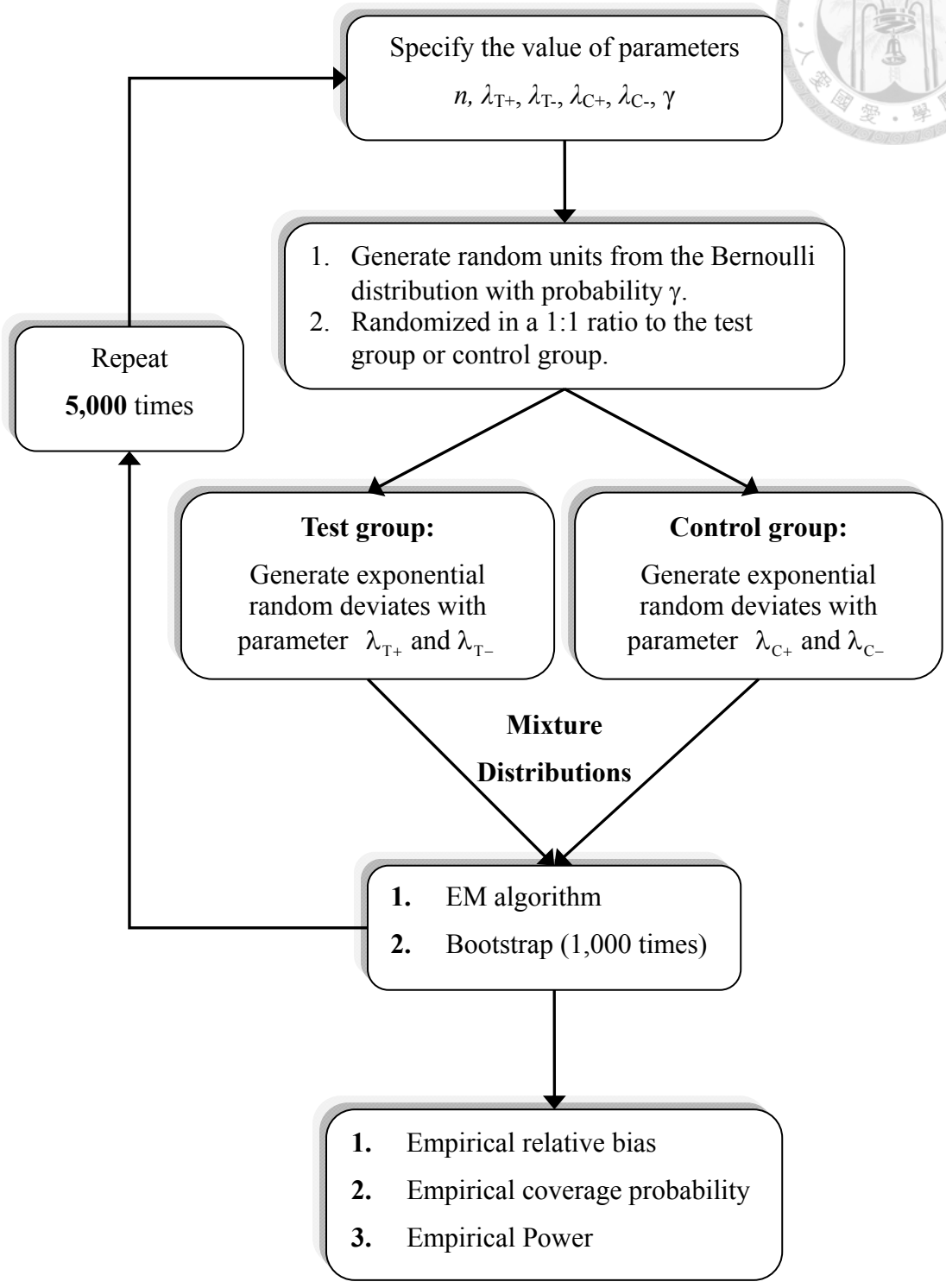


Figure 5.1 Flow chart of the simulation study to exponential distribution

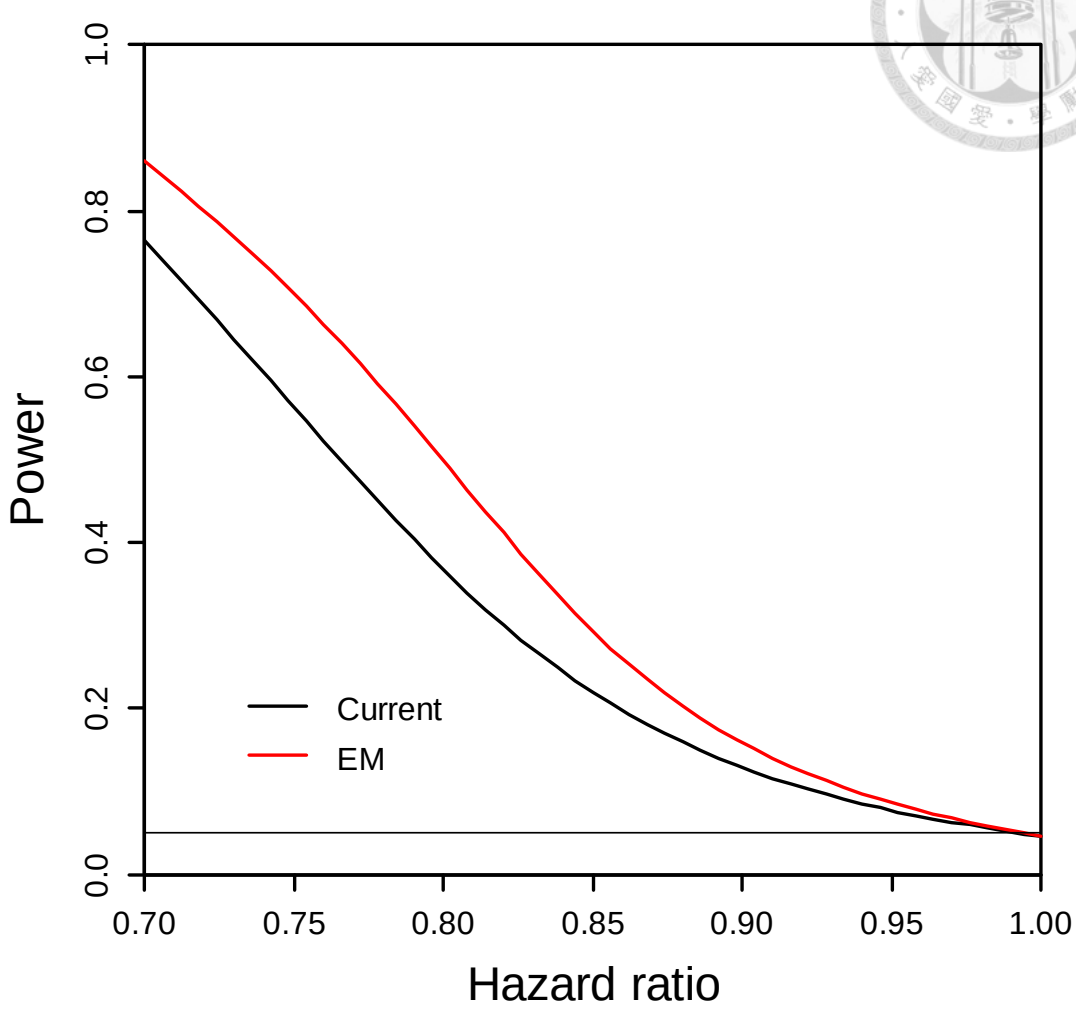


Figure 5.2 The empirical power curve when the PPV is 0.6, n=300 and CR=10%

Table 5.1 Relative bias (%) and the coverage probability to exponential distribution

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
300	0.85	0.0	8.21[a]	-2.41	6.60	-1.87	4.92	-1.51	3.05	-1.31
			0.8316[b]	0.9682	0.8800	0.9648	0.9150	0.9640	0.9318	0.9562
		0.1	8.01	-3.11	6.38	-2.61	4.76	-2.00	3.11	-1.54
			0.8538	0.9742	0.8804	0.9636	0.9102	0.9592	0.9318	0.9594
		0.2	7.96	-3.73	6.42	-3.14	4.60	-2.52	3.13	-1.74
			0.8632	0.9768	0.8942	0.9732	0.9202	0.969	0.9360	0.9610
		0.3	8.24	-4.35	6.20	-3.85	4.74	-2.88	3.19	-1.76
			0.8652	0.9758	0.9094	0.9736	0.9238	0.963	0.9396	0.9630
		0.4	8.13	-4.99	6.38	-3.94	4.82	-2.88	3.19	-1.98
			0.8820	0.9758	0.9146	0.9732	0.9256	0.9662	0.9440	0.9646
	0.80	0.0	11.28	-1.51	8.59	-1.60	6.29	-1.50	3.96	-1.46
			0.7386	0.9672	0.8132	0.9628	0.8772	0.9598	0.9178	0.9536
		0.1	11.29	-2.41	8.70	-2.26	6.56	-1.61	4.33	-1.45
			0.7620	0.9730	0.8398	0.9684	0.8822	0.9662	0.917	0.9566
		0.2	11.01	-3.78	8.52	-2.95	6.37	-2.46	4.12	-2.05
			0.7874	0.9760	0.8548	0.9732	0.8978	0.9696	0.9290	0.9604
		0.3	11.05	-4.84	8.75	-3.74	6.46	-2.84	4.24	-2.26
			0.8100	0.9782	0.8602	0.9682	0.9048	0.9674	0.9322	0.9630
		0.4	11.39	-5.14	8.77	-4.56	6.27	-3.64	4.30	-2.58
			0.8266	0.9752	0.8738	0.9748	0.9052	0.9660	0.927	0.9628
	0.75	0.0	14.19	-0.92	11.04	-0.73	7.92	-1.04	5.16	-1.11
			0.6264	0.9648	0.7594	0.9572	0.8432	0.9610	0.9018	0.9566
		0.1	14.39	-1.63	11.03	-1.88	8.07	-1.39	5.49	-1.17
			0.6494	0.9758	0.7680	0.9720	0.8510	0.9700	0.8974	0.9592
		0.2	14.33	-2.93	11.13	-2.33	8.21	-2.32	5.12	-2.04
			0.6870	0.9752	0.7926	0.9724	0.8564	0.9674	0.9174	0.966
		0.3	14.43	-4.29	11.13	-3.28	7.99	-2.81	5.25	-2.47
			0.7140	0.9750	0.8064	0.9738	0.8760	0.9680	0.9136	0.9662
		0.4	14.24	-5.72	10.85	-4.76	7.95	-3.96	5.45	-2.80
			0.7530	0.9764	0.8352	0.9762	0.8822	0.9688	0.9162	0.9622
	0.70	0.0	17.70	0.24	13.53	-0.21	9.77	-0.57	6.41	-0.73
			0.4860	0.9566	0.6498	0.9598	0.7870	0.9582	0.8796	0.9568
		0.1	17.73	-1.00	13.57	-1.19	9.87	-1.07	6.47	-1.09
			0.5292	0.9676	0.6826	0.9662	0.7996	0.9630	0.8848	0.9604
		0.2	17.90	-1.76	13.59	-2.13	10.17	-1.79	6.36	-1.79
			0.5610	0.9772	0.7070	0.9742	0.9096	0.9690	0.8930	0.9604
		0.3	17.87	-3.76	13.61	-3.11	9.9	-2.87	6.41	-2.54
			0.6114	0.9780	0.7390	0.9766	0.8348	0.9682	0.8986	0.9616
		0.4	17.67	-5.89	13.93	-4.99	9.61	-4.49	6.53	-2.86
			0.6558	0.9740	0.7648	0.9740	0.8562	0.9636	0.9020	0.9688

HR: hazard ratio; CR: censoring rate

a: Relative bias (%); b: Coverage probability

Table 5.1 Relative bias (%) and the coverage probability to exponential distribution (continued)

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
600	0.85	0.0	8.27[a]	-1.85	6.32	-1.72	4.76	-1.29	3.12	-1.06
			0.7118[b]	0.9510	0.8126	0.9504	0.8668	0.9484	0.9162	0.9470
		0.1	8.12	-2.56	6.35	-2.06	5.07	-1.38	3.12	-1.22
			0.7414	0.9482	0.8228	0.9444	0.8750	0.9510	0.9246	0.9524
		0.2	8.06	-2.99	6.49	-2.25	4.72	-1.95	3.08	-1.41
			0.7684	0.9488	0.8318	0.9486	0.8934	0.9482	0.9202	0.9484
		0.3	8.02	-3.67	6.31	-2.86	4.72	-2.44	2.99	-1.53
			0.7972	0.9412	0.8606	0.9382	0.8990	0.9200	0.9286	0.9470
		0.4	8.19	-3.92	6.29	-3.29	4.54	-2.58	3.16	-1.53
			0.8126	0.938	0.8798	0.9314	0.9090	0.9374	0.9312	0.9508
	0.80	0.0	11.09	-0.93	8.67	-0.91	6.44	-0.86	4.21	-0.88
			0.5532	0.9550	0.6856	0.9494	0.8100	0.9502	0.8940	0.9522
		0.1	11.1	-1.73	8.89	-1.39	6.49	-1.29	4.25	-1.03
			0.5914	0.9554	0.7094	0.9586	0.8180	0.9516	0.8944	0.9536
		0.2	11.1	-2.71	8.65	-2.29	6.50	-1.63	4.05	-1.53
			0.6314	0.9532	0.7468	0.9506	0.834	0.9512	0.9084	0.9546
		0.3	10.94	-3.65	8.64	-3.00	6.40	-2.54	4.04	-1.81
			0.6778	0.9460	0.7706	0.9494	0.8496	0.9446	0.9152	0.9504
		0.4	10.90	-4.66	8.52	-3.89	6.30	-2.98	4.39	-1.90
			0.7158	0.9350	0.8084	0.9312	0.8690	0.9410	0.9110	0.9448
	0.75	0.0	14.36	-0.12	11.2	-0.24	8.07	-0.48	5.17	-0.85
			0.3564	0.9494	0.5452	0.954	0.7340	0.9576	0.8610	0.9514
		0.1	14.37	-1.12	11.28	-0.95	8.12	-1.12	5.27	-0.93
			0.3968	0.9582	0.5768	0.9514	0.7446	0.9472	0.8628	0.9474
		0.2	14.39	-2.05	11.27	-1.65	8.01	-1.71	5.19	-1.37
			0.4478	0.9602	0.6164	0.9484	0.7794	0.9514	0.8762	0.9538
		0.3	14.21	-3.6	11.05	-2.83	8.23	-2.19	5.35	-1.65
			0.5084	0.9478	0.6644	0.9478	0.7862	0.9432	0.8854	0.954
		0.4	8.27	-1.85	6.32	-1.72	4.76	-1.29	3.12	-1.06
			0.7118	0.9510	0.8126	0.9504	0.8668	0.9484	0.9162	0.947
	0.70	0.0	17.63	0.36	13.61	0.21	9.79	-0.20	6.43	-0.30
			0.2002	0.9470	0.4054	0.9464	0.6304	0.9502	0.8106	0.9478
		0.1	17.59	-0.93	13.41	-0.84	9.96	-0.63	6.36	-0.81
			0.2382	0.9480	0.4554	0.9504	0.6500	0.9554	0.8260	0.9474
		0.2	17.66	-1.71	13.67	-1.67	9.86	-1.50	6.36	-1.34
			0.2864	0.9572	0.4886	0.9556	0.6896	0.9516	0.8382	0.9514
		0.3	17.77	-3.13	13.71	-2.63	10.13	-2.04	6.43	-1.69
			0.3426	0.9508	0.5460	0.9460	0.7146	0.9522	0.8464	0.9496
		0.4	17.76	-4.84	13.56	-4.09	9.87	-3.26	6.59	-2.26
			0.4156	0.9368	0.6050	0.9344	0.7544	0.9384	0.8606	0.9474

HR: hazard ratio; CR: censoring rate
a: Relative bias (%); b: Coverage probability

Table 5.1 Relative bias (%) and the coverage probability to exponential distribution (continued)

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
900	0.85	0.0	8.18[a]	-1.34	6.44	-1.27	4.67	-1.16	2.99	-1.05
			0.6178[b]	0.9456	0.7394	0.9420	0.8416	0.9458	0.9000	0.9452
		0.1	7.99	-2.27	6.49	-1.61	4.75	-1.40	3.07	-1.04
			0.6524	0.9374	0.7538	0.9374	0.8480	0.9402	0.9092	0.9406
		0.2	8.02	-2.61	6.51	-2.13	4.73	-1.71	3.20	-1.06
			0.6844	0.931	0.7752	0.9286	0.8698	0.936	0.9096	0.9486
		0.3	8.09	-3.25	6.32	-2.61	4.6	-1.99	3.02	-1.39
			0.7136	0.9128	0.81	0.9278	0.8748	0.9332	0.9158	0.9442
		0.4	8.16	-3.56	6.46	-2.81	4.78	-2.06	3.14	-1.41
			0.748	0.9182	0.8264	0.9214	0.8786	0.9346	0.9156	0.9404
	0.80	0.0	11.05	-0.88	8.67	-0.88	6.47	-0.7	4.1	-0.8
			0.3978	0.9466	0.5818	0.947	0.731	0.946	0.8624	0.949
		0.1	11.09	-1.4	8.84	-1.16	6.44	-1.61	4.16	-1.03
			0.4382	0.946	0.6076	0.9488	0.777	0.9392	0.8706	0.9478
		0.2	11.08	-2.18	8.65	-1.96	6.52	-1.40	4.11	-1.29
			0.4830	0.9354	0.6536	0.9358	0.796	0.9540	0.8800	0.9442
		0.3	11.21	-2.86	8.69	-2.71	6.19	-2.25	4.30	-1.49
			0.5234	0.9252	0.6764	0.9238	0.8090	0.9388	0.8836	0.9476
		0.4	10.93	-4.28	8.65	-3.46	6.40	-2.48	4.05	-2.04
			0.6060	0.9180	0.7236	0.9188	0.8302	0.9322	0.8996	0.9410
	0.75	0.0	14.16	-0.16	11.09	-0.33	8.15	-0.33	5.27	-0.48
			0.2006	0.945	0.3938	0.9466	0.6142	0.9464	0.8042	0.9478
		0.1	14.39	-0.91	11.19	-0.68	8.20	-0.81	5.28	-0.91
			0.2364	0.9416	0.4360	0.9410	0.6454	0.9506	0.828	0.9474
		0.2	14.21	-1.80	11.08	-1.69	8.11	-1.41	5.19	-1.24
			0.2962	0.9372	0.4906	0.9410	0.6702	0.9466	0.8336	0.9466
		0.3	14.29	-2.93	10.92	-2.65	8.08	-1.95	5.31	-1.59
			0.3436	0.9342	0.5450	0.9298	0.7128	0.9414	0.8434	0.9406
		0.4	14.32	-4.15	11.08	-3.37	8.12	-2.60	5.40	-1.97
			0.4036	0.9194	0.5892	0.9170	0.7516	0.9278	0.8542	0.9388
	0.70	0.0	17.56	0.26	13.61	0.19	9.79	-0.13	6.47	-0.14
			0.0720	0.9376	0.2310	0.9382	0.4928	0.9386	0.7314	0.9502
		0.1	17.56	-0.79	13.64	-0.66	10.07	-0.36	6.41	-0.64
			0.1032	0.9444	0.2696	0.9492	0.5170	0.9540	0.7560	0.9496
		0.2	17.71	-1.77	13.49	-1.60	9.94	-1.31	6.39	-1.19
			0.1364	0.9350	0.3300	0.9396	0.5666	0.9454	0.7824	0.944
		0.3	17.6	-3.01	13.77	-2.40	9.86	-2.14	6.44	-1.53
			0.1814	0.9290	0.3798	0.9346	0.6100	0.9310	0.7938	0.9472
		0.4	17.70	-4.59	13.71	-3.66	9.84	-3.01	6.36	-2.26
			0.2324	0.9130	0.4350	0.9216	0.6600	0.9314	0.8220	0.9360

HR: hazard ratio; CR: censoring rate
a: Relative bias (%); b: Coverage probability

Table 5.2 Comparison of empirical sizes to exponential distribution

		PPV							
		0.5		0.6		0.7		0.8	
N	CR	Current	EM	Current	EM	Current	EM	Current	EM
300	0.0	0.0490	0.0486	0.0568	0.0582	0.0488	0.0490	0.0476	0.0474
	0.1	0.0504	0.0496	0.0454	0.0456	0.0506	0.0494	0.0484	0.0476
	0.2	0.0550	0.0540	0.0482	0.0490	0.0524	0.0532	0.0488	0.0482
	0.3	0.0536	0.0514	0.0508	0.0484	0.0514	0.0518	0.0520	0.0516
	0.4	0.0534	0.0516	0.0474	0.0488	0.0480	0.0494	0.0496	0.0480
600	0.0	0.0468	0.0474	0.0518	0.0518	0.0480	0.0474	0.0534	0.0544
	0.1	0.0538	0.0540	0.0502	0.0506	0.0464	0.0452	0.0540	0.0558
	0.2	0.0510	0.0514	0.0492	0.0488	0.0484	0.0492	0.0540	0.0550
	0.3	0.0480	0.0484	0.0520	0.0518	0.0508	0.0496	0.0506	0.0508
	0.4	0.0538	0.0538	0.0610	0.0600	0.0512	0.0516	0.0478	0.0484
900	0.0	0.0500	0.0500	0.0496	0.0492	0.0512	0.0502	0.0544	0.0534
	0.1	0.0564	0.0552	0.0534	0.0526	0.0502	0.0496	0.0494	0.0490
	0.2	0.0464	0.0466	0.0476	0.0484	0.0506	0.0516	0.0572	0.0568
	0.3	0.0536	0.0538	0.0572	0.0580	0.0526	0.0546	0.0546	0.0546
	0.4	0.0526	0.0514	0.0540	0.0530	0.0560	0.0556	0.0536	0.0532

CR: censoring rate

Table 5.3 Comparison of empirical powers to exponential distribution

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
300	0.85	0.0	0.1808	0.2980	0.2294	0.3498	0.2840	0.3920	0.3774	0.4610
		0.1	0.1626	0.2312	0.2198	0.2932	0.2662	0.3414	0.3334	0.3998
		0.2	0.1586	0.1592	0.1998	0.2248	0.2492	0.2934	0.3046	0.3418
		0.3	0.1412	0.1280	0.1866	0.1904	0.2214	0.2490	0.2574	0.2852
		0.4	0.1290	0.0984	0.1576	0.1418	0.1996	0.2038	0.2324	0.2480
	0.80	0.0	0.2964	0.4840	0.4190	0.5800	0.5110	0.6570	0.6178	0.7182
		0.1	0.2712	0.3836	0.3672	0.4984	0.4546	0.5650	0.5506	0.6372
		0.2	0.2626	0.2870	0.3384	0.3760	0.4242	0.4900	0.5138	0.5824
		0.3	0.2262	0.2378	0.2946	0.3170	0.3724	0.4170	0.4536	0.5064
		0.4	0.1956	0.1670	0.2676	0.2650	0.3382	0.3546	0.4032	0.4360
	0.75	0.0	0.4764	0.6802	0.6112	0.7768	0.7372	0.8472	0.8252	0.8946
		0.1	0.4288	0.5726	0.5636	0.7006	0.6722	0.7786	0.7730	0.8396
		0.2	0.3930	0.4300	0.5118	0.5684	0.6268	0.6976	0.7432	0.8042
		0.3	0.3500	0.3504	0.4574	0.4968	0.5774	0.6222	0.6836	0.7286
		0.4	0.3100	0.2902	0.4224	0.4230	0.5144	0.5540	0.5994	0.6534
	0.70	0.0	0.6558	0.8348	0.7940	0.9102	0.8970	0.9558	0.9466	0.9740
		0.1	0.6166	0.7396	0.7638	0.8592	0.8556	0.9204	0.9270	0.9600
		0.2	0.5552	0.5864	0.7008	0.7324	0.8056	0.8470	0.8930	0.9252
		0.3	0.5044	0.5098	0.6496	0.6760	0.7600	0.7906	0.8562	0.8822
		0.4	0.4560	0.4366	0.5652	0.5888	0.7100	0.7504	0.7944	0.8256
600	0.85	0.0	0.3060	0.5256	0.4202	0.6082	0.5204	0.6788	0.6292	0.7398
		0.1	0.2866	0.4538	0.3846	0.5452	0.4550	0.6080	0.5796	0.6882
		0.2	0.2652	0.3822	0.3364	0.4700	0.4292	0.5688	0.5340	0.6342
		0.3	0.2372	0.3154	0.3106	0.4040	0.3790	0.4840	0.4914	0.5762
		0.4	0.1980	0.2384	0.2738	0.3296	0.3524	0.4184	0.4210	0.4962
	0.80	0.0	0.5300	0.7722	0.6690	0.8442	0.7974	0.9042	0.8864	0.9472
		0.1	0.4952	0.7032	0.6174	0.8052	0.7444	0.8692	0.8486	0.9136
		0.2	0.4506	0.6464	0.5838	0.7576	0.6944	0.8236	0.8138	0.8936
		0.3	0.4072	0.5608	0.5264	0.6876	0.6436	0.7778	0.7540	0.8402
		0.4	0.3638	0.4570	0.4732	0.5976	0.5960	0.7080	0.6706	0.7628
	0.75	0.0	0.7514	0.9120	0.8830	0.9634	0.9482	0.9874	0.9834	0.9948
		0.1	0.7178	0.8858	0.8430	0.9440	0.9276	0.9770	0.9704	0.9882
		0.2	0.6544	0.8366	0.7982	0.9188	0.9052	0.9642	0.9530	0.9814
		0.3	0.6072	0.7736	0.7450	0.8842	0.8502	0.9374	0.9278	0.9694
		0.4	0.5598	0.7148	0.6756	0.823	0.7974	0.8972	0.885	0.9398
	0.70	0.0	0.9194	0.9796	0.9760	0.9952	0.9954	0.9984	0.9990	1.0000
		0.1	0.8952	0.9692	0.9622	0.9906	0.9886	0.9978	0.9984	0.9998
		0.2	0.8534	0.9440	0.9450	0.9862	0.9814	0.9958	0.9944	0.9988
		0.3	0.8000	0.9056	0.9046	0.9650	0.9670	0.9890	0.9904	0.9978
		0.4	0.7298	0.8622	0.8656	0.9498	0.9410	0.9800	0.9732	0.9914

HR: hazard ratio; CR: censoring rate

Table 5.3 Comparison of empirical powers to exponential distribution (continued)

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
900	0.85	0.0	0.4276	0.6570	0.5654	0.7578	0.6956	0.8370	0.7986	0.8824
		0.1	0.4002	0.6146	0.5158	0.6992	0.6330	0.7856	0.7568	0.8478
		0.2	0.3718	0.5296	0.4772	0.6358	0.5938	0.7364	0.6986	0.7888
		0.3	0.3310	0.4442	0.4308	0.5552	0.5530	0.6632	0.6488	0.7386
		0.4	0.2802	0.3310	0.3674	0.4422	0.4742	0.5604	0.5722	0.6472
	0.80	0.0	0.7086	0.8928	0.8374	0.9460	0.9238	0.9762	0.9710	0.9908
		0.1	0.6564	0.8574	0.8014	0.9270	0.8614	0.9478	0.9508	0.9814
		0.2	0.6076	0.8142	0.7582	0.9010	0.8630	0.9600	0.9340	0.9704
		0.3	0.5420	0.7332	0.6914	0.8448	0.8218	0.9196	0.8962	0.9494
		0.4	0.5054	0.6522	0.6354	0.7784	0.7516	0.8612	0.8536	0.9156
	0.75	0.0	0.9116	0.9796	0.9720	0.9954	0.9936	0.9990	0.9982	0.9996
		0.1	0.8642	0.9662	0.9510	0.9890	0.9884	0.9974	0.9972	0.9990
		0.2	0.8248	0.9460	0.9272	0.9854	0.9750	0.9952	0.9946	0.9984
		0.3	0.7800	0.9246	0.9004	0.9740	0.9600	0.9918	0.9852	0.9954
		0.4	0.7156	0.8792	0.8504	0.9434	0.9280	0.9752	0.9714	0.9906
	0.70	0.0	0.9836	0.9980	0.9982	1.0000	1.0000	1.0000	1.0000	1.0000
		0.1	0.9734	0.9942	0.9980	0.9998	1.0000	1.0000	0.9998	1.0000
		0.2	0.9536	0.9902	0.9932	0.9990	0.9990	1.0000	0.9998	1.0000
		0.3	0.9254	0.9810	0.9786	0.9958	0.9960	0.9996	0.9992	1.0000
		0.4	0.8890	0.9686	0.9646	0.9918	0.9894	0.9984	0.9978	0.9996

HR: hazard ratio; CR: censoring rate

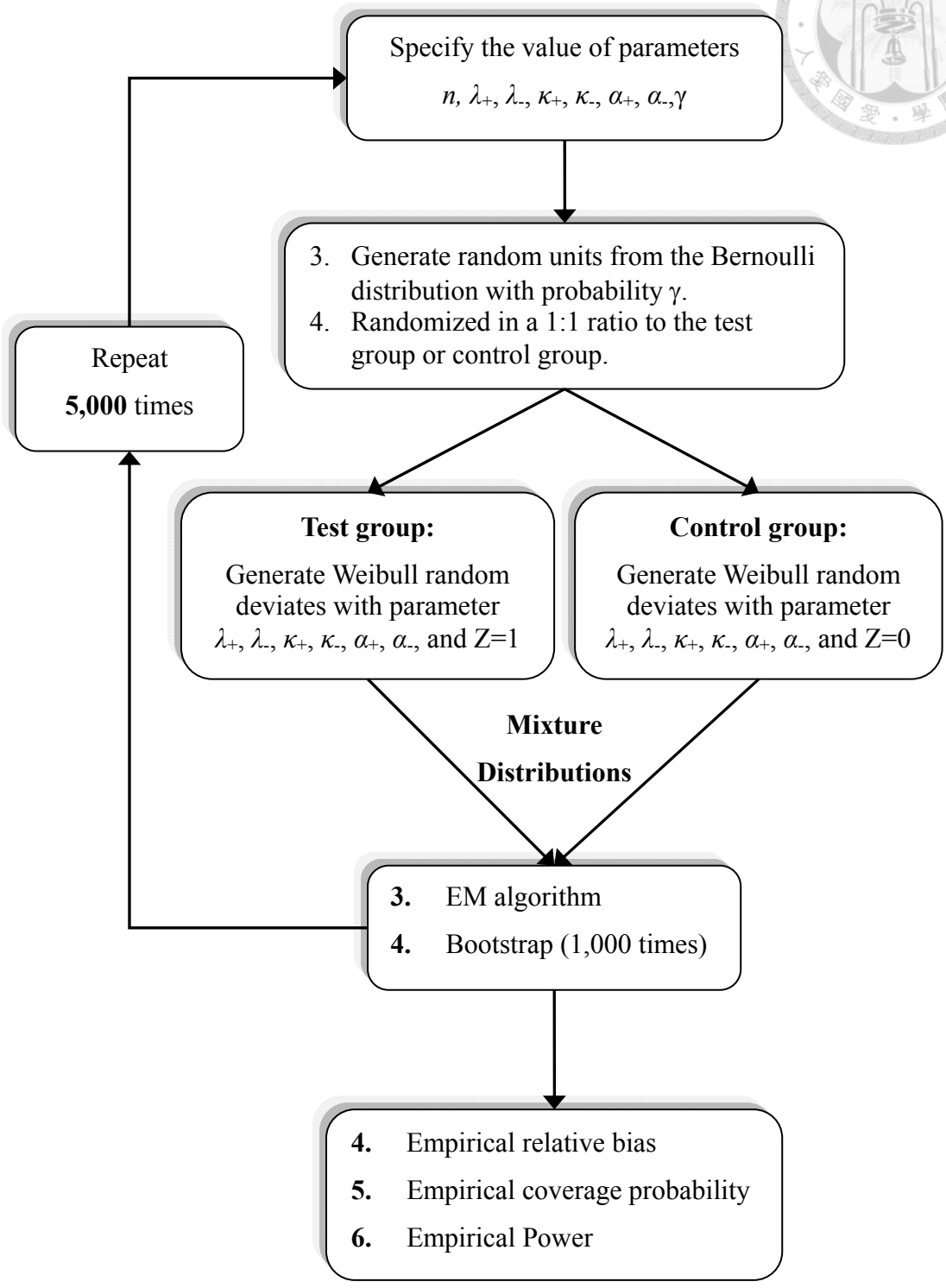


Figure 5.3 Flow chart of the simulation study to parametric proportional hazard regression model

Table 5.4 Relative bias (%) and coverage probability under the parametric proportional hazard model

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
600	0.8	0.0	11.95[a]	-0.86	8.6	-0.95	6.34	-0.86	4.95	-0.88
			0.5832[b]	0.9492	0.6856	0.9494	0.8174	0.9504	0.8962	0.9446
	0.8	0.2	11.31	-2.17	8.79	-2.06	6.55	-1.71	4.49	-1.41
			0.6190	0.9540	0.7628	0.9576	0.8376	0.9506	0.9062	0.9532
	0.75	0.0	14.39	-0.09	11.28	-0.26	8.54	-0.48	5.60	-0.02
			0.3306	0.9530	0.5060	0.9594	0.7602	0.9616	0.8938	0.9592
	0.75	0.2	16.09	-2.15	11.62	-1.54	8.36	-1.78	6.22	-2.00
			0.4724	0.9548	0.6292	0.9479	0.7930	0.9544	0.8538	0.9526
	0.7	0.0	17.31	0.52	13.87	0.26	9.07	-0.19	6.27	-0.23
			0.2224	0.9428	0.4234	0.9438	0.6458	0.9522	0.8284	0.9414
	0.7	0.2	17.45	-1.65	13.76	-1.64	9.39	-1.59	6.45	-1.28
			0.2472	0.9548	0.4536	0.9572	0.6546	0.9522	0.8356	0.9534

HR: hazard ratio; CR: censoring rate
a: Relative bias (%); b: Coverage probability

Table 5.5 Comparison of empirical sizes under the parametric proportional hazard model

		PPV							
		0.5		0.6		0.7		0.8	
N	CR	Current	EM	Current	EM	Current	EM	Current	EM
600	0.0	0.0516	0.0478	0.0474	0.0506	0.0498	0.0488	0.0536	0.0544
	0.1	0.0498	0.0574	0.0492	0.0528	0.0488	0.0478	0.0542	0.0592
	0.2	0.0524	0.0542	0.0496	0.0502	0.0492	0.0458	0.0544	0.0584
	0.3	0.0498	0.0518	0.0516	0.0486	0.0516	0.0496	0.0546	0.0526
	0.4	0.0526	0.0514	0.0546	0.0540	0.0522	0.0512	0.0498	0.0488

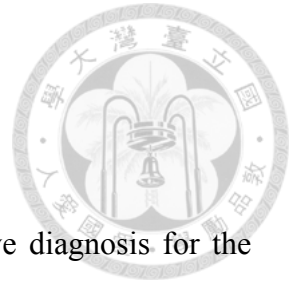
CR: censoring rate

Table 5.6 Comparison of empirical power under the parametric proportional hazard model

			PPV							
			0.5		0.6		0.7		0.8	
N	HR	CR	Current	EM	Current	EM	Current	EM	Current	EM
600	0.80	0.0	0.5494	0.7614	0.6696	0.8642	0.7858	0.8954	0.8934	0.9426
	0.80	0.2	0.4512	0.6676	0.5704	0.7394	0.6932	0.8524	0.8506	0.8914
	0.75	0.0	0.7352	0.9390	0.8734	0.9602	0.9476	0.9754	0.9836	0.9876
	0.75	0.2	0.6336	0.8390	0.7832	0.9104	0.9236	0.9636	0.9566	0.9850
	0.70	0.0	0.9162	0.9726	0.9738	0.9944	0.9962	0.9982	0.9988	1.0000
	0.70	0.2	0.8714	0.9476	0.9436	0.9758	0.9746	0.9924	0.9918	0.9970

CR: censoring rate

Chapter 6 Discussion



Under the enrichment design, all patients must have a positive diagnosis for the molecular targets by the diagnostic device to be randomized to receive either the targeted drug or the control treatment in the targeted clinical trials. However, no diagnostic device is perfect with 100% PPV. The positive predictive value is an increasing function of prevalence. For example, the molecular target Xalkori is approved by the US FDA issued guidance *Draft Guidance on In Vitro Companion Diagnostic Devices* 2011 for the treated of non-small-cell lung cancer. However, the prevalence rate of each patient with the target is only 5%. The PPV of the diagnostic devices for the target will be low. Therefore, the treatment effect of Xalkori may be underestimated. Since the PPV of the *in vitro* companion devices will not be high if the prevalence rate of the molecular target is low even the device is approved by the US FDA. The treatment effect of the patients truly with the molecular target is underestimated. The magnitude of underestimation is inversely proportional to the magnitude of the PPV.

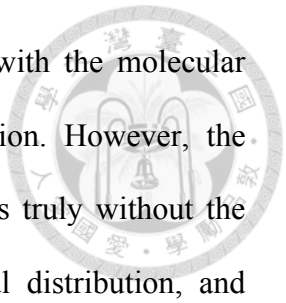
As a result, the current estimation method may produce a biased estimator for the treatment effects of the targeted test drug in the patients truly with molecular target. Hence, we propose estimation and testing procedures by application of the EM algorithm to incorporate information of the PPV for inference of the treatment effects in the patient population truly with the molecular target. The results of our simulations show that the proposed estimation method can provide a sufficient coverage probability. The proposed testing procedure also can adequately control the type I error rate at the

nominal level and is uniformly more powerful than the current method.

In the application of the EM algorithm, selection of initial values for μ_{i+} and μ_{i-}, i = T, C is important. The estimates of efficacy for the current available therapies (control) are known for most of diseases. In addition, the expected magnitude of increment of efficacy over the control provided by the targeted drug is also specified in the protocol for the sample size determination. Therefore, a range of reasonable initial values can be determined for the EM algorithm from this information. One method for selection of initial values to generate μ_{i+} and μ_{i-} from a exponential distribution is to employ the sample mean survival time of the observed data, $\bar{y}_T$ and $\bar{y}_C$. In other words, the traditional sample mean survival times, $\bar{y}_T$ and $\bar{y}_C$, of the test and control treatments are reasonable initial values for the proposed method.

We consider the exponential parametric model that satisfies the proportional hazard assumption. However, the proportionality assumption may not hold in practice. We conducted an additional simulation study to investigate the impact of violation of the proportional hazard assumption on performance of our proposed method. Because the hazard function of the lognormal distribution is nonmonotonic and changes over time, we generated the survival times from the log-normal distribution for the situation when the proportional hazard assumption is violated. We further consider the following two cases in the simulation study:

Case 1: Because we assume that the molecular targeted test drug is ineffective in the patients truly without the target and the placebo is ineffective in the patients truly with and without the target either. Therefore, the survival

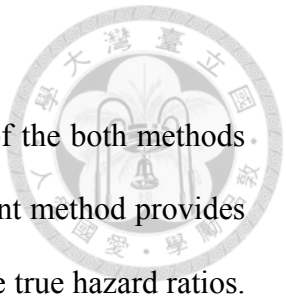


times of the test drug group for the patients truly with the molecular target are generated from the exponential distribution. However, the survival times of the test drug group for the patients truly without the molecular target are generated from the log-normal distribution, and those of the control group are also generated from the log-normal distribution.

Case 2: The survival times of the patients truly with and without the target assigned either to the molecular target test drug or to placebo were all generated from the log-normal distribution.

The sample size for the additional simulation study is 600 per group with a censoring proportion of 0.2. The hazard ratios are 1.0, 0.80, and 0.75 and the PPVs are 0.5, 0.6, 0.7, and 0.8. The results of this additional simulation are given in Tables 6.1 and 6.2. Table 6.1 presents the relative bias and coverage probability of the current and the proposed EM method. The results in Table 6.1 reveal that both the current and EM methods are biased.

For Case 1, when the hazard ratio is 1, both methods underestimate the true hazard ratio with a relative bias ranging from –3.16% to –4.8%. When the hazard ratio is 0.8 or 0.75, the current method over-estimates the true hazard ratios, whereas the EM method produces under-estimated estimates. The relative bias of the two methods increases as the PPV decreases. The magnitude of the bias of the current method decreases from around 10.0% when PPV is 0.5 to around 1.0% when PPV is 0.8. On the other hand, the bias of the EM method changes from –15.1% when PPV is 0.5 to –8.44% when PPV is 0.8.



For Case 2, when the hazard ratio is 1, the absolute relative biases of the both methods do not exceed 0.26%. When the hazard ratio is 0.8 or 0.75, the current method provides overestimated estimates, whereas the EM method under-estimates the true hazard ratios. The relative bias of the two methods increases as the PPV decreases. The magnitude of the bias of the current method decreases from 14.43% when PPV is 0.5 to around 4.0% when PPV is 0.8 On the other hand, the bias of the EM method changes from –38.77% when PPV is 0.5 to -12.93% when PPV is 0.8.

For case 1, when the hazard ratio is 1, the fluctuation of the magnitude of the empirical coverage probability over PPV is quite small. However, when the hazard ratio is 0.75 or 0.8, the empirical coverage probability of both methods is an increasing function of PPV. For the current method, the empirical coverage probability increase from around 0.65 when PPV is 0.5 to 0.94 when PPV is 0.8. On the other hand, the empirical coverage probability of the EM method increases from about 0.77 for PPV being 0.5 to 0.81 for PPV being 0.8.

For case 2, when the hazard ratio is 1, the empirical coverage probabilities of the both methods range from 0.9314 to 0.9422 with very minor fluctuations. When the hazard ratio is 0.8 or 0.75, the empirical coverage probability of both methods is again an increasing function of PPV. However, the empirical coverage probabilities of the both methods are comparable for different values of PPVs.

Table 6.2 presents the empirical powers of the current and the proposed EM method. For Case 1, when the hazard ratio is 1, all empirical sizes of both the current method

and the proposed EM procedure are inflated with a range from 0.0858 to 0.1288. These results demonstrate that both methods cannot control the size at its nominal level. Both of the empirical sizes and powers are increasing functions of the PPV. For Case 2, when the hazard ratio is 1, all empirical sizes of both the current method and the proposed EM procedure are from 0.0578 to 0.0686. The sizes of both methods are slightly inflated. Similar to Case 1, the empirical sizes and powers are increasing functions of the PPV. In summary, when the proportional hazard assumption is violated, both the current and EM methods produce biased estimates. The magnitude of the bias of the EM method seems to be larger than that of the current method. In addition, both methods can not control the empirical size.

The most popular survival data regression model is the Cox (1972) proportional hazards model, in which the hazard function $h(y|\mathbf{z})$ for an individual with covariate vector $\mathbf{z} \in \mathbb{R}^P$ is modeled as

$$h(y|z) = h_0(y) \exp(\beta^T \mathbf{Z})$$

where $\mathbf{Z}$ is the p-vector of true covariate values. The function $h_0(t)$ is a baseline hazard function of unspecified form, so that the model is semi-parametric. There might be two approaches to hazard ratio under the Cox's proportional hazard model when the true status of the molecular target is not completely known. The first approach is to employ the observed status of the molecular target as a covariate Z_2 in the Cox's proportional hazard model. However, the observed Z_2 is a surrogate measure for the true but latent covariate X. The second approach is to apply the EM algorithm to the produce of the partial likelihood. Under the assumption of the proportional hazards, the partial likelihoods can be formulated for the patients truly with the molecular target and truly without the molecular target, respectively. Then the latent variable for the true status of

the molecular target may be introduced to form the complete-data partial likelihood. Let $h_+(y|z)$ be the hazard rate for the patients truly with the molecular target and $h_-(y|z)$ for those truly without the molecular target, respectively.

$$h_+(y \mid z) = h_{0+}(y \mid z)\exp(\lambda_+ X)$$
$$h_-(y \mid z) = h_{0-}(y \mid z)\exp(\lambda_- X)$$

where $h_{0+}(y|z)$ and $h_{0-}(y|z)$ are an arbitrary baseline hazard rates and Z is an indicator variable for the treatment. The likelihood based on the hazard function as specified above is expressed by

$$L = \prod \left\{\gamma h_{0+}(y \mid z)\exp(\lambda_+ X)\right\}^{x_i} \left\{(1-\gamma)h_{0-}(y \mid z)\exp(\lambda_- X)\right\}^{(1-x_i)}$$

where x_i is the latent variable indicating the true status of the molecular target for patient i.

The partial likelihood may be formulated as

$$L = \prod_{i=1}^{n} \left\{\frac{\gamma\exp(\lambda_+ X)}{\sum\exp(\lambda_+ X)}\right\}^{x_i} \left\{\frac{(1-\gamma)\exp(\lambda_- X)}{\sum\exp(\lambda_- X)}\right\}^{(1-x_i)}$$

The inferential procedures for the treatment effects of the targeted drug based on the censored endpoints such as overall survival (OS) or progression free survival (PFS) in the patients truly with the molecular target with the Cox's proportional hazard model require further research.

Table 6.1 Relative bias and the coverage probability when censored rate = 20%, n = 600

		PPV							
		0.5		0.6		0.7		0.8	
HR		Current	EM	Current	EM	Current	EM	Current	EM
Case 1									
1.00	RB	-3.16	-3.22	-3.37	-3.42	-4.18	-4.21	-4.77	-4.80
	CP	0.9142	0.9114	0.9052	0.9056	0.889	0.8872	0.8732	0.8712
0.80	RB	7.74	-15.10	4.58	-11.88	1.75	-9.70	-1.03	-8.44
	CP	0.7770	0.7782	0.8838	0.7932	0.9280	0.8068	0.9374	0.8116
0.75	RB	10.45	-15.03	6.57	-11.92	3.28	-9.64	0.09	-8.57
	CP	0.6562	0.7956	0.8202	0.8012	0.9116	0.8122	0.9458	0.8054
Case 2									
1.00	RB	0.26	0.19	-0.03	-0.09	0.16	0.12	0.05	0.02
	CP	0.9422	0.9418	0.9360	0.9340	0.9330	0.9314	0.9362	0.9350
0.80	RB	11.24	-38.66	8.90	-28.86	6.44	-20.31	4.18	-13.03
	CP	0.6172	0.6676	0.7338	0.7250	0.8262	0.7918	0.8854	0.8344
0.75	RB	14.43	-38.77	11.16	-29.59	8.28	-22.23	5.32	-12.93
	CP	0.4520	0.6632	0.6262	0.7376	0.7538	0.7900	0.8628	0.8484

RB: Relative bias (%); CP: Coverage probability;

Table 6.2 Comparison of empirical powers when censored rate = 20%, n = 600

		PPV							
		0.5		0.6		0.7		0.8	
	HR	Current	EM	Current	EM	Current.	EM	Current	EM
Case 1	1.00	0.0858	0.0886	0.0948	0.0944	0.1110	0.1128	0.1268	0.1288
	0.80	0.6270	0.8074	0.7828	0.8978	0.8780	0.9374	0.9454	0.9686
	0.75	0.8250	0.9424	0.9234	0.9760	0.9742	0.9908	0.9906	0.997
Case 2	1.00	0.0578	0.0582	0.0640	0.0660	0.0670	0.0686	0.0638	0.0650
	0.80	0.4380	0.7400	0.5694	0.7834	0.6960	0.8286	0.7902	0.8574
	0.75	0.6542	0.8722	0.7922	0.9126	0.8870	0.9542	0.9458	0.9728

REFERENCES

1. Basford K.E., Greenway D.R., McLachlan G.J. and Peel D. (1997). Standard errors of fitted component means of normal mixtures, *Computational Statistics*. 12(1):1-17.
2. Buyse, M., Loi, S., van't Veer, L., et al. (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer, *Journal of the National Cancer Institute, 98*, 1183-1192.
3. Casciano D.A. and Woodcock J. (2006). Empowering microarrays in the regulatory setting, *Nature Biotechnology*. 24: 1103.
4. Chow S.C. and Liu J.P. (2004). *Design and Analysis of Clinical Trials*, 2nd Ed., John Wiley and Sons, 3rd Ed., New York, USA.
5. Collett D. (2003). *Modeling Survival Data in Medical Research*, 2nd Ed., Chapman and Hall, London, UK.
6. Dalton W.S. and Friend S.H. (2006). Cancer Biomarkers – an invitation to the table, *Science*. 312: 1165-8.
7. Dempster A.P., Laird N.M. and Rubin D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*. 39(1): 1-38
8. Efron B. and Tibshirani R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York, USA.
9. Fleiss J.L., Levin B. and Paik M.C. (2003). *Statistical Methods for Rates and Proportions*, 3rd Ed. John Wiley and Sons, New York.
10. Freidlin B. and Simon R. (2005). Evaluation of randomized discontinuation design.

Journal of Clinical Oncology. 23: 5094-5098.

11. Liu J.P. and Chow S.C. (2008). Issues on the diagnostic multivariate index assay and targeted clinical trials, *Journal of Biopharmaceutical Statistics*. 18: 167-182.
12. Liu J.P. and Lin J.R. (2008). Statistical methods for targeted clinical trials under enrichment design. *Journal of the Formosan Medical Association*.107: S35-S42.
13. Liu J.P., Lin J.R. and Chow S.C. (2009). Inference on treatment effects for targeted clinical trials under enrichment design. *Pharmaceutical Statistics*. 8: 356-370
14. Maitournam A. and Simon R. (2005). On the efficiency of targeted clinical trials. *Statistics in Medicine*. 24: 329-339.
15. McLachlan G.J. and Krishnan T. (1997). *The EM algorithm and Extensions*, Wiley, New York
16. McLachlan G.J. and Peel D. (2000). *Finite Mixture Models*, Wiley, New York, USA.
17. More, J.J., Garbow, B.S. and Hillstrom, K.E. (1980). *User Guide for MINPACK-1*, ANL-80-74. Argonne National Laboratory, Chicago.
18. Ravdin, P.M. and Chamness, G.C. (1995). The c-erbB-2 proto-oncogene as a prognostic and predictive markers in breast cancer: A paradigm for the development of other macromolecular markers-a review. *Gene*, 159, 19-27.
19. Romond, E.H., Perez, E.A., Bryant, J. et al. (2005). Trastuzumab plus chemotherapy for operable *HER*2-positive breast cancer, *New England Journal of Medicine*. 353: 1673-84.
20. Ng, S.K., McLachlan, G.J., Yau, K.K. W. and Lee, A.H. (2004). Modelling the distribution of ischaemic stroke-specific survival time using an EM-based mixture approach with random effects adjustment. *Statistics in Medicine*. 23: 2729-2744.
21. Seshadri, R., Figaira, E.A., Horsfall, D.J., McCaul, K., Setlur, V., and Kitchen, P.

(1993). Clinical significance of HER-2/neu oncogene amplification in primary breast cancer. The South Australian Breast Cancer Study Group. *Journal of Clinical Oncology*. 11: 1936-1942.

22. Simon R. (2008). The use of genomics in clinical trial design. *Clinical Cancer Research*. 14: 5984-93.
23. Simon R. and Maitournam A. (2004). Evaluating the efficiency of targeted designs for randomized clinical trials. *Clinical Cancer Research*. 10: 6759- 6763.
24. Slamon, D.J., Leyland-Jones B., Shak, S., et al. (2001). Use of chemotherapy plus a monoclonal antibody against *HER*2 for metastatic breast cancer that overexpresses *HER*2, *New England Journal of Medicine*. 344: 783-792.
25. U.S. FDA (1998). *Decision Summary P980018*. Rockville, Maryland, U.S.A.
26. U.S. FDA (2001). *Decision Summary P980024/S001*. Rockville, Maryland, U.S.A.
27. U.S. FDA (2004). *Decision Summary P040030*. Rockville, Maryland, U.S.A.
28. U.S. FDA (2005). *The draft concept paper on Drug-Diagnostic Co-Development*. The U.S. Food and Drug Administration, Rockville, Maryland, USA.
29. U.S. FDA (2006). Annotated Redlined Draft Package Insert for Herceptin®. The U.S. Food and Drug Administration, Rockville, Maryland, U.S.A.
30. U.S. FDA (2007a). *Decision Summary k062694*. Rockville, Maryland, U.S.A.
31. U.S. FDA (2007b). Draft Guidance on *In Vitro Diagnostic Multivariate Index Assays*. The U.S. Food and Drug Administration, Rockville, Maryland, USA.
32. U.S. FDA (2007c). *Guidance on Gene Expression Profiling Test System for Breast Cancer Prognosis*. The U.S. Food and Drug Administration, Rockville, Maryland, USA.
33. U.S. FDA (2007d). *Guidance on Drug Metabolizing Enzyme Genotyping System*. The U.S. Food and Drug Administration, Rockville, Maryland, USA.

34. The U.S. FDA (2007e). *Decision Summary k062694*, Rockville, Maryland, U.S.A.
35. The U.S. FDA (2011). Draft Guidance on *In Vitro Companion Diagnostic Devices*. The U.S. Food and Drug Administration, Rockville, Maryland, USA.
36. Varmus H. (2006). The new era in cancer research, *Science*. 312: 1162-5.
37. Wu C.F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*. 11: 95-103.
38. Zucker, D.M. and Spiegelman, D. (2004) Inference for the proportional hazards model with misclassified discrete-valued covariates, *Biometrics*. 60: 324–334.
39. Zucker, D. M. and Spiegelman, D. Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in Medicine* 2008, 27: 1911–1933.

Appendix A Fortran Codes for Simulation

Fortran Codes to the Exponential Distribution Model

```
PROGRAM EMEXP
PARAMETER (MN = 2000)
PARAMETER (MM = 1000)
PARAMETER (N1 = 300)
!PARAMETER (BB = 1000)
PARAMETER (W1 = 0.8) ! PPV
PARAMETER (W2 = 0.8) ! EVENT RATE = 1 - CENSORED RATE
PARAMETER (A1 = 0.75) ! LAMBDA = 0.75 = 1/MEAN
PARAMETER (A2 = 1) ! LAMBDA = 0.75 = 1/MEAN

IMPLICIT DOUBLE PRECISION (a-h, o-Z)
DOUBLE PRECISION IT
INTEGER ISEED, IPER(MN), IB(MN)

REAL P, PD, XTM1, XTM2, XCM1, XCM2, D, PDT, PDC, &
      TXT1, CXT1, TXT2, CXT2, TXC1, CXC1, TXC2, CXC2, TXX(MN),
CXX(MN), &
      TUT(2), TDT(2), TUC(2), TDC(2), TROU(2), TSE(2), TOT(2), &
      OUT(2), ODT(2), OUC(2), ODC(2), ROU(2), OSE(2), OT(2), OLCI(2),
OUCI(2), OCI(2), &
      SOUT(2), SOUC(2), SROU(2), SOSE(2), SOLCI(2), SOUCI(2), &
      SUT(2), SUC(2), SPT(2), SPC(2), SRU(1), &
      T(1), CI(2), LCI(1), UCI(1), SEB(2), SUSE(2)

COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
        DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
        XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
        BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
        DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB


EXTERNAL RNPER, RNEXP, RNSET, FIN, DRNUN, RNBIN

NR1 = N1+N1
N = N1
NR = NR1
ND = N1


 OPEN(8, FILE='emd11.txt')
```

```
ISEED=0

CALL RNSET(ISEED)

!---------- BERNOULLI ----------
P = W1
PD = W2
!---------- EXPONENTIAL DATA T1 ----------
XTM1 = 1/A1
!---------- EXPONENTIAL DATA T2 ----------
XTM2 = 1/A2
!---------- EXPONENTIAL DATA C1 ----------
XCM1 = 1/A2
!---------- EXPONENTIAL DATA C2 ----------
XCM2 = 1/A2
!---------- "INVERSE" HAZARD RATIO ----------
D = LOG(XCM1 / XTM1)
!================================================
  CALL EXPMIN(DLIM)
!---------- SIMULATION LOOP START ----------
DO 11 L = 1, MM
WRITE(*,139) L
139 FORMAT(6X, 'SIMUL_C1', 4X, I6)
!--------------------
U1(1) = XTM1
U1(2) = XTM2
U2(1) = XCM1
U2(2) = XCM2
P1(1) = P
P1(2) = 1-P
P2(1) = P
P2(2) = 1-P
!---------- DATA SET ----------
CALL RNBIN(NR, 1, P, IR)
!---------- PSEUDORANDOM PERMUTATION ----------
CALL RNPER(NR, IPER)
!---------- SORT RANDOM NUMBERS ----------
CALL SVIGP(NR, IPER, IB, IR)
!--------------------
!---------- Generate pseudorandom numbers from a standard exponential distribution
----------
!---------- Deviates from the exponential distribution with mean THETA ----------

PDT = 1 / (A1 * (1-W2) / W2)
PDC = 1 / (A2 * (1-W2) / W2)

DO 701 I = 1, NR
!---------- TEST ----------
```

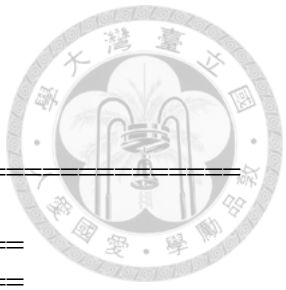
```
IF (I .LE. N .AND. IR(I) .EQ. 1) THEN
CALL RNEXP(1, TXT1)
CALL SSCAL(1, XTM1, TXT1, 1)
TXX(I) = TXT1
CALL RNEXP(1, CXT1)
CALL SSCAL(1, PDT, CXT1, 1)
CXX(I) = CXT1
!--------------------
    IF ( TXT1 .LE. CXT1 ) THEN
    XX(I) = TXT1
    ID(I) = 1
    ELSE IF ( TXT1 .GT. CXT1 ) THEN
    XX(I) = CXT1
    ID(I) = 0
    END IF
!--------------------
ELSE IF (I .LE. N .AND. IR(I) .EQ. 0) THEN
CALL RNEXP(1, TXT2)
CALL SSCAL(1, XTM2, TXT2, 1)
TXX(I) = TXT2
CALL RNEXP(1, CXT2)
CALL SSCAL(1, PDC, CXT2, 1)
CXX(I) = CXT2
!--------------------
    IF ( TXT2 .LE. CXT2 ) THEN
    XX(I) = TXT2
    ID(I) = 1
    ELSE IF ( TXT2 .GT. CXT2 ) THEN
    XX(I) = CXT2
    ID(I) = 0
    END IF
!---------- CONTROL ----------
ELSE IF (I .GT. N .AND. IR(I) .EQ. 1) THEN
CALL RNEXP(1, TXC1)
CALL SSCAL(1, XCM1, TXC1, 1)
TXX(I) = TXC1
CALL RNEXP(1, CXC1)
CALL SSCAL(1, PDC, CXC1, 1)
CXX(I) = CXC1
!--------------------
    IF ( TXC1 .LE. CXC1 ) THEN
    XX(I) = TXC1
    ID(I) = 1
    ELSE IF ( TXC1 .GT. CXC1 ) THEN
    XX(I) = CXC1
    ID(I) = 0
    END IF
!--------------------
ELSE IF (I .GT. N .AND. IR(I) .EQ. 0) THEN
```

```
CALL RNEXP(1, TXC2)
CALL SSCAL(1, XCM2, TXC2, 1)
TXX(I) = TXC2
CALL RNEXP(1, CXC2)
CALL SSCAL(1, PDC, CXC2, 1)
CXX(I) = CXC2
!--------------------
    IF ( TXC2 .LE. CXC2 ) THEN
    XX(I) = TXC2
    ID(I) = 1
    ELSE IF ( TXC2 .GT. CXC2 ) THEN
    XX(I) = CXC2
    ID(I) = 0
    END IF
!--------------------
END IF
701 CONTINUE

!---------- END NR ----------
DO I = 1, N
XXT(I) = XX(I)
XXC(I) = XX(I+N)
DXT(I) = ID(I)
DXC(I) = ID(I+N)
END DO
!=====================================================

!=====================================================
!---------- TRUE TEST GROUP SAMPLE MEAN ----------
TUT(1) = 0.
TDT(1) = 0.
DO I = 1, N
TUT(1) = TUT(1) + XXT(I) * IR(I)
TDT(1) = TDT(1) + DXT(I) * IR(I)
END DO
TUT(1) = TUT(1) / TDT(1)
!---------- TRUE CONTROL GROUP SAMPLE MEAN ----------
TUC(1) = 0.
TDC(1) = 0.
DO I = 1, N
TUC(1) = TUC(1) + XXC(I) * IR(I+N)
TDC(1) = TDC(1) + DXC(I) * IR(I+N)
END DO
TUC(1) = TUC(1) / TDC(1)
!---------- TRUE MEAN RATE ----------
TROU(1) = TUC(1) / TUT(1)
TSE(1) = (TDT(1) + TDC(1)) / (TDT(1) * TDC(1))
TOT(1) = ( LOG(TROU(1) ) / SQRT(TSE(1)) )
TOT(1) = ABS(TOT(1))
```

```
!---------- ALL TEST GROUP SAMPLE MEAN ----------
OUT(1) = 0.
ODT(1) = 0.
DO I = 1, N
OUT(1) = OUT(1) + XXT(I)
ODT(1) = ODT(1) + DXT(I)
END DO
OUT(1) = OUT(1) / ODT(1)
!WRITE(*,*) OUT(1)
!---------- ALL CONTROL GROUP SAMPLE MEAN ----------
OUC(1) = 0.
ODC(1) = 0.
DO I = 1, N
OUC(1) = OUC(1) + XXC(I)
ODC(1) = ODC(1) + DXC(I)
END DO
OUC(1) = OUC(1) / ODC(1)
ODTT = ODT(1) + ODC(1)
ODTT = ODTT / NR
!---------- MEAN RATE ----------
ROU(1) = OUC(1) / OUT(1)
OSE(1) = ((ODT(1) + ODC(1)) / (ODT(1) * ODC(1)))
OT(1) = ( LOG(ROU(1)) / SQRT(OSE(1)) )
OT(1) = ABS(OT(1))

II=0
IF (OT(1) .GE. 1.96) II=1
OCI(1) = OCI(1) + II
!--------------------
CC1 = 0
OLCI(1) = EXP( LOG(ROU(1)) - 1.96 * SQRT( OSE(1) ) )
OUCI(1) = EXP( LOG(ROU(1)) + 1.96 * SQRT( OSE(1) ) )
IF ((OLCI(1) .LE. EXP(D)) .AND. (OUCI(1) .GE. EXP(D))) CC1 = 1
OCI(2) = OCI(2) + CC1
!--------------------
ROU(1) = LOG(ROU(1))
OSE(1) = SQRT(OSE(1))
SOUT(1) = SOUT(1) + OUT(1)
SOUC(1) = SOUC(1) + OUC(1)
SROU(1) = SROU(1) + ROU(1)
SOSE(1) = SOSE(1) + OSE(1)
SOLCI(1) = SOLCI(1) + OLCI(1)
SOUCI(1) = SOUCI(1) + OUCI(1)
!=================================================

!---------- EM METHOD ----------
!---------- CALL NOCO ----------
CALL NOCOT
```

```
!---------- EM METHOD END ----------
DO J=1,2
SUT(J) = SUT(J) + U1(J)
SUC(J) = SUC(J) + U2(J)
SPT(J) = SPT(J) + PC(J)
SPC(J) = SPT(J)
END DO
!---------- R(+) RATIO ----------
DU(1) = LOG(U2(1) / U1(1))
SRU(1) = SRU(1) + DU(1)
!--------------------
!==================================================
!---------- CALL BOOTSTRAPS ----------
CALL BOOT
!---------- HOP TEST ----------
SEB(1) = SQRT(SE(1))
SUSE(1) = SUSE(1) + SEB(1)
T(1) = (DU(1)) / SEB(1)
T(1) = ABS(T(1))

JJ1 = 0
IF (T(1) .GE. 1.96)    JJ1 = 1
CI(1) = CI(1) + JJ1
!---------- CONFERENCE INTERVAL ----------
CC2 = 0
UCI(1) = DU(1) + (1.96 * SEB(1))
LCI(1) = DU(1) - (1.96 * SEB(1))
IF( (UCI(1) .GE. D) .AND. (LCI(1) .LE. D) )    CC2=1
CI(2) = CI(2) + CC2
!--------------------
11 CONTINUE
!--------------------

!---------- SIMULATION LOOP END ----------
!---------- SIMULATION MEAN ----------
DO J = 1, 2
SUT(J) = SUT(J) / MM
SUC(J) = SUC(J) / MM
SPT(J) = SPT(J) / MM
END DO

SRU(1) = SRU(1) / MM

SOUT(1) = SOUT(1) / MM
SOUC(1) = SOUC(1) / MM
SROU(1) = SROU(1) / MM
SOSE(1) = SOSE(1) / MM
SUSE(1) = SUSE(1) / MM
```

```
CI(1) = CI(1) / MM
CI(2) = CI(2) / MM
OCI(1) = OCI(1) / MM
OCI(2) = OCI(2) / MM

SRU(1) = EXP(SRU(1))
SROU(1) = EXP(SROU(1))

WRITE(8, "(1X, A79)") '      SUT(1)      SUC(1)      SRU(1)      RB
SUSE(1)    CI(1)      CI(2)      SPT(1)'
WRITE(8, "(3F10.4, F10.2, 4F10.4)") SUT(1), SUC(1), SRU(1), (SRU(1)-A1)/A1*100,
SUSE(1), CI(1), CI(2), SPT(1)
WRITE(8,                                                                        *)
'-------------------------------------------------------------------------------------------'
WRITE(8, "(1X, A69)") '      SOUT(1)      SOUC(1)      SROU(1)      RB
SOSE(1)    OCI(1)      OCI(2)'
WRITE(8, "(3F10.4, F10.2, 3F10.4)") SOUT(1), SOUC(1), SROU(1),
(SROU(1)-A1)/A1*100, SOSE(1), OCI(1), OCI(2)
WRITE(8,                                                                        *)
'-------------------------------------------------------------------------------------------'
WRITE(8, "(1X, A29)") '     SUT(2)      SUC(2)      SPT(2)'
WRITE(8, "(7F10.4)") SUT(2), SUC(2), SPT(2)

WRITE(*, "(1X, A77)") '     SUT(1)      SUC(1)      SRU(1)      RB      SUSE(1)
CI(1)      CI(2)      SPT(1)'
WRITE(*, "(3F10.4, F8.2, 4F10.4)") SUT(1), SUC(1), SRU(1), (SRU(1)-A1)/A1*100,
SUSE(1), CI(1), CI(2), SPT(1)
WRITE(*, *) '-------------------------------------------------------------------------------'
WRITE(*, "(1X, A67)") '     SOUT(1)      SOUC(1)      SROU(1)      RB      SOSE(1)
OCI(1)      OCI(2)'
WRITE(*,    "(3F10.4,    F8.2,    3F10.4)")    SOUT(1),    SOUC(1),    SROU(1),
(SROU(1)-A1)/A1*100, SOSE(1), OCI(1), OCI(2)
WRITE(*, *) '-------------------------------------------------------------------------------'
WRITE(*, "(1X, A29)") '     SUT(2)      SUC(2)      SPT(2)'
WRITE(*, "(7F10.4)") SUT(2), SUC(2), SPT(2)

!=================================================

END

!--------------------
!-------------- SUBROUTINE --------------
!--------------------
SUBROUTINE EXPMIN(rrmax)
implicit double precision (a-h, o-z)
rrmax = 1.0
120 rrmax = rrmax + 1.0
if (exp(-rrmax) .eq. 0.0) go to 130
go to 120
```



```
130 rrmax = rrmax - 1.0
return
end
!=============================================================

!=================================================
!=================================================
! SUBROUTINE BOOTSTRAPS METHOD FOR STANDARD ERROR
!=================================================
!=================================================
SUBROUTINE BOOT
PARAMETER( MN = 2000 )
PARAMETER (BB = 1000 )
implicit double precision (a-h, o-Z)
INTEGER BIR(MN), BIPER(MN), ISEED, KK, NRN, BIB(MN)
INTEGER I, BL
REAL BPP
REAL P, PD, BXTM1, BXTM2, BXCM1, BXCM2, D, PDT, PDC, &
      TXT1, CXT1, TXT2, CXT2, TXC1, CXC1, TXC2, CXC2, TXX(MN),
CXX(MN), &
      TUT(2), TDT(2), TUC(2), TDC(2), TROU(2), TSE(2), TOT(2), &
      OUT(2), ODT(2), OUC(2), ODC(2), ROU(2), OSE(2), OT(2), OLCI(2),
OUCI(2), OCI(2), &
      SOUT(2), SOUC(2), SROU(2), SOSE(2), SOLCI(2), SOUCI(2), &
      SUT(2), SUC(2), SPT(2), SPC(2), SRU(1), &
      T(1), CI(2), LCI(1), UCI(1), SEB(2), SUSE(2), SSE1(1)

COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
        DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
        XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
        BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
        DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB

EXTERNAL RNPER, DRNNOA, RNSET, TIN, EQTIL, DRNUN, RNBIN
!--------------------
DO J = 1,2
BUT(J) = 0
BUC(J) = 0
SUUT(J) = 0
SUUC(J) = 0
END DO
KB = 0
!=================================================
!---------- BERNOULLI ----------
BPP = PC(1)
!---------- EXPONENTIAL DATA ----------
BXTM1 = U1(1)
```

```
BXTM2 = U1(2)
BXCM1 = U2(1)
BXCM2 = U2(2)
!--------------------
BD = LOG(BXCM1 / BXTM1)
!--------------------
CALL EXPMIN(DLIM)
PDT = 1 / ((1 / BXTM1) * (1-ODTT) / ODTT)
PDC = 1 / ((1 / BXCM1) * (1-ODTT) / ODTT)
!---------- SIMULATION LOOP START ----------
DO 12 BL = 1, BB

CALL RNBIN(NR, 1, BPP, BIR)
!---------- PSEUDORANDOM PERMUTATION ----------
CALL RNPER(NR, BIPER)
!---------- SORT RANDOM NUMBERS ----------
CALL SVIGP(NR, BIPER, BIB, BIR)
!--------------------

DO 706 I = 1, NR
!---------- TEST ----------
IF (I .LE. N .AND. BIR(I) .EQ. 1) THEN
CALL RNEXP(1, TXT1)
CALL SSCAL(1, BXTM1, TXT1, 1)
TXX(I) = TXT1
CALL RNEXP(1, CXT1)
CALL SSCAL(1, PDT, CXT1, 1)
CXX(I) = CXT1
!--------------------
    IF ( TXT1 .LE. CXT1 ) THEN
    XX(I) = TXT1
    ID(I) = 1
    ELSE IF ( TXT1 .GT. CXT1 ) THEN
    XX(I) = CXT1
    ID(I) = 0
    END IF
!--------------------
ELSE IF (I .LE. N .AND. BIR(I) .EQ. 0) THEN
CALL RNEXP(1, TXT2)
CALL SSCAL(1, BXTM2, TXT2, 1)
TXX(I) = TXT2
CALL RNEXP(1, CXT2)
CALL SSCAL(1, PDC, CXT2, 1)
CXX(I) = CXT2
!--------------------
    IF ( TXT2 .LE. CXT2 ) THEN
    XX(I) = TXT2
    ID(I) = 1
    ELSE IF ( TXT2 .GT. CXT2 ) THEN
```

```
    XX(I) = CXT2
    ID(I) = 0
    END IF
!---------- CONTROL ----------
ELSE IF (I .GT. N .AND. BIR(I) .EQ. 1) THEN
CALL RNEXP(1, TXC1)
CALL SSCAL(1, BXCM1, TXC1, 1)
TXX(I) = TXC1
CALL RNEXP(1, CXC1)
CALL SSCAL(1, PDC, CXC1, 1)
CXX(I) = CXC1
!--------------------
    IF ( TXC1 .LE. CXC1 ) THEN
    XX(I) = TXC1
    ID(I) = 1
    ELSE IF ( TXC1 .GT. CXC1 ) THEN
    XX(I) = CXC1
    ID(I) = 0
    END IF
!--------------------
ELSE IF (I .GT. N .AND. BIR(I) .EQ. 0) THEN
CALL RNEXP(1, TXC2)
CALL SSCAL(1, BXCM2, TXC2, 1)
TXX(I) = TXC2
CALL RNEXP(1, CXC2)
CALL SSCAL(1, PDC, CXC2, 1)
CXX(I) = CXC2
!--------------------
    IF ( TXC2 .LE. CXC2 ) THEN
    XX(I) = TXC2
    ID(I) = 1
    ELSE IF ( TXC2 .GT. CXC2 ) THEN
    XX(I) = CXC2
    ID(I) = 0
    END IF
!--------------------
END IF
706 CONTINUE
!---------- END NR ----------
DO I = 1, N
XXT(I) = XX(I)
XXC(I) = XX(I+N)
DXT(I) = ID(I)
DXC(I) = ID(I+N)
END DO
!================================================

!================================================
!-------------- EM METHOD ----------------
```

```
U1(1) = BXTM1
U1(2) = BXTM2
U2(1) = BXCM1
U2(2) = BXCM2
P1(1) = BPP
P1(2) = (1-BPP)
P2(1) = BPP
P2(2) = (1-BPP)
!---------- CALL NOCO ----------
CALL NOCOB
!---------- EM METHOD END ----------
BOT(1) = LOG(U2(1) / U1(1))
!--------------------
BUT(1) = BUT(1) + BOT(1)
SUUT(1) = SUUT(1) + ( BOT(1) * BOT(1) )
!--------------------
KB = KB + 1
!--------------------
12 CONTINUE
!---------- SIMULATION LOOP END ----------
!---------- SIMULATION MEAN ----------
BUT(1) = BUT(1) / BB
!--------------------
BBIS = BUT(1) - BD
!---------- STANDARD ERROR ----------
SSE1(1) = (SUUT(1) - BB * BUT(1) * BUT(1)) / (BB - 1)
SE(1) = SSE1(1)
!--------------------
RETURN
END


!--------------------
!---------- NOCOT ----------
!--------------------
SUBROUTINE NOCOT
PARAMETER( MN = 2000 )
implicit double precision (a-h, o-z)
!LOGICAL FOUT
REAL XT1(1), XT2(1), PS(2), PS1(2), PS2(2), XLNC, D, BBIS
COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
        DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
        XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
        BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
        DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB

DATA TT, IEND /.01, 200/
```

```
XL1 = -1.0E30
ITT = 0

101 IF (ITT .NE. 0) GO TO 275
      GO TO 301
275 DO J = 1, 2
          P1(J) = XM1(3,J) / N
          P2(J) = XM2(3,J) / N
          PC(J) = (P1(J) + P2(J)) / 2
          U1(J) = XM1(1,J) / XM1(2,J)
          U2(J) = XM2(1,J) / XM2(2,J)
          R1(J) = XM1(2,J)
          R2(J) = XM2(2,J)
      END DO

DO 287 I = 1, 3
    DO 287 J = 1, 2
        XM1(I, J) = 0.
        XM2(I, J) = 0.
287 CONTINUE
!-------------- COMPUTE LIKELIHOOD --------------
301 CALL LNKLT
      T1 = ABS(XL1 - XLNT)
      IF (T1 .LE. TT) GO TO 326
304 XL1 = XLNT
      ITT = ITT + 1

      IF( (ITT - IEND) .LT. 0) GOTO 101
!       GO TO 101

326 IF( (T1 - TT) .LE. 0 ) GOTO 601
      IF( (ITT - IEND) .LT. 0) GOTO 304

601 DO 296 I = 1, 3
      DO 296 J = 1, 2
      XM1(I, J) = 0.
      XM2(I, J) = 0.
296 CONTINUE
RETURN
END
!=====================================================================


!=====================================================================
!--------------------
!-------------- LNKLT --------------
!--------------------
SUBROUTINE LNKLT
PARAMETER( MN = 2000 )
```

```
implicit double precision (a-h, o-z)
REAL PS(2), PS1(2), PS2(2), XLNC, D, BBIS
COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
       DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
       XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
       BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
       DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB

XLNT=0

DO 271 K = 1, N

DMI1=1.E+30
DMI2=1.E+30

DO 210 J = 1, 2
D1 = -(LOG(P1(J)) - DXT(K) * LOG(U1(J)) - XXT(K) / U1(J))
IF(D1 .GE. DMI1) GO TO 210
DMI1 = D1
JJ1 = J
210 DV1(J) = D1

DO 211 J = 1, 2
D2 = -(LOG(P2(J)) - DXC(K) * LOG(U2(J)) - XXC(K) / U2(J))
IF(D2 .GE. DMI2) GO TO 211
DMI2 = D2
JJ2 = J
211 DV2(J) = D2


PTOT1=0.
PTOT2=0.

DO 250 J = 1, 2
IF (J - JJ1) 220, 230, 220
220 D1 = DMI1 - DV1(J)
IF (D1 + DLIM) 222, 225, 225
222 PROB1(J) = 0.
GO TO 250
225 PROB1(J) = EXP(D1)
GO TO 240
230 PROB1(J) = 1.0
240 PTOT1 = PTOT1 + PROB1(J)
250 CONTINUE

DO 251 J = 1, 2
IF (J - JJ2) 221, 231, 221
```

```
221 D2 = DMI2 - DV2(J)
IF (D2 + DLIM) 223, 226, 226
223 PROB2(J) = 0.
GO TO 251
226 PROB2(J) = EXP(D2)
GO TO 241
231 PROB2(J) = 1.0
241 PTOT2 = PTOT2 + PROB2(J)
251 CONTINUE

XLNT = XLNT + ( LOG(PTOT1) - DMI1 ) + ( LOG(PTOT2) - DMI2 )

DO 271 J = 1, 2
XP1 = PROB1(J) / PTOT1
XP2 = PROB2(J) / PTOT2
XM1(1,J) = XM1(1,J) + XP1 * XXT(K)
XM1(2,J) = XM1(2,J) + XP1 * DXT(K)
XM1(3,J) = XM1(3,J) + XP1
XM2(1,J) = XM2(1,J) + XP2 * XXC(K)
XM2(2,J) = XM2(2,J) + XP2 * DXC(K)
XM2(3,J) = XM2(3,J) + XP2
271 CONTINUE

RETURN
END
!=================================================================


!--------------------
!-------------- NOCOB --------------
!--------------------
SUBROUTINE NOCOB
PARAMETER( MN = 2000 )
implicit double precision (a-h, o-z)
!LOGICAL FOUT
REAL XT1(1), XT2(1), PS(2), PS1(2), PS2(2), XLNC, D, BBIS
COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
        DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
        XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
        BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
        DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB

DATA TT, IEND /.01, 200/
XL1 = -1.0E30
ITT = 0

101 IF (ITT .NE. 0) GO TO 275
```

```
      GO TO 301
275 DO J = 1, 2
          P1(J) = XM1(3,J) / N
          P2(J) = XM2(3,J) / N
          PC(J) = (P1(J) + P2(J)) / 2
          U1(J) = XM1(1,J) / XM1(2,J)
          U2(J) = XM2(1,J) / XM2(2,J)
          R1(J) = XM1(2,J)
          R2(J) = XM2(2,J)
      END DO

DO 287 I = 1, 3
    DO 287 J = 1, 2
        XM1(I, J) = 0.
        XM2(I, J) = 0.
287 CONTINUE
!-------------- COMPUTE LIKELIHOOD --------------
301 CALL LNKLB
      T1 = ABS(XL1 - XLNT)
      IF (T1 .LE. TT) GO TO 326
304 XL1 = XLNT
      ITT = ITT + 1

      IF( (ITT - IEND) .LT. 0) GOTO 101
!       GO TO 101

326 IF( (T1 - TT) .LE. 0 ) GOTO 601
      IF( (ITT - IEND) .LT. 0) GOTO 304

601 DO 296 I = 1, 3
      DO 296 J = 1, 2
      XM1(I, J) = 0.
      XM2(I, J) = 0.
296 CONTINUE
RETURN
END
!=====================================================================


!=====================================================================
!--------------------
!-------------- LNKLB --------------
!--------------------
SUBROUTINE LNKLB
PARAMETER( MN = 2000 )
implicit double precision (a-h, o-z)
REAL PS(2), PS1(2), PS2(2), XLNC, D, BBIS
COMMON U1(2), U2(2), P1(2), P2(2), R1(2), R2(2), XX(MN), ID(MN), IR(MN), &
          DXT(MN), XXT(MN), DXC(MN), XXC(MN), &
```

```
      XM1(4,2), XM2(4,2), PROB1(2), PROB2(2), PC(2), DV1(2), DV2(2), XLNT,
&
      BXXT(MN), BXXC(MN), BDXT(MN), BDXC(MN), BR(1), BJ, DLIM, &
      DU(1), SE(1), BD, BOT(1), BUT(2), BUC(2), SUUT(2), SUUC(2), ODTT,
NS, NC, NEWF, N, NR, KB

XLNT=0

DO 271 K = 1, N

DMI1=1.E+30
DMI2=1.E+30

DO 210 J = 1, 2
D1 = -(LOG(P1(J)) - DXT(K) * LOG(U1(J)) - XXT(K) / U1(J))
IF(D1 .GE. DMI1) GO TO 210
DMI1 = D1
JJ1 = J
210 DV1(J) = D1

DO 211 J = 1, 2
D2 = -(LOG(P2(J)) - DXC(K) * LOG(U2(J)) - XXC(K) / U2(J))
IF(D2 .GE. DMI2) GO TO 211
DMI2 = D2
JJ2 = J
211 DV2(J) = D2

PTOT1=0.
PTOT2=0.

DO 250 J = 1, 2
IF (J - JJ1) 220, 230, 220
220 D1 = DMI1 - DV1(J)
IF (D1 + DLIM) 222, 225, 225
222 PROB1(J) = 0.
GO TO 250
225 PROB1(J) = EXP(D1)
GO TO 240
230 PROB1(J) = 1.0
240 PTOT1 = PTOT1 + PROB1(J)
250 CONTINUE

DO 251 J = 1, 2
IF (J - JJ2) 221, 231, 221
221 D2 = DMI2 - DV2(J)
IF (D2 + DLIM) 223, 226, 226
223 PROB2(J) = 0.
GO TO 251
226 PROB2(J) = EXP(D2)
```

```
GO TO 241
231 PROB2(J) = 1.0
241 PTOT2 = PTOT2 + PROB2(J)
251 CONTINUE

XLNT = XLNT + ( LOG(PTOT1) - DMI1 ) + ( LOG(PTOT2) - DMI2 )

DO 271 J = 1, 2
XP1 = PROB1(J) / PTOT1
XP2 = PROB2(J) / PTOT2
XM1(1,J) = XM1(1,J) + XP1 * XXT(K)
XM1(2,J) = XM1(2,J) + XP1 * DXT(K)
XM1(3,J) = XM1(3,J) + XP1
XM2(1,J) = XM2(1,J) + XP2 * XXC(K)
XM2(2,J) = XM2(2,J) + XP2 * DXC(K)
XM2(3,J) = XM2(3,J) + XP2
271 CONTINUE

RETURN
END
!=============================================================
```

Appendix B Publish Papers

MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1568

Published online 19 April 2013 in Wiley Online Library

Statistical inference on censored data for targeted clinical trials under enrichment design

Chen-Fang Chen,[a†] Jr-Rung Lin,[c†] and Jen-Pei Liu[a,b,d*]

For the traditional clinical trials, inclusion and exclusion criteria are usually based on some clinical endpoints; the genetic or genomic variability of the trial participants are not totally utilized in the criteria. After completion of the human genome project, the disease targets at the molecular level can be identified and can be utilized for the treatment of diseases. However, the accuracy of diagnostic devices for identification of such molecular targets is usually not perfect. Some of the patients enrolled in targeted clinical trials with a positive result for the molecular target might not have the specific molecular targets. As a result, the treatment effect may be underestimated in the patient population truly with the molecular target. To resolve this issue, under the exponential distribution, we develop inferential procedures for the treatment effects of the targeted drug based on the censored endpoints in the patients truly with the molecular targets. Under an enrichment design, we propose using the expectation–maximization algorithm in conjunction with the bootstrap technique to incorporate the inaccuracy of the diagnostic device for detection of the molecular targets on the inference of the treatment effects. A simulation study was conducted to empirically investigate the performance of the proposed methods. Simulation results demonstrate that under the exponential distribution, the proposed estimator is nearly unbiased with adequate precision, and the confidence interval can provide adequate coverage probability. In addition, the proposed testing procedure can adequately control the size with sufficient power. On the other hand, when the proportional hazard assumption is violated, additional simulation studies show that the type I error rate is not controlled at the nominal level and is an increasing function of the positive predictive value. A numerical example illustrates the proposed procedures. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: targeted clinical trials; enrichment design; exponential hazard; EM algorithm

1. INTRODUCTION

The disease targets at the molecular level can be identified using the state-of-the-art biotechnology such as microarray, single nucleotide polymorphisms, or next-generation sequencing. Treatments specific for the patients with the identified molecular targets can be developed and patients benefit from the treatment without suffering serious or even fatal toxicity. As a result, personalized medicines finally become a reality. Development of targeted treatments is a translational science, which involves translation from the accuracy of diagnostic devices for the molecular targets to the effectiveness and safety of the treatment modality for the patient population with the targets. Therefore, evaluation of targeted treatments consists of evaluation not only the efficacy and safety of the treatment modality but also the accuracy of diagnostic device for the molecular targets. To address these issues, the United States Food and Drug Administration (US FDA) issued *Draft Drug-Diagnostic Co-Development Concept Paper* and *Draft Guidance In Vitro Diagnostic Multivariate Index Assays*, respectively, in April 2005 and in July 2007 [1, 2].

One of the designs introduced in the US FDA *Drug-Diagnostic Co-development Concept Paper* for evaluation of the targeted treatments is the enrichment design [3]. Unlike the traditional trials, under the enrichment design, the targeted clinical trials consist of two phases. The first phase is the enrichment phase in which in addition to the inclusion/exclusion criteria based on some clinical endpoints, signs or symptoms, each patient is screened by a diagnostic device for detection of the predefined molecular targets. Then, the patients with a positive result by the diagnostic device are randomized to receive either the targeted treatment or the concurrent control.

However, no diagnostic test is perfect with 100% positive predictive value (PPV). In addition, measures for diagnostic accuracy such as PPV are in fact estimators with variability. Liu and Chow [4] pointed out that some patients enrolled in targeted clinical

[†] *Equal contribution*

[a] *Division of Biometry, Department of Agronomy, National Taiwan University, Taipei, Taiwan*

[b] *Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan*

[c] *Clinical Informatics and Medical Statistics Research Center and Graduate Institute of Clinical Medical, Chang Gung University, Tao-Yuan, Taiwan*

[d] *Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan*

**Correspondence to: Jen-Pei Liu, Division of Biometry, Department of Agronomy, National Taiwan University, Taipei, Taiwan.*
E-mail: jpliu@ntu.edu.tw

trials under the enrichment design might not have the specific targets and hence the treatment effects of the drug for the molecular targets could be underestimated. Liu, Lin and Chow [5] and Liu and Lin [6] proposed to apply the expectation–maximization (EM) algorithm in conjunction with the bootstrap technique to incorporate the uncertainty on the inaccuracy of the diagnostic device in detection of the molecular targets for the inference of the treatment effects for the targeted therapy for the binary and continuous endpoints under the enrichment design. On the other hand, most of the current targeted drugs are for the treatment of cancers such as breast cancer, lung cancer, or colorectal cancer. The efficacy endpoints for evaluation of targeted therapies in cancer trials are censored endpoints such as overall survival or progression free survival. Currently, literature for the statistical methods taking into account the variability and accuracy of the diagnostic device for molecular targets for the inference based on censored endpoints is scarce.

Therefore, under the assumptions enrichment design and one-parameter exponential distribution, we propose to apply EM algorithm [7, 8] with the bootstrap method [9] to incorporate the uncertainty of the PPV of the diagnostic device for statistical inference of the treatment effect of the targeted drugs based on the censored data. In the next section, under the one-parameter exponential distribution, the currents method for the inference of the treatment effect between the targeted therapy and control based on censored data are reviewed. Our proposed EM-bootstrap method for incorporation of the variability and inaccuracy of the diagnostic devices for molecular targets are provided in Section 3. A simulation study was conducted to empirically investigate the performance of the proposed procedures in terms of the bias and coverage probability of confidence intervals (CIs) for estimation, and size and power for testing procedures. The results of the simulation study are given in Section 4. Section 5 provides a numerical example to illustrate the proposed method. Discussion and final remarks are provided in the last section.

2. CURRENT METHODS

In the following, the situation where a particular molecular target involved with the pathway in pathogenesis of the disease has been identified, and a validated diagnostic device is available for detection of the identified molecular target is considered. Furthermore, we suppose that (i) this device is only for detection of the molecular target and is not for prognosis of clinical outcomes of patients, (ii) and a test drug for the particular molecular target is available and is currently being developed. Under enrichment design, to evaluate the treatment effects of the targeted test treatment in the patient population truly with the molecular target is one of the objectives of targeted clinical trials. Following the enrichment design, a two-group parallel design is considered. The patients with a positive result by the diagnostic device are randomized to receive either the molecular targeted test treatment (T) or an untargeted concurrent control treatment (C). The primary endpoint considered here is the censored data. The data set that we have interest in analyzing consists of observations of the random variables $(Y_{ij}, \delta_{ij}), j = 1, \ldots, n_i; i = T, C$, where Y_{ij} is the observed survival time for the event or censored, and δ_{ij} is an indicator variable that takes 1 = event and 0 = censored. Therefore, under the assumption of the one-parameter exponential distribution, the probability density functions of the two exponential distributions are $f_i(y_{ij}) = \lambda_i e^{-\lambda_i y_{ij}}$, i $= T, C$ with a mean survival time $\mu_i = \lambda_i^{-1}$, where λ_i is the hazard rate in the ith treatment group. The corresponding hazard functions are $h_i(y_{ij}) = \lambda_i$, and the survival functions are $S_i(y_{ij}) = e^{-\lambda_i y_{ij}}$, for $0 \leqslant y_{ij} < \infty$. The maximum likelihood estimators for the mean survival time μ_i are $\hat{\mu}_i = T_i^*/r_i$, where $T_i^* = \sum_{j=1}^{n_i} y_{ij}$ and $r_i = \sum_{j=1}^{n_i} \delta_{ij}$, $i = T, C$ are the totals of uncensored and censored survival times and the numbers of actual death time in each group [10]. Table I gives the expected values of Y_{ij} by treatment and diagnostic result of the molecular target. In Table I, μ_{T+}, μ_{C+} (μ_{T-}, μ_{C-}) are the mean survival times of test and control groups for the patients with (without) the molecular target. The hypothesis for detection of treatment difference in the patient population truly with the molecular target is the hypothesis of interest:

$$\mathrm{H}_0 : \mu_{T+} - \mu_{C+} = 0 \quad \text{versus} \quad \mathrm{H}_a : \mu_{T+} - \mu_{C+} \neq 0 \qquad (1)$$

Let $\bar{y}_T$ and $\bar{y}_C$ be the sample mean survival time of test and control treatments, respectively. In addition, some patients with a positive diagnostic result may not have the molecular target, in fact, because no diagnostic test is perfect for diagnosis of the molecular target of interest without error. It follows that

$$\begin{aligned} E(\bar{y}_T - \bar{y}_C) &= \gamma(\mu_{T+} - \mu_{C+}) + (1-\gamma)(\mu_{T-} - \mu_{C-}) \\ &= \gamma\left(\lambda_{T+}^{-1} - \lambda_{C+}^{-1}\right) + (1-\gamma)\left(\lambda_{T-}^{-1} - \lambda_{C-}^{-1}\right) \end{aligned} \qquad (2)$$

where γ is the PPV.

The expected values of the difference in sample mean survival time consists of two parts. The first part is the treatment effects of the molecular target drug in the positive diagnosis patients who truly have the molecular target of interest. The second part is the treatment effects of the patients with a positive diagnosis, but they do not have the molecular target in fact. The assumption based on that the efficacy of the targeted treatment in the patients truly with the molecular target is greater than those without the target is the reason for developing the targeted treatment. Moreover, the targeted treatment is also expected to be

Table I. Population mean survival time by treatment and diagnosis.

Positive diagnosis	True target condition	Indicator of diagnostic	Test group	Control group	Difference
+	+	γ	μ_{T+}	μ_{C+}	$\mu_{T+} - \mu_{C+}$
	−	$1-\gamma$	μ_{T-}	μ_{C-}	$\mu_{T-} - \mu_{C-}$

γ is the positive predictive value.

more efficacious than the untargeted control in the patient population truly with the molecular targets. It follows that $\mu_{T+}-\mu_{C+} > \mu_{T-} - \mu_{C-}$. As a result, the difference in sample mean survival times obtained under the enrichment design for targeted clinical trials actually underestimates the true treatment effect of the molecular targeted drug in the patient population truly with the molecular target of interest. As it can be seemed from (2), the bias of the difference in sample mean survival times decreases as the PPV increases.

The current approach without considering the PPV and the variability of its estimate is based on the following hypothesis in terms of the log of the hazard ratio, $\theta = \log(\lambda_T/\lambda_C) = \log(\mu_C/\mu_T)$:

$$\mathrm{H}_0 : \theta = 0 \text{ versus } \mathrm{H}_a : \theta \neq 0.$$

The maximum likelihood estimate $\hat{\theta}$ has an approximate normal distribution with estimated variance of $\frac{1}{r_T} + \frac{1}{r_C}$. The traditional *z*-test approach is to reject the null hypothesis at the α significance level if

$$z = \left| \frac{\hat{\theta}}{se\left(\hat{\theta}\right)} \right| \geqslant z_{\alpha/2}$$

On the basis of the aforementioned *z*-statistic, the corresponding 100 $(1-\alpha)$% CI can be obtained as follows

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{1}{r_T} + \frac{1}{r_C}}$$

3. THE PROPOSED PROCEDURE

All patients randomized under the enrichment design have a positive diagnosis, but the true status of the molecular target for individual patients in the target clinical trial is in fact unknown. It follows that Y_{ij} are independently distributed as a mixture of two exponential distributions with hazard λ_{i+} and λ_{i-}, respectively

$$\varphi(y_{ij},\ \delta_{ij}|\lambda_{i+})^{\gamma}\varphi(y_{ij},\ \delta_{ij}|\lambda_{i-})^{1-\gamma} \qquad i = T, C\ ;\ j = 1, \ldots, n_i$$

where $\varphi(.|.)$ denotes the density function of an exponential variable with events happened or survival function of an exponential variable with events censored. However, γ is an unknown PPV that must be estimated from the data. Therefore, the data obtained from the targeted clinical trials are incomplete because the true status of the molecular target of the patients is unknown. We apply the EM algorithm to estimate the treatment effects for the patients truly with the molecular target by incorporating the estimates of the PPV of the device obtained from the diagnostic effectiveness trials as the initial values.

For each patient, we have a set of variables $(Y_{ij},\ \delta_{ij},\ X_{ij})$, where Y_{ij} is the observed survival time for the event or the censored time and δ_{ij} is an indicator variable that takes 1 = event and 0 = censored for patient *j* in treatment *i*; X_{ij} is the latent variable indicting the true status of the molecular target of patient *j* in treatment *i*; $j = 1, \ldots, n_i$, $i = T, C$. X_{ij} is an indicator variable with value of 1 for the patients with the molecular target and with a value of 0 for the patients without the target. In addition, X_{ij} are assumed i.i.d. Bernoulli random variables with probability with the molecular target being γ. Let $\boldsymbol{\Psi}$ be the vector containing all unknown parameters and $(\mathbf{y}_{\text{obs}},\ \boldsymbol{\delta}_{\text{obs}})$ denote the vectors of the observed primary efficacy endpoints from the targeted clinical trial, where

$$\boldsymbol{\Psi} = (\gamma, \lambda_{T+}, \lambda_{T-}, \lambda_{C+}, \lambda_{C-})'$$

and

$$\mathbf{y}_{\text{obs}} = \left(y_{T1}, \ldots, y_{Tn_T}, y_{C1}, \ldots, y_{Cn_C}\right)'$$

$$\boldsymbol{\delta}_{\text{obs}} = \left(\delta_{T1}, \ldots, \delta_{Tn_T}, \delta_{C1}, \ldots, \delta_{Cn_C}\right)'$$

It follows that the complete-data log-likelihood function for $\boldsymbol{\Psi}$ is given by

$$\begin{aligned}\log L_c(\boldsymbol{\Psi}) = &\sum_{j=1}^{n_T} x_{Tj}\left[\log\gamma + \log\varphi(y_{Tj},\ \delta_{Tj}|\lambda_{T+})\right] \\ &+ \sum_{j=1}^{n_T}(1-x_{Tj})\left[\log(1-\gamma) + \log\varphi(y_{Tj},\ \delta_{Tj}|\lambda_{T-})\right] \\ &+ \sum_{j=1}^{n_C} x_{Cj}\left[\log\gamma + \log\varphi(y_{Cj},\ \delta_{Cj}|\lambda_{C+})\right] \\ &+ \sum_{j=1}^{n_C}(1-x_{Cj})\left[\log(1-\gamma) + \log\varphi(y_{Cj},\ \delta_{Cj}|\lambda_{C-})\right]\end{aligned}$$

Furthermore, from the previous diagnostic effectiveness trials, an estimate of the PPV of the device is known. Therefore, at the initial step of the EM algorithm for estimation, the treatment effects in the patients truly with the molecular target, the observed latent variable X_{ij} are generated as i.i.d. Bernoulli random variables with γ estimated from the diagnostic effectiveness trial. The procedures for implementation of the EM algorithm in conjunction with the bootstrap procedure for inference of θ in the patient population truly with the molecular target are provided in the Appendix.

Let $\hat{\theta}$ be the estimator for the treatment effects in the patients truly with the molecular target obtained from the EM algorithm. Let S_B^2 denote the estimator of the variance of $\hat{\theta}$ obtained by the bootstrap procedure as demonstrated in the Appendix. It follows that the null hypothesis is rejected and the efficacy of the molecular targeted test drug is different from that of the control in the patient population truly with the molecular target at the α level if

$$z = \left| \frac{\hat{\theta}}{\sqrt{S_B^2}} \right| \geqslant z_{\alpha/2}$$

where $z_{\alpha/2}$ is the upper 100 $(\alpha/2)$ percentile of a standard normal distribution. The corresponding 100 $(1-\alpha)$% asymptotic CI for $\theta = \log(\lambda_{T+}/\lambda_{C+})$ can be constructed as

$$\hat{\theta} \pm z_{1-\alpha/2}\sqrt{S_B^2}$$

4. SIMULATION STUDY

The random samples of patients with or without the molecular target were generated from the Bernoulli distribution with probability γ. Then, the units are randomized in a 1:1 ratio to the test group or control group. We generate the survival times of the groups from exponential distributions with the specified parameters λ_{i+} and λ_{i-} according to the status of the molecular target,

Table II. Relative bias and the coverage probability.

			PPV							
			0.5		0.6		0.7		0.8	
n	CR	HR	Current	EM	Current	EM	Current	EM	Current	EM
300	20%	0.85	7.96†	−3.73	6.42	−3.14	4.60	−2.52	3.13	−1.74
			0.8632‡	0.9768	0.8942	0.9732	0.9202	0.9690	0.9360	0.9610
		0.80	11.01	−3.78	8.52	−2.95	6.37	−2.46	4.12	−2.05
			0.7874	0.9760	0.8548	0.9732	0.8978	0.9696	0.9290	0.9604
		0.75	14.33	−2.93	11.13	−2.33	8.21	−2.32	5.12	−2.04
			0.6870	0.9752	0.7926	0.9724	0.8564	0.9674	0.9174	0.9660
		0.70	17.90	−1.76	13.59	−2.13	10.17	−1.79	6.36	−1.79
			0.5610	0.9772	0.7070	0.9742	0.9096	0.9690	0.8930	0.9604
	30%	0.85	8.24	−4.35	6.20	−3.85	4.74	−2.88	3.19	−1.76
			0.8652	0.9758	0.9094	0.9736	0.9238	0.9630	0.9396	0.9630
		0.80	11.05	−4.84	8.75	−3.74	6.46	−2.84	4.24	−2.26
			0.8100	0.9782	0.8602	0.9682	0.9048	0.9674	0.9322	0.9630
		0.75	14.43	−4.29	11.13	−3.28	7.99	−2.81	5.25	−2.47
			0.7140	0.9750	0.8064	0.9738	0.8760	0.9680	0.9136	0.9662
		0.70	17.87	−3.76	13.61	−3.11	9.90	−2.87	6.41	−2.54
			0.6114	0.9780	0.7390	0.9766	0.8348	0.9682	0.8986	0.9616

PPV, positive predictive value; CR, censored rate; HR, hazard ratio; EM, expectation–maximization.
† Relative bias (%).
‡ Coverage probability.

$i = T,\ C$, respectively, and also the censoring times from exponential distributions with hazards corresponding to a common censoring proportion. We assume the placebo control is employed in the targeted clinical trial. In addition, the molecularly targeted test drug is not effective in the patients truly without the target either. Therefore, in the simulation, λ_{T-}, λ_{C+}, and λ_{C-} are assumed equal and set to be a generic value of 1. The PPV is set to be 0.5, 0.6, 0.7, and 0.8. We employed random right-censoring mechanism in the simulation study. Each unit has a potential censoring time C_i' and a potential lifetime T_i', which are assumed to be independent random variables. Consider $Y_i = \min\{C_i',\ T_i'\}$ and an indicator δ_i for the type of event (censored or death). If T' and C' are independent exponential random variables with parameters λ_1 and λ_2, respectively, then $P\{T' < C'\} = \lambda_1/(\lambda_1 + \lambda_2)$. By matching survival and censoring times in n pairs in sequence, we generated the observed times, $y = \min(t',\ c')$. The censoring proportions considered in the simulation study are 0, 0.1, 0.2, 0.3, and 0.4. The sample sizes are set as 300, 600, and 900 per group. The power of the proposed testing procedure was investigate at $\lambda_{T+} = 0.70$, 0.75, 0.80, and 0.85. The bias of the estimators and the coverage probability of the 95% CI were investigated. A total of 5000 random samples were generated, and the number of the

Table III. Comparison of empirical powers when censored rate = 20%.

		PPV							
		0.5		0.6		0.7		0.8	
n	HR	Current	EM	Current	EM	Current	EM	Current	EM
300	0.85	0.1586	0.1592	0.1998	0.2248	0.2492	0.2934	0.3046	0.3418
	0.80	0.2626	0.2870	0.3384	0.3760	0.4242	0.4900	0.5138	0.5824
	0.75	0.3930	0.4300	0.5118	0.5684	0.6268	0.6976	0.7432	0.8042
	0.70	0.5552	0.5864	0.7008	0.7324	0.8056	0.8470	0.8930	0.9252
600	0.85	0.2652	0.3822	0.3364	0.4700	0.4292	0.5688	0.5340	0.6342
	0.80	0.4506	0.6464	0.5838	0.7576	0.6944	0.8236	0.8138	0.8936
	0.75	0.6544	0.8366	0.7982	0.9188	0.9052	0.9642	0.9530	0.9814
	0.70	0.8534	0.9440	0.9450	0.9862	0.9814	0.9958	0.9944	0.9988
900	0.85	0.3718	0.5296	0.4772	0.6358	0.5938	0.7364	0.6986	0.7888
	0.80	0.6076	0.8142	0.7582	0.9010	0.8630	0.9600	0.9340	0.9704
	0.75	0.8248	0.9460	0.9272	0.9854	0.9750	0.9952	0.9946	0.9984
	0.70	0.9536	0.9902	0.9932	0.9990	0.9990	1.0000	0.9998	1.0000

HR, hazard ratio; EM, expectation–maximization; PPV, positive predictive value.

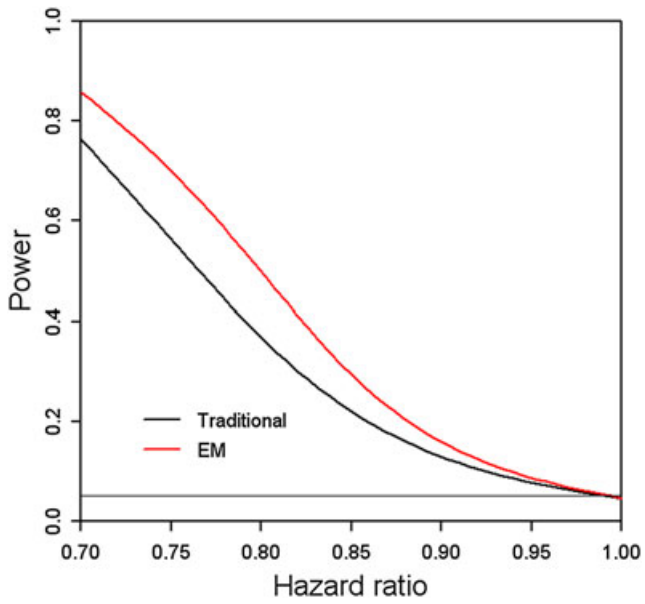


Figure 1. The empirical power curve when the PPV is 0.6, $n = 300$ and CR = 10%.

bootstrap samples was set to be 1000. FORTRAN 95 and IMSL's STAT/LIBRARY FORTRAN subroutines were used in the simulation study.

For estimation, we investigate the relative bias of the estimators and the coverage probability of the 95% CI. For hypothesis testing, the performance measures include empirical size and power. The relative bias is estimated as the average of 100 times the differences between the estimates and the true value of θ divided by the true value over 5000 simulated samples. The coverage probability is calculated as the proportion of the 5000 95% CIs that contain θ. The size and power were computed as the proportion of the 5000 samples that the null hypothesis is rejected for the two-sided test at the 5% significance level. For a 95% confidence level, with 5000 simulation random samples implies that 95% of the empirical coverage probabilities will be within 0.94396 and 0.95604 if the proposed methods provide sufficient coverage probability. In addition, for a 5% nominal significance level, a simulation study with 5000 random samples implies that 95% of empirical sizes will be within 0.04396 and 0.05604 if the proposed methods can adequately control the size at the nominal level of 0.05.

The simulation results on estimation for censored rate of 20% and 30% for sample size of 300 are provided in Table II. The simulation results of other combinations are similar. Hence, they are not presented here. The results in Table II demonstrate that the absolute relative bias of the estimator for traditional hazard ratio for the patients truly with the molecular target by the current method ranges from 3.1% to more than 17%. It increases as the PPV decreases. On the other hand, the absolute relative bias of the estimator for hazard ratio for the patients truly with the molecular target obtained by the EM algorithm does not exceed 5.0%, although most of them are smaller than 3.0%. The variability has little impact on the bias of both methods. Consequently, the empirical coverage probabilities of the 95% CI of the current method can be as low as 56.1% when the PPV is 50%, censored rate is 20%, hazard ratio is 0.70 and n is 300. The coverage probability of the 95% CI by the current method is an increasing function of the PPV. None of the 32 coverage probabilities of the 95% CIs by the current method in Table II exceed 0.95. On the contrary, no coverage probability of the EM method is below 0.95. Therefore, the proposed procedures for estimation of the treatment effects in the patients population truly with the molecular target by the EM algorithm is not only nearly unbiased but also provide sufficient coverage probability.

The simulation results on sizes reveal that all empirical sizes of both the current method and the proposed EM procedure for testing the hypothesis are within 0.0452 and 0.0610. These results demonstrate that both methods can adequately control the size at its nominal level of 5% under the null hypothesis. The empirical powers of the simulation for the hazard ratio when the censored rate is 20% are given in Table III. The results of power for the other combinations are consistent with the other presented in Table III. They are not provided here. In addition, Figure 1 presents the power curves when $n = 300$, censored rate = 10%, and PPV is 0.6. In Table III, we observe that the power of the current method is an increasing function of the PPV. For both the methods, the power increases as the sample size increases. However, the simulation results clearly demonstrate that the proposed testing procedure for the treatment effects based on the EM algorithm in the patient population truly with the molecular target is uniformly more powerful than the current method as depicted in Figure 1.

5. NUMERIC EXAMPLE

A targeted drug is developed for the treatment of the patient with a certain cancer whose specific biomarker is over-expressed as measured by an immunohistochemical assay. Suppose that the immunohistochemical assay has a PPV of 0.75. From previous studies, the hazard ratios for the patients truly with and without the biomarker are 0.7 and 1.26, respectively, and are given in Table IV. Under the enrichment design, 480 patients with positive test results were randomized in 1:1 ratio to receive either the targeted drug plus the standard chemotherapy or the standard chemotherapy. The censored rate is assumed to be 30%. Table V provides the point estimates of hazard ratio for mortality between the two groups, and their standard error and 95% CIs for the risk when PPV is 0.75.

Table IV. Treatment effects as a function of a specific biomarker overexpression.

IHC assay result	No. of patients	Hazard ratio for mortality (95% CI)
Test result +	469	0.80 (0.64, 1.00)
True status −	120	1.26 (0.82, 1.94)
True status +	349	0.70 (0.51, 0.90)

IHC, immunohistochemical.

Table V. Point and interval estimator of hazard ratio for mortality.

	PPV = 0.75	
Results	Current	EM
Hazard ratio for mortality	0.8097	0.7108
SE	0.1082	0.1626
95% Lower confidence limit	0.6550	0.5168
95% Upper confidence limit	1.0009	0.9777

PPV, positive predictive value; EM, expectation–maximization.

When PPV is 0.75, the traditional approach without consideration of inaccuracy of diagnostic device yields the estimate of hazard ratio for mortality of 0.8097 with a 95% CI from 0.6550 to 1.0009. Because the 95% CI contains 1, the observed hazard ratio of death is not statistically significant and the targeted drug does not prove its superior efficacy over chemotherapy alone at the 5% level. The reason for the failure of the targeted drug is that 25% positive patients randomized do not have the molecular targets. On the other hand, our proposed EM method provides the estimated hazard ratio of mortality is 0.7108. The 95% CI for the hazard ratio of mortality is (0.5168, 0.9777), which does not contain 1. As a result, the efficacy of the targeted drug can be concluded superior to the control group based on the hazard ratio.

6. DISCUSSION

Under the enrichment design, all patients must have a positive diagnosis for the molecular target by the diagnostic device to be randomized to receive either the targeted drug or the control treatment in the targeted clinical trials. However, no diagnostic device is perfect with 100% PPV. The PPV is an increasing function of prevalence. For example, the molecular targeted drug Xalkori is approved by the US FDA for the treatment of non-small cell lung cancer according to the *Draft Guidance on In Vitro Companion Diagnostic Devices* [11]. However, the prevalence rate for patients with the target is only 5%. The PPV of the diagnostic devices for the target will be low. Therefore, the treatment effect of Xalkori may be underestimated. Because the PPV of the *in vitro* companion devices will not be high if the prevalence rate of the molecular target is low even the device is approved by the US FDA. The treatment effect of the patients truly with the molecular target is underestimated. The magnitude of underestimation is inversely proportional to the magnitude of the PPV.

As a result, the current estimation method may produce a biased estimator for the treatment effects of the targeted test drug in the patients truly with the molecular target. Hence, we propose estimation and testing procedures by application of the EM algorithm to incorporate information of the PPV for inference of the treatment effects in the patient population truly with the molecular target. On the results of our simulations, the proposed estimation method is nearly unbiased and can provide a sufficient coverage probability. The proposed testing procedure also can adequately control the type I error rate at the nominal level and is uniformly more powerful than the current method.

In the application of the EM algorithm, selection of initial values for μ_{i+} and μ_{i-}, $i = T,\ C$ is important. However, the estimates of efficacy for the current available therapies (control) are known for most of diseases. In addition, the expected magnitude of increment of efficacy over the control provided by the targeted drug is also specified in the protocol for the sample size determination. Therefore, a range of reasonable initial values can be determined for the EM algorithm from this information. One method for selection of initial values to generate μ_{i+} and μ_{i-} from a exponential distribution is to employ the sample mean survival time of the observed data, $\bar{y}_T$ and $\bar{y}_C$. In other words, the traditional sample mean survival times, $\bar{y}_T$ and $\bar{y}_C$, of the test and control treatments are reasonable initial values for the proposed method.

We consider the exponential parametric model that satisfies the proportional hazard assumption. However, the proportionality assumption may not hold in practice. We conducted an additional simulation study to investigate the impact of violation of the proportional hazard assumption on performance of our proposed method. Because the hazard function of the log-normal distribution is nonmonotonic and changes over time, we generated the survival times from the log-normal distribution for the situation when the proportional hazard assumption is violated. We further consider the following two cases in the simulation study:

Case 1: Because we assume that the molecular targeted test drug is ineffective in the patients truly without the target and the placebo is ineffective in the patients truly with and without the target either. Therefore, the survival times of the test drug group for the patients truly with the molecular target are generated from the exponential distribution. However, the survival times of the test drug group for the patients truly without the molecular target are generated from the log-normal distribution, and those of the control group are also generated from the log-normal distribution.

Case 2: The survival times of the patients truly with and without the target assigned either to the molecular target test drug or to placebo were all generated from the log-normal distribution.

The sample size for the additional simulation study is 600 per group with a censoring proportion of 0.2. The hazard ratios are 1.0, 0.80, and 0.75 and the PPVs are 0.5, 0.6, 0.7, and 0.8. The results of this additional simulation are given in Tables VI and VII. Table VI presents the relative bias and coverage probability of the current and the proposed EM method. The results in Table VI reveal that both the current and EM methods are biased.

For Case 1, when the hazard ratio is 1, both methods underestimate the true hazard ratio with a relative bias ranging from −3.16% to −4.8%. When the hazard ratio is 0.8 or 0.75, the current method over-estimates the true hazard ratios, whereas the EM method produces under-estimated estimates. The relative bias of the two methods increases as the PPV decreases. The magnitude of the bias of the current method decreases from around 10.0% when PPV is 0.5 to around 1.0% when PPV is 0.8. On the other hand, the bias of the EM method changes from −15.1% when PPV is 0.5 to −8.44% when PPV is 0.8.

For Case 2, when the hazard ratio is 1, the absolute relative biases of the both methods do not exceed 0.26%. When the hazard ratio is 0.8 or 0.75, the current method provides over-estimated estimates, whereas the EM method under-estimates the true hazard ratios. The relative bias of the two methods increases as the PPV decreases. The magnitude of the bias of

Table VI. Relative bias and the coverage probability when censored rate = 20%, $n = 600$.

		PPV							
		0.5		0.6		0.7		0.8	
HR		Current	EM	Current	EM	Current	EM	Current	EM
Case 1									
1.00	Relative bias	−3.16	−3.22	−3.37	−3.42	−4.18	−4.21	−4.77	−4.80
	Coverage probability	0.9142	0.9114	0.9052	0.9056	0.8890	0.8872	0.8732	0.8712
0.80	Relative bias	7.74	−15.10	4.58	−11.88	1.75	−9.70	−1.03	−8.44
	Coverage probability	0.7770	0.7782	0.8838	0.7932	0.9280	0.8068	0.9374	0.8116
0.75	Relative bias	10.45	−15.03	6.57	−11.92	3.28	−9.64	0.09	−8.57
	Coverage probability	0.6562	0.7956	0.8202	0.8012	0.9116	0.8122	0.9458	0.8054
Case 2									
1.00	Relative bias	0.26	0.19	−0.03	−0.09	0.16	0.12	0.05	0.02
	Coverage probability	0.9422	0.9418	0.9360	0.9340	0.9330	0.9314	0.9362	0.9350
0.80	Relative bias	11.24	−38.66	8.90	−28.86	6.44	−20.31	4.18	−13.03
	Coverage probability	0.6172	0.6676	0.7338	0.7250	0.8262	0.7918	0.8854	0.8344
0.75	Relative bias	14.43	−38.77	11.16	−29.59	8.28	−22.23	5.32	−12.93
	Coverage probability	0.4520	0.6632	0.6262	0.7376	0.7538	0.7900	0.8628	0.8484

PPV, positive predictive value; HR, hazard ratio; EM, expectation–maximization; RB, Relative bias; CP, Coverage probability.

Table VII. Comparison of empirical powers when censored rate = 20%, $n = 600$.

		PPV							
		0.5		0.6		0.7		0.8	
	HR	Current	EM	Current	EM	Current.	EM	Current	EM
Case 1	1.00	0.0858	0.0886	0.0948	0.0944	0.1110	0.1128	0.1268	0.1288
	0.80	0.6270	0.8074	0.7828	0.8978	0.8780	0.9374	0.9454	0.9686
	0.75	0.8250	0.9424	0.9234	0.9760	0.9742	0.9908	0.9906	0.9970
Case 2	1.00	0.0578	0.0582	0.0640	0.0660	0.0670	0.0686	0.0638	0.0650
	0.80	0.4380	0.7400	0.5694	0.7834	0.6960	0.8286	0.7902	0.8574
	0.75	0.6542	0.8722	0.7922	0.9126	0.8870	0.9542	0.9458	0.9728

PPV, positive predictive value; HR, hazard ratio; EM, expectation–maximization.

the current method decreases from 14.43% when PPV is 0.5 to around 4.0% when PPV is 0.8 On the other hand, the bias of the EM method changes from −38.77% when PPV is 0.5 to -12.93% when PPV is 0.8.

For case 1, when the hazard ratio is 1, the fluctuation of the magnitude of the empirical coverage probability over PPV is quite small. However, when the hazard ratio is 0.75 or 0.8, the empirical coverage probability of both methods is an increasing function of PPV. For the current method, the empirical coverage probability increase from around 0.65 when PPV is 0.5 to 0.94 when PPV is 0.8. On the other hand, the empirical coverage probability of the EM method increases from about 0.77 for PPV being 0.5 to 0.81 for PPV being 0.8.

For case 2, when the hazard ratio is 1, the empirical coverage probabilities of the both methods range from 0.9314 to 0.9422 with very minor fluctuations. When the hazard ratio is 0.8 or 0.75, the empirical coverage probability of both methods is again an increasing function of PPV. However, the empirical coverage probabilities of the both methods are comparable for different values of PPVs.

Table VII presents the empirical powers of the current and the proposed EM method. For Case 1, when the hazard ratio is 1, all empirical sizes of both the current method and the proposed EM procedure are inflated with a range from 0.0858 to 0.1288. These results demonstrate that both methods cannot control the size at its nominal level. Both of the empirical sizes and powers are increasing functions of the PPV. For Case 2, when the hazard ratio is 1, all empirical sizes of both the current method and the proposed EM procedure are from 0.0578 to 0.0686. The sizes of both methods are slightly inflated. Similar to Case 1, the empirical sizes and powers are increasing functions of the PPV. In summary, when the proportional hazard assumption is violated, both the current and EM methods produce biased estimates. The magnitude of the bias of the EM method seems to be larger than that of the current method. In addition, both methods can not control the empirical size.

The Cox proportional hazard model is a semi-parametric method which does not need to assume a particular form of probability distribution for the survival data. As a result, the hazard function is not restricted to a specific functional form, and hence the model has flexibility and widespread applications. On the other hand, if the assumption of a particular probability distribution for the data is valid, inference based on such an assumption will be more precise. In particular, estimates of relative hazards will tend to have smaller standard errors than they would in the absence of a distribution assumption. The exponential distribution is the simplest parametric model that satisfied the assumption of proportional hazard. This is one of the seasons why the exponential distribution is considered in this paper. However, the inference procedures for the treatment effects for the patients truly with the molecular targets based on the Cox semi-parametric proportional hazard model requires further research.

REFERENCES

[1] US FDA. *The Draft Concept Paper on Drug-Diagnostic Co-Development*. The US Food and Drug Administration: Rockville, MD, 2005.
[2] US FDA. *Draft Guidance on In Vitro Diagnostic Multivariate Index Assays*. The US Food and Drug Administration: Rockville, MD, 2007.
[3] Chow SC, Liu JP. *Design and Analysis of Clinical Trials* (2nd Ed). Wiley: New York, 2004.
[4] Liu JP, Chow SC. Issues on the diagnostic multivariate index assay and targeted clinical trials. *Journal of Biopharmaceutical Statistics* 2008; **18**:167–182.
[5] Liu JP, Lin JR, Chow SC. Inference on treatment effects for targeted clinical trials under enrichment design. *Pharmaceutical Statistics* 2009; **8**:356–370.
[6] Liu JP, Lin JR. Statistical methods for targeted clinical trials under enrichment design. *Journal of the Formosan Medical Association* 2008; **107**:S35–S42.
[7] Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society, Series B* 1977; **39**:1–38.
[8] McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. Wiley: New York, 1997.
[9] Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman and Hall: New York, 1993.
[10] Collett D. *Modeling Survival Data in Medical Research* (2nd Ed). Chapman and Hall: London, 2003.
[11] US FDA. *Draft Guidance on In Vitro Companion Diagnostic Devices*. The US Food and Drug Administration: Rockville, MD, 2011.

APPENDIX

At the initial step of the EM algorithm, the observed latent variable X_{ij} are generated as i.i.d. Bernoulli random variables with the PPV γ estimated by that obtained from the diagnostic effectiveness trial. At the $(k+1)$st iteration, the E-step requires the calculation of the conditional expectation of the complete-data log-likelihood $L_c(\Psi)$, given the observed data $(y_{\text{obs}}, \delta_{\text{obs}})$, using currently fitting $\widehat{\Psi}^{(k)}$ for Ψ.

$$Q\left(\Psi; \widehat{\Psi}^{(k)}\right) = E_{\Psi(k)}\{\log L_c(\Psi)|y_{\text{obs}}, \delta_{\text{obs}}\}$$

Because $\log L_c(\Psi)$ is a linear function of the unobservable component labeled variables x_{ij}, the E-step is calculated by replacing x_{ij}, by its conditional expectation given y_{ij}, using $\widehat{\Psi}^{(k)}$ for Ψ. That is, x_{ij} is replaced by

$$\hat{x}_{ij}^{(k)} = E_{\Psi(k)}\{x_{ij}|y_{ij}, \delta_{ij}\} = \frac{\hat{\gamma}_i^{(k)} \varphi\left(y_{ij}, \delta_{ij} \left| \hat{\lambda}_{i+}^{(k)} \right.\right)}{\hat{\gamma}_i^{(k)} \varphi\left(y_{ij}, \delta_{ij} \left| \hat{\lambda}_{i+}^{(k)} \right.\right) + \left(1-\hat{\gamma}_i^{(k)}\right) \varphi\left(y_{ij}, \delta_{ij} \left| \hat{\lambda}_{i-}^{(k)} \right.\right)}$$
$$i = T, C$$

which is the estimate of the posterior probability of the observation y_{ij}, δ_{ij} with molecular targets after the kth iteration . The M-step requires the computation of $\hat{\gamma}_i^{(k+1)}$, $\hat{\lambda}_{i+}^{(k+1)}$, and $\hat{\lambda}_{i-}^{(k+1)}$; $i = T,\ C$, by maximizing $\log L_c(\Psi)$. Because $\log L_c(\Psi)$ is linear in the x_{ij}, it follows that x_{ij} are replaced by their conditional expectations $\hat{x}_{ij}^{(k)}$. On the $(k+1)$th iteration, the intent is to choose the value of Ψ, say $\widehat{\Psi}^{(k+1)}$, that maximizes $Q\left(\Psi; \widehat{\Psi}^{(k)}\right)$. It follows that on the M-step of the $(k+1)$st iteration, the current fit for the PPV of test drug group and control group is given by

$$\hat{\gamma}_i^{(k+1)} = \frac{\sum_{j=1}^{n_i} \hat{x}_{ij}^{(k)}}{n_i}, i = T, C.$$

Under the assumption of $n_T = n_C$, it follows that the overall PPV is estimated by

$$\hat{\gamma}^{(k+1)} = \left(\hat{\gamma}_T^{(k+1)} + \hat{\gamma}_C^{(k+1)}\right) \Big/ 2.$$

The hazard rates of the molecularly target test drug and control can then be estimated respectively as

$$\hat{\lambda}_{T+}^{(k+1)} = \frac{\sum_{j=1}^{n_T} \hat{x}_{Tj}^{(k)} \delta_{Tj}}{\sum_{j=1}^{n_T} \hat{x}_{Tj}^{(k)} y_{Tj}},\ \hat{\lambda}_{T-}^{(k+1)} = \frac{\sum_{j=1}^{n_T} \left(1-\hat{x}_{Tj}^{(k)}\right) \delta_{Tj}}{\sum_{j=1}^{n_T} \left(1-\hat{x}_{Tj}^{(k)}\right) y_{Tj}}$$

$$\hat{\lambda}_{C+}^{(k+1)} = \frac{\sum_{j=1}^{n_C} \hat{x}_{Cj}^{(k)} \delta_{Cj}}{\sum_{j=1}^{n_C} \hat{x}_{Cj}^{(k)} y_{Cj}}, \text{ and } \hat{\lambda}_{C-}^{(k+1)} = \frac{\sum_{j=1}^{n_C} \left(1-\hat{x}_{Cj}^{(k)}\right) \delta_{Cj}}{\sum_{j=1}^{n_C} \left(1-\hat{x}_{Cj}^{(k)}\right) y_{Cj}}$$

Therefore, the estimator for the treatment effect in the patients with the molecular target θ obtained from the EM algorithm is given as

$$\hat{\theta} = \log\left(\hat{\lambda}_{T+}/\hat{\lambda}_{C+}\right) = \log\left(\hat{\mu}_{C+}/\hat{\mu}_{T+}\right).$$

We propose to apply the parametric bootstrap method to estimate the standard error of $\hat{\theta}$.

Step 1: Choose a large bootstrap sample size, say $B = 1000$ or above. For $1 \leqslant b \leqslant B$, generate the bootstrap sample y_{obs}^b, δ_{obs}^b according to the probability model. The parameters for generating bootstrap samples y_{obs}^b and δ_{obs}^b are substituted by the estimators obtained from the EM algorithm based on the original observations of primary efficacy endpoints from the targeted clinical trial.
Step 2: The EM algorithm is applied to the bootstrap sample y_{obs}^b and δ_{obs}^b to obtain estimates $\widehat{\theta}_b^*$, $b = 1, .., B$.

Step 3: An estimator for the variance of $\hat{\theta}$ by the parametric bootstrap procedure is given as

$$S_B^2 = \frac{\sum_{b=1}^{B} \left(\hat{\theta}_b^* - \bar{\hat{\theta}}^*\right)^2}{B-1},$$

where

$$\bar{\hat{\theta}}^* = \frac{\sum_{b=1}^{B} \hat{\theta}_b^*}{B}.$$

Let $\hat{\theta}$ be the estimator for the treatment effects in the patients truly with the molecular target obtained from the EM algorithm. Let S_B^2 denote the estimator of the variance of $\hat{\theta}$ obtained by the bootstrap procedure. The null hypothesis is rejected and the efficacy of the molecular targeted test drug is different from that of the control in the patient population truly with the molecular target at the α significance level if

$$z = \left| \frac{\hat{\theta}}{\sqrt{S_B^2}} \right| \geqslant z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper percentile of a standard normal distribution.

The corresponding 100 $(1-\alpha)$% asymptotic CI for $\theta = \log(\lambda_{T+}/\lambda_{C+})$ can be constructed as $\hat{\theta} \pm z_{1-\alpha/2}\sqrt{S_B^2}$.