The Library of the Department of Statistics
North Carolina State University

A MODIFIED SCORE TEST FOR HIGHLY STRATIFIED SURVIVAL DATA IN RANDOMIZED CLINICAL TRIALS OF TUMOR

by

Yue-Mei L. Cheng

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1869T

A MODIFIED SCORE TEST FOR HIGHLY STRATIFIED SURVIVAL DATA IN RANDOMIZED CLINICAL TRIALS

by

Yue-Mei L. Cheng

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill, 1989

Approved by:

Advisor

Reader

Reader

ABSTRACT

YUE-MEI LIN CHENG. A Modified Score Test for Highly Stratified Survival Data in Randomized Clinical Trials. (Under of the direction of Dr. TIMOTHY M. MORGAN)

Many randomized clinical trials compare treatments which are intended to prolong life time following a particular disease. The log-rank test (LRT) given by Mantel (1966) is optimal in the comparison of treatment effect in randomized clinical trials assuming proportional hazards and a homogeneous population. However, the LRT is inefficient if the comparison is made among heterogeneous populations. One can control for this heterogeneity by including prognostic factors in a model or stratifying patients into smaller homogeneous group. A generalization of the LRT to heterogeneous populations is the stratified log-rank test (SLRT). However, if the number of patients in each stratum is small, such as in matched designs, the SLRT can be very inefficient. For instance, if there is no censoring and no stratum effect, the efficiency of the SLRT when the number of patients within each stratum is 2, 4, 6 or 10 is 50%, 64%, 71%, 78%, respectively when compared to the LRT. Schoenfeld and Tsiatis (1987) recently proposed a modified log-rank test (MLRT) for analyzing highly stratified data that is fully efficient when there is no strata effect but can be less efficient than the SLRT for large strata effects.

This research proposes and investigates the properties of a new stratified test for treatment effect where there are large number of strata. The proposed test is based on the score statistic of Cox's proportional hazards model with indicator strata covariates. The asymptotic distribution and efficiency are derived and compared to the LRT, the

SLRT and the MLRT. Conceptual advantages of the proposed test over the other three tests are discussed and a comparison is made with respect to the asymptotic relative efficiency (Pitman efficiency) at close alternative hypotheses. It is found that the proposed test is as efficient as the SLRT when there are very "large" strata effects and is as efficient as the LRT when there are no strata effects and when stratum size is large. Simulations are presented to compare the AREs of the proposed test with the LRT, the SLRT and the MLRT for some "moderate" strata effects. The test statistic requires the computation of a large number of parameter estimates. The computation involved in the estimation of the large number of parameters is impossible using existing software packages such as SAS PHGLM or BMDP2L and has been an obstacle in persuing the proposed test by previous researchers. A feasible algorithm and FORTRAN program is developed for the computation of the proposed test.

The distribution of the strata parameter estimates where the number of parameters increases proportional to the increasing sample sizes is derived. It is common in clinical trials to have quite large numbers of small groups of observations where the group sizes are 2 (corresponding to matched pair design), 3 or 4. A problem in estimating these parameters is that the maximum likelihood theory does not apply.

An example of a 2 to 1 matched randomized trial comparing the effect of two catheters on time to infection is presented to illustrate the use of the proposed test for testing of treatment effects.

ACKNOWLEDGEMENTS

I gratefully acknowledge the tremendous help of my advisor Dr. Timothy Morgan. He has always been available to me and has given high priority to supervising this project in his busy schedule. His precise guidance and continuous encouragement kept alive my interest in this project. His friendship also made this study such a wonderful experience.

I also wish to express my appreciation to the members of my committee, Drs. C. Ed. Davis, Gerardo Heiss, Ronald Helms and P. K. Sen, for their comments, suggestions and support. I appreciate the valuable advice of Dr. Sen, especially when he suggested an alternative method to one of my problems. Special thanks go to Dr. Ronald Helms and Dr. Paul Stewart for their continuous encouragement and friendship and for allowing me to use the computer equipments in their offices. I also extend my thanks to those with whom I worked on the SCOR project for their friendship and unreserved help.

Free computer time for this research was provided by UNC Academic Computing Service, Cornell National Super Computer Facility and the computing facilities of the research computing program of the Department of Public Health Sciences of Bowman Gray School of Medicine, Wake Forest University. Considerable help came from Mr. Lewyckyj, the supercomputing consultant at the UNC's ACS, for optimizing the program code to be more efficient for this project.

I deeply appreciate the unending love, support and encouragement from my parents, especially my mother. Without her support and sacrifice, this thesis would not have been completed. It is needless to mention how much I appreciate my husband and two sons for their support and love which makes all the efforts in this work worthwhile.

Finally I would like give my thanks to God with all my heart. I truly believe that "The fear of the Lord is the beginning of knowledge" (Proverbs 1:7). Therefore, I wish to use the knowledge gained through this thesis work to glorify God.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Summary of Symbols and Notation

$\mathfrak{R}_i$ or $\mathfrak{R}(t_{(i)})$: risk set at (ordered) time i, i.e. the set of individuals at risk at $t_{(i)} - 0$

n : total number of subjects

d or $\mathfrak{D}$: total number of uncensored (deaths)

n_i : number of individuals in $\mathfrak{R}_i$

d_i : number of deaths at time i

n_{Aij} : th number of patients in group A at risk at time i in stratum j

n_g : total number of strata

d_{s_j} : the number of deaths in stratum j, j=1, $\cdots$, n_g

n_{s_j} : the number of subjects in stratum j, j=1, $\cdots$, n_g

$n_{s_{Aj}}$: the number of subjects in group A in stratum j, j=1, $\cdots$, n_g

$\mathfrak{D}$: total number of deaths, $\mathfrak{D} = \sum_{i=1}^{n} \mathbb{E}\delta_i$

$\mathfrak{D}_i$: number of deaths may occur at and after time i, $\mathfrak{D}_i = \sum_{l=i}^{n} \mathbb{E}\delta_l$

$\mathfrak{D}_{ji}$: number of deaths may occur at and after time i in stratum j, $\mathfrak{D}_{ij} = \sum_{l=i}^{n} \mathbb{E}\delta_{lj}$

T : the survival time; C : the censoring time

Y= min(T, C) : the observed portion of survival time

δ : failure indicator, $\delta = 1$ if Y = T (subject failed), $\delta = 0$ if Y = C (subject censored)

X : treatment indicator; $\underset{\sim}{Z}$: vector of p covariables; observed data (Y, δ, X, $\underset{\sim}{Z}$)

α : treatment parameter

β_j : the stratum parameter of the j^{th} strata

K_j : the exponential form of the stratum parameter of the j^{th} strata, $K_j = e^{\beta_j}$

$f_T(t)$: density function of random variable (r.v.) T

$F_T(t)$: distribution function (d.f.) of r.v. T

$\overline{F}_T(t)$: survival function of r.v. T, $\overline{F}_T(t) = 1 - F_T(t) = \int_t^{\infty} f_T(u)\, du$

$\lambda(t)$: the hazard function, $\lambda(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} P[t<T\leq t+\delta t | T>t] = \frac{f(t)}{\overline{F}(t)}$

$\Lambda(t)$: the cumulative hazard, $\Lambda(t) = \int_0^t \lambda(u)\, du$

$\mathbb{E}K_j(r) = \mathbb{E}(\hat{K}_j |$ one of the member of stratum j died in r^{th} order)

p_A (p_B) : the probability of death for treatment A (treatment B), p = Pr(X=1)

p_j : the proportion of patients in stratum j with treatment A

LRT : log rank test

SLRT : stratified log rank test

MLRT : modified log rank test (Schoenfeld & Tsiatis, 1987)

MST : modified score test

ARE : asymptotic relative efficiency (Pitman efficiency)

Chapter I

INTRODUCTION

§ 1.1. *Introduction*

A clinical trial is defined as a prospective study in human subjects comparing the effect and value of interventions against a control in human subjects (Friedman, Furberg and DeMets, 1982). Clinical trials are the most definitive method of comparing two or more groups and determining whether interventions have the postulated effect. A randomized controlled trial is defined as "a carefully and ethically designed experiment with the aim of answering some precisely framed question" (Hill, 1951). Bulpitt (1983) modified this definition as "a carefully and ethically designed experiment which includes the provision of adequate and appropriate controls by a process of randomization, so that precisely framed questions can be answered". Randomization is widely accepted as an important element in the design of clinical trials (Byar, et al., 1976). In randomized designs, the control group is comparable to the intervention group in every way except for the intervention being studied and provides the baseline against which treatment or intervention can be assessed. Randomization procedures also act to prevent conscious or unconscious bias in the allocation of patients to treatments (Efron, 1971). The other advantage of randomization is that it provides a valid basis for randomization tests which can be used to test the null hypothesis of no treatment effect without any additional model assumptions and without knowledge of the prognostic factors as discussed by Fisher (1966), Kempthorne (1977) and Lehmann (1975). Note that the χ^2 test for 2 x 2 tables and t-test for comparing two means can be justified on the basis of

randomization alone without making further assumptions concerning the distribution of the baseline variables. In addition, randomization leads to unbiased estimates of an additive treatment effect, whether the important covariates are known or not (Cox, 1958). A randomized clinical trial (RCT) provides a sounder rationale for intervention than is obtainable by other methods of investigation.

Many RCTs compare treatments which are intended to prevent or delay death from a particular disease. If the course of the disease is very rapid and it is not important to take into account the duration before death, then a total count of the numbers of deaths and survivors on each treatment is sufficient for making treatment comparisons. However, if the disease is fatal, but death may take some considerable time, then it is important to not only look at how many patients died, but also when they died. In general, this is a problem in the analysis of data that have as a principal end point the time until an event occurs. Such events are usually referred to as failures (e.g. in industrial life testing) or deaths (e.g. in medical studies on chronic diseases, the endpoint of interest is survival time, the time from randomization to death), therefore the names, failure time data or survival data. In fact, the definition of survival analysis is broad enough to encompass the study of the duration between any two events, or even of the times of transitions among several states or conditions.

The distinguishing factor between survival analysis and other fields of statistics is censoring. When there is no censoring, common parametric and nonparametric two sample tests (t-test and Wilcoxon rank test) can be used when comparing two groups. Clearly, a censored observation contains only partial information about the value of the variable of interest.

Let $t_1, t_2, \cdots, t_n$ be the variables for survival time, which are independent and identically distributed (*i.i.d.*) with density $f(t) \geq 0$ and distribution function (d.f.) $F(t)$ for

a random sample of n subjects. The survival function $\overline{F}(t)$ is defined as:

$$\overline{F}(t) = 1 - F(t) = \int_t^\infty f(u)\, du\,, \quad t \geq 0.$$

The hazard function, variously known as the risk, the force or mortality, or the force of transition, specifies the instantaneous rate of failure at time t conditional upon survival to time t and is defined by:

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}$$

The cumulative hazard is $\Lambda(t) = \int_0^t \lambda(u)\, du$, so that integrating with respect to time and assuming $\overline{F}(0) = 1$ will lead to the following:

$$\overline{F}(t) = \exp\left\{ - \int_0^t \lambda(u)\, du \right\} = \exp\left\{ - \Lambda(t) \right\}. \tag{1.1.1}$$

Thus an estimate of the cumulative or integrated hazard function can be transformed to the survival curve and vice versa using equation (1.1.1).

Let $C_1, C_2, \cdots, C_n$ be the variables for censoring time for the n subjects and assume the C are *i.i.d.*, each with d.f. $F_C(t)$, and suppose that there is a period of observation c_i such that observation on subject i is stopped at c_i if failure has not occurred by then. Usually we will require that the censoring distribution does not depend on the treatment group and that censoring is independent of the principle response variate under investigation.

Three censoring mechanisms are commonly discussed in the literature, namely, fixed Type I censoring in which all the c_i are equal, c_i=c, a constant time under the

control of the investigator, Type II censoring where the trial is stopped when a prespecified number of failures occurred so that c is a random variable, and random Type I censoring where every subject is stopped at a fixed time c_i known or unknown time, with the c_i's being observed values from a random variable. An example of random type I censoring occurs when we have staggered entry and the trial is stopped at a predetermined time. The potential censoring time occurs randomly and is different for each person and dependent on the random entry time. Both Type I and Type II censoring arise in industrial and engineering applications and random censoring arises more than in clinical trials and medical applications. In a RCT, censoring may occur when patients are lost to follow-up, drop out(e.g. the therapy has increased side effects) or when the study is terminated.

Another classification of censoring appearing in the literature is the following: Let Y_i be the observed outcome variable, then,

right censoring: If $Y_i = \min(T_i, C_i)$

left censoring: If $Y_i = \max(T_i, C_i)$

An example of right censoring is the survival data observed in a clinical trial, where the death times of the censored subjects are only known to be greater than the censored times. Left censoring may be found in morbidity data in a typical census experiment. Both right and left censoring are special cases of *interval censoring*, where the random variable of interest falls in an interval. This research will be concerned with right censoring.

If the censoring distribution depends on the unknown parameters Θ of the distribution of survival times, the censoring mechanism is said to be informative, otherwise noninformative (Kalbfleisch and Prentice, 1980). If the censoring is

independent of the survival distribution, it is noninformative, but the converse is not true. It is assumed throughout this research that the censoring mechanisms are independent, unless otherwise specified.

It is important in survival analysis to estimate the survival curve $\overline{F}(t)$ defined as the probability of surviving from time 0 to t. For example, in cancer therapy, we are often interested in the 5-year survival probability. In the engineering literature, the survival function is called the reliability function and we are interested in estimating the reliability of a machine after a certain time. Censoring usually complicates the distribution theory for the estimators and survival methods typically rely on asymptotic methods for estimation and testing (Kalbfleisch & Prentice, 1980). As in most survival analyses, we consider survival studies where there are n subjects and the data can be represented by the form (Y, δ, X, $\underline{Z}$), where Y is the observed portion of survival time, δ is a failure indicator, X is the treatment indicator, and $\underline{Z}$ is a vector of p covariables. We define Y=min(T,C), where T is the survival time and C is the censoring time, and δ=1 if Y=T (subject failed), δ=0 if Y=C (subject censored). If a parametric model for survival time is specified up to a parameter vector Θ, the likelihood function for Θ can be constructed under the assumption that the censoring times are stochastically independent of each other and of the failure times.

Maximum likelihood estimation is a method commonly used in parametric models. For nonparametric situations, Cox's proportional hazards model and its non proportional extensions with time dependent covariates are often used in survival analysis. The proportional hazards model (Cox, 1972) is "non-parametric" (actually, "distribution free") in the sense that it involves an unspecified, arbitrary base-line hazard function. As a consequence, this model is more flexible, but requires different approaches for testing and estimation. The proportional hazards model (PHM) is specified by the hazard relationship,

$$\lambda(t \mid x, \underset{\sim}{z}) = \lambda(t) \exp(\alpha x + \underset{\sim}{\beta}\underset{\sim}{z}), \qquad (1.1.2)$$

where α is the parameter for treatment effect, $\underset{\sim}{\beta}$ is the parameter row vector for covariate effects, and $\lambda(t)$ is an arbitrary and unspecified base-line hazard function. This model says that the risk of dying in one group is a constant multiple of that in the other, or that the relative risk is constant, no matter the form of the risk function across time.

The classical method of estimating $\bar{F}(t)$ is the Kaplan-Meier estimator (or product-limit estimator). The estimate $\hat{\bar{F}}(t)$ is the direct generalization of the sample survivor function for censored data and was derived by Kaplan and Meier (1958). This estimator of $\bar{F}(t)$ is derived using products of conditional probabilities of surviving intervals, using the fact that $\bar{F}(t) = \Pi_{j=1}^{i} \bar{F}(t_j)/\bar{F}(t_{j-1})$. Let $Y_1 < Y_2 < \cdots < Y_d$ represent the ordered observed failure times. Let $\Re(t)$ denote the risk set at time t and let

n_i = number of individuals in $\Re(t_i)$

d_i = number of individuals who died at time Y_i

the product limit estimate of the survival function is:

$$\hat{\bar{F}}(t_i) = \prod_{j=1}^{i} \left(\frac{n_j - d_j}{n_j} \right)$$

The standard error of this at time t is estimated by Greenwood's formula (Greenwood, 1926) as:

$$\hat{\bar{F}}(t) \left\{ \sum_{t_j \le t} \frac{d_j}{n_j(n_j - d_j)} \right\}^{1/2}.$$

When there is no censoring, it reduces to the usual empirical distribution function (a step function with jump of size 1/n at each observation). Fig. 1.1 displays an example of the Kaplan-Meier estimate for the carcinogenesis data which came from the study of the times from insult with the carcinogen DMBA to mortality of vaginal cancer in rats (Pike, 1966). This figure is taken from Kalbfleisch & Prentice (1980).

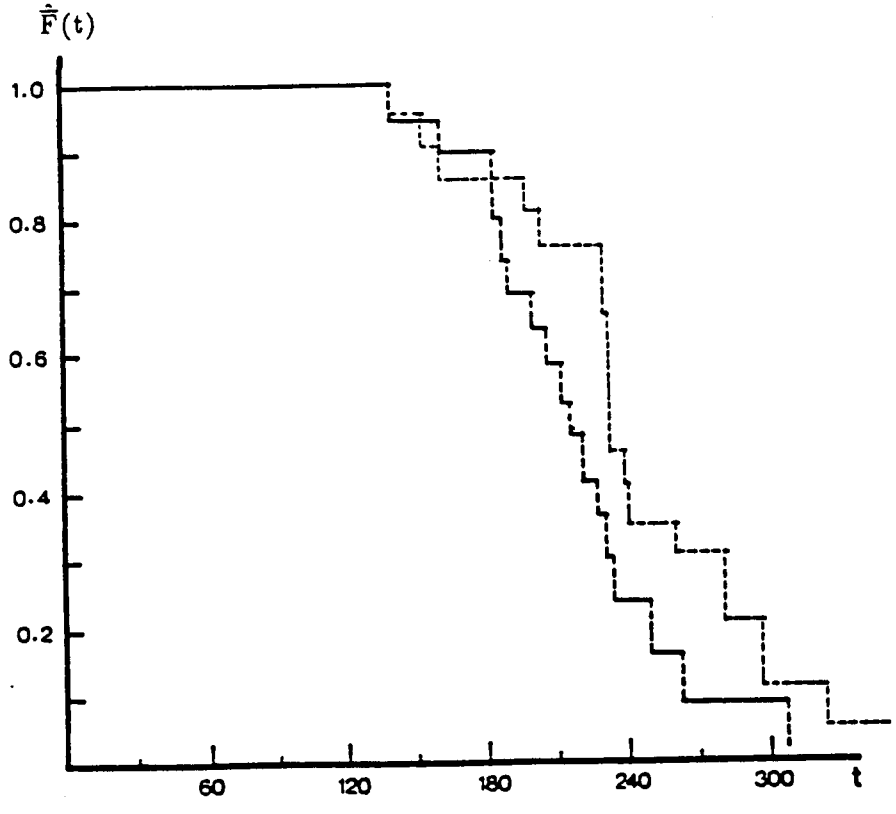

Fig. 1.1 Kaplan-Meier survivor function estimates for the carcinogenesis data: solid line, group A; broken line, group B. (from Kalbfleisch & Prentice, 1980)

This graph shows for any specified time the estimates of proportion of patients in the population (assumed homogeneous) whose survival time would exceed t, without making any assumption about the form of the survival function.

§ 1.2 *Hypothesis Testing - Homogeneous Population*

Considering the comparison of two survival curves, the problem is to devise a statistical test of the hypothesis that the two survival curves are the same. For testing, Pearson first derived a statistic for comparing the proportion of events (e.g. deaths) between two homogeneous groups which is commonly referred to as Pearson's chi-square statistic. Let p_A and p_B be the probability of death for group A and B respectively:

	Death	Alive	
Group A	a		n_A
Group B			n_B
	m_1	m_2	n

then,

$$\text{Pearson's } \chi^2 = \sum \frac{(O-E)^2}{E} = \frac{n(a-n_A m_1)^2}{n_A n_B m_1 m_2} \ ,$$

which has an asymptotic chi-square distribution under the null hypothesis of $P_A = P_B$. Mantel and Haenszel (1959) derived a χ^2 test (the M-H test) for comparing dichotomous outcomes between 2 groups across strata by combining results from a series of independent contingency tables. Let p_{Ai} and p_{Bi} be the probabilities of death for group A and B in each stratum ($i = 1, \cdots, n_g$; n_g : number of strata) respectively:

	Dead	Alive	
Group A	a_i		n_{Ai}
Group B			n_{Bi}
	m_{1i}	m_{2i}	n_i

For testing Ho : $p_{Ai} = p_{Bi}$ $i = 1, 2, \cdots, n_g$

$$\chi^2_{MH} = \frac{\left(\sum_i \left(a_i - E_o(a_i)\right)\right)^2}{\sum_i V_o(a_i)} = \frac{\left(\sum_{i=1}^{n_g}\left(a_i - \frac{n_{Ai}m_{1i}}{n_i}\right)\right)^2}{\sum_{i=1}^{n_g} \frac{n_{Ai}n_{Bi}m_{1i}m_{2i}}{n_i^2(n_i - 1)}}$$

Mantel (1966) extended these results on comparing proportions of events between groups across strata to the analysis of comparing the time to the dichotomous event (e.g. death) by forming a sequence of 2x2 tables, one at each death time. Mantel (1966) then proposed a chi-square procedure for comparing survival distributions (life table type data) of two treatment groups that was generated by applying the M-H test to this sequence of 2 x 2 tables. Following Mantel's derivation, the statistic is defined as follows: Suppose we have n individuals in 2 treatment groups, among them n_d were uncensored (i.e. death), if there are no ties (i.e. $d_i=1$), then there are n_d ordered survival times, $t_{(1)}<t_{(2)}<\cdots\cdot t_{(n_d)}$. In case of ties (i.e. $d_i \geq 1$), then there may be k ordered survival times, $t_{(1)}<t_{(2)}<\cdots\cdot t_{(k)}$, where $k \leq n_d$. For each death time (i) we have a 2 x 2 table like the one below:

	Dead	Alive	
Group A	a_i		n_{Ai}
Group B			n_{Bi}
	d_i		n_i

Here, a_i = number of deaths of group A at death time (i), d_i = number of deaths at death time (i) and n_i = number of subjects at risk at time (i), then,

$$E_o(a_i) = d_i \cdot \frac{n_{Ai}}{n_i} \quad \text{and} \quad V_o(a_i) = \frac{n_{Ai}n_{Bi}d_i(n_i-d_i)}{n_i^2(n_i-1)}$$

The statistic proposed by Mantel, subsequently called the log-rank test (LRT) is given by:

$$\text{LRT} = \frac{\left\{\sum_i \left(a_i - E_o(a_i)\right)\right\}^2}{\sum_i V_o(a_i)} = \frac{\left\{\sum_{i=1}^{k}\left(a_i - \frac{n_{Ai}d_i}{n_i}\right)\right\}^2}{\sum_{i=1}^{k}\frac{n_{Ai}n_{Bi}d_i(n_i-d_i)}{n_i^2(n_i-1)}} \qquad (1.2.1)$$

The LRT statistic for this sequence of tables is used to test the null hypothesis H_o: $F_A = F_B$, or equivalently $f_A = f_B$, $\overline{F}_A = \overline{F}_B$, $\lambda_A = \lambda_B$. Under the null hypothesis, the LRT statistic has an asymptotic chi-square distribution with one degree of freedom. Peto & Peto (1972) formalized this statistic by writing it as a rank statistic for censored data. They also showed that it was asymptotically efficient and a rank invariant test for proportional hazards (PH) alternatives and named this the log-rank test statistic. Peto (1972b) proved that it had optimal properties under the null hypothesis provided that censorship patterns were equal in all groups.

The LRT can also be derived from nonparametric rank score randomization tests. The log-rank scores are suggested by Mantel (1966), Cox (1972) and Peto and Peto (1972) for comparing the survival curves of two or more groups in the presence of right censoring. For simplicity, assuming no censoring and no ties, the ranked data : $T_{(1)}, T_{(2)}, \cdots, T_{(n)}$, have the form :

$$\zeta_{(i)} = 1 - \mathbb{E}(T_{(i)})$$

$$= 1 - \sum_{k=1}^{i} (n - k + 1)^{-1},$$

where $T_{(i)}$ denotes the i^{th} order statistic from the unit exponential distribution. Let $T = -\log(1 - R)$, where the random variable R has the uniform distribution on (0,1), then,

$$\zeta_{(i)} = 1 + \mathbb{E}(\log(1 - R_{(i)})).$$

Using the scores $\{\zeta_{(i)}\}$, the corresponding nonparametric test (log-rank test) is optimal in the sense of being the locally most powerful rank invariant (LMPR) test against Lehmann alternatives. Tests for tied and censored data are reviewed by Koch, Sen and Amara (1985).

Another derivation was given by Cox (1972) in the context of his general proportional hazards regression model. This is modeled as the product of an arbitrary baseline hazard, $\lambda(t)$, and an exponential form involving the covariates, as was shown in equation (1.1.2) and is characterized by the fact that the ratio of hazards for any two individuals does not depend on time. Assume there are d ordered survival times, $t_{(1)} < t_{(2)} < \cdots t_{(d)}$ and there are no ties ($d_i = 1$), the conditional, marginal or partial likelihoods for the proportional hazards model (PHM) with the indicator variable for treatment effect X (X=1 if treatment A, X=0 if treatment B) are proportional to:

$$L(\alpha) = \prod_{i=1}^{d} \left(\frac{e^{\alpha x_{(i)}}}{\sum_{j \in \Re(t_{(i)})} e^{\alpha x_j}} \right)$$

and the log-likelihood,

$$\mathcal{L}(\alpha) = \log L(\alpha) = \sum_{i=1}^{n} \delta_i \left(\alpha x_{(i)} - \log \left(\sum_{j \in \Re(t_{(i)})} e^{\alpha x_j} \right) \right)$$

where $\Re(t_{(i)})$ is the set of individuals at risk at $t_{(i)} - 0$. The first derivative of this log likelihood function is given by:

$$\frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = U(\alpha) = \sum_i \delta_i \left(x_{(i)} - \frac{\sum_{j \in \Re(t_{(i)})} x_j e^{\alpha x_j}}{\sum_{j \in \Re(t_{(i)})} e^{\alpha x_j}} \right)$$

and the second derivative is the sample information given by:

$$i(\alpha) = - \frac{\partial^2 \mathcal{L}(\alpha)}{\partial \alpha^2} = \sum_i \delta_i \left\{ \frac{\sum_{\Re_i} x_j^2 e^{\alpha x_j}}{\sum_{\Re_i} e^{\alpha x_j}} - \left(\frac{\sum_{\Re_i} x_j e^{\alpha x_j}}{\sum_{\Re_i} e^{\alpha x_j}} \right)^2 \right\}$$

The score test for testing H_o: $\alpha = 0$, or equivalently $F_A = F_B$, assuming PH alternatives, is given by:

$$S_{PH} = \frac{\{U(0)\}^2}{i(0)} \qquad (1.2.2)$$

The log-rank test can be rewritten to show that it is the same as the score test proposed by Cox for his proportional hazards model for the two-sample problem with censored data. The derivation is shown below:

If, under H_o: $\alpha = 0$ with $x_i = \begin{cases} 1 & \text{if } i \text{ is in group A} \\ 0 & \text{if } i \text{ is in group B} \end{cases}$, one may use

the notation for the LRT to express the score statistic (Rao, 1973, p.417), U(0), by:

$$U(0) = \frac{\partial}{\partial \alpha} \log L\,(0) = \sum_{i=1}^{n} \delta_i \Big(x_{(i)} - \frac{\sum_{\Re_i} x_j}{n_i} \Big) = \sum_{i=1}^{n_d} \Big(a_i - \frac{n_{Ai}}{n_i} \Big).$$

In the above, n_i = number of individuals in $\Re(t_{(i)})$, a_i = number of deaths of group A at death time (i) and it is assumed that there are no ties (d_i=1). The sample information is given by,

$$i(0) = \sum_{i=1}^{n} \delta_i \left(\frac{1}{n_i} \sum_{j \in \Re(t_{(i)})} x_j^2 - \Big(\frac{\sum_{\Re_i} x_j}{n_i} \Big)^2 \right) = \sum_{i=1}^{n} \delta_i \frac{n_{Ai}}{n_i} \left(1 - \frac{n_{Ai}}{n_i} \right) = \sum_{i=1}^{n_d} \frac{n_{Ai} n_{Bi}}{n_i}.$$

Similarly, in the case of no ties (d_i=1), equation (1.2.1) can be written as:

$$LRT = \frac{\left(\sum_{i=1}^{d} \Big(a_i - \frac{n_{Ai}}{n_i} \Big) \right)^2}{\sum_{i=1}^{d} \frac{n_{Ai} n_{Bi}}{n_i}} = \frac{\left(\sum_{i=1}^{n} \delta_i \Big(x_i - \frac{n_{Ai}}{n_i} \Big) \right)^2}{\sum_{i=1}^{n} \delta_i \frac{n_{Ai} n_{Bi}}{n_i}} \qquad (1.2.3)$$

Clearly, from equation (1.2.2) and (1.2.3),

$$S_{PH} = \frac{\left(\sum_i \delta_i \Big(x_i - \frac{n_{Ai}}{n_i} \Big) \right)^2}{\sum_i \delta_i \frac{n_{Ai} n_{Bi}}{n_i}} = LRT \qquad \underset{\sim}{H_o} \quad \chi^2_{(1)}$$

Thus the score statistic, $[U(0)]^2/i(0)$ coincides with the log-rank test in the case where

there are no ties.

In this paper, emphasis is on the log-rank test, its weighted version as proposed by Taron & Ware (1977), and modified versions of the log-rank test. In the 2 x 2 table of each uncensored observation, Taron and Ware suggest weighting the statistic as:

$$\left(\sum_{i=1}^{d} w_i \left(a_i - E_O(a_i) \right) \right)^2$$

for the numerator and

$$\sum_{i=1}^{d} w_i^2 \, V_O(a_i)$$

for the denominator, where, as in the log-rank test,

$$E_O(a_i) = d_i \cdot \frac{n_{Ai}}{n_i} \quad \text{and} \quad V_O(a_i) = \frac{n_{Ai} n_{Bi} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} .$$

The choices for the weight functions give rise to different test statistics. Examples are listed below (SAS PROC SURDIFF), where $\hat{\bar{F}}$ is the Kaplan-Meier (1958) estimator of the survival function based on the combined sample:

Weight function	Test statistic
n_i	Gehan-Wilcoxon (Gehan 1965)
$[\hat{\bar{F}}(t)]^{\rho}[1-\hat{\bar{F}}(t)]^{\gamma}$	$G^{\rho,\gamma}$ class
1	log-rank
$\hat{\bar{F}}(x)$	Peto-Peto-Wilcoxon (Peto & Peto 1972)
$\{\hat{\bar{F}}(x)\}^{\rho}$	G^{ρ} (Harrington and Fleming 1982)

Since n_i and $\hat{\bar{F}}(x)$ are decreasing weight functions, the Gehan-Wilcoxon and Peto-Peto-Wilcoxon tests are more sensitive to earlier survival differences than the log-rank test. The $G^{\rho,\gamma}$ test has the flexibility of being able to place weight at small values of time to detect earlier differences by taking $\rho>0$, γ close to 0, or at large value of time to detect late differences by taking ρ close to 0, $\gamma>0$, or to middle differences by taking ρ and γ nearly equal. The $G^{\rho,\gamma}$ class places the most weight at times where $\bar{F}(t) = \rho/(\rho+\tau)$. The G^{ρ} class of test procedures is the $G^{\rho,\gamma}$ family with $\gamma=0$, and includes the log-rank ($\rho=0$) and Peto-Peto-Wilcoxon ($\rho=1$) tests as special cases. The log-rank test turns out to be the optimal test within these tests when the proportional hazards model holds.

§ 1.3 *Stratification*

Section 1.1 discussed the case of homogeneous populations. If instead, we have heterogeneous populations, for example, where groups of subjects may differ by race, age, blood pressure, performance status, family history etc. We can control for this heterogeneity by including prognostic factors into a model or stratifying patients into

smaller groups of homogeneous populations. Even when reasonable balance has been achieved by randomization it may be desirable to perform a treatment comparison adjusted for prognostic factors in an attempt to achieve greater precision and power. Many randomized clinical trails are designed using methods such as permuted block randomization within strata to ensure that important prognostic factors are equally distributed among the different treatments. The importance of prognostic stratification in reducing the chance of spuriously significant results in treatment comparisons was demonstrated by Feinstein (1972) and Feinstein & Landis (1976).

§ 1.4 *Proposed Research*

The focus of this dissertation is to develop a new statistic for comparing survival distributions when there are a large number of strata. The new test performs as well as the stratified log-rank test (SLRT) and better than either the LRT or the MLRT when the stratum effect is large. When there is no stratum effect, the proposed test is more efficient than the SLRT and is fully efficient if the stratum sizes are large.

A review of the stratification literature will be presented in Chapter 2. Section 2.1 considers the asymptotic properties of test procedures (the LRT and the SLRT) that allow for a fixed number of strata. Section 2.2 investigates the asymptotic properties of the LRT, the SLRT and the MLRT when there is a fixed number of subjects within each stratum and the number of strata is proportional to the sample size as in the case of matched pair designs.

The asymptotic variance and efficiency of the recently proposed MLRT (Shoenfeld & Tsiatis, 1987) is derived in chapter 3. The asymptotic relative efficiency (ARE) of the LRT, the SLRT and the MLRT are evaluated.

Section 4.1 proposes a new stratified test, the modified score test (MST). The properties of the test when there are large strata effects are given in Section 4.2. In section 4.3 the properties of the test conditional upon estimated strata effects are studied. The unconditional distribution of the proposed test is given in Section 4.4 when there are no strata effects. The estimate of the variance of the test statistic is proposed in Section 4.5. Simulation results for the comparisons of the AREs of the LRT, the SLRT, the MLRT and the MST are given in Section 4.6. The distribution of strata effects for large stratum sizes is derived in Section 4.7. To derive the distribution of strata effects estimates for small stratum sizes, the expected value of strata effects conditioned on the order of one of the members death time is necessary and is given in section 4.8. Based on this, the distribution of the estimated stratum effect is given in section 4.9. Section 4.10 summarizes the results.

The uncensored results derived in Chapter 4 are extended in Chapter 5 to allow for type I censoring. Simulation results for various proportion of Type I censoring are given in Section 5.3.

A computer algorithm developed to compute the proposed test statistic is given in chapter 6. An example of an application of the method is given for comparing times to infection between two types of catheters in a two to one matched study.

Finally, discussions of the plan of future research and possible extensions are given in Chapter 7.

Chapter II

LITERATURE REVIEW

§ 2.1 *Stratification (Fixed number of stratum)*

2.1.1. *Introduction*

Two kinds of stratification appear in the literature, namely, *pre-stratification* which occurs prior to balanced randomization and *post-stratification* which occurs after patients are entered into study and the randomization did not take strata into account. Pocock and Simon (1975) emphasized the use of pre-stratification to balance the effect of stratification characteristics across treatments. Peto et al (1976) advocates post-stratification to adjust for the effect of prognostic factors. Asymptotically, pre-stratification and post-stratification are equivalent (ARE=1). Grizzle (1982) examined the efficiency of stratifying on a risk factor before randomization as opposed to complete randomization and adjustment for the risk-factor in the analysis stage. He showed that, in small samples, relying on post-stratification can lead to significant loss of efficiency compared to pre-stratification. Nevertheless, in moderate and large samples the two design strategies should have similar efficiency. Meier (1981) showed that if 2n is the number of subjects in a given stratum, and for simplicity, assuming that event rates are near 0.5, the expected relative efficiency of the random allocation compared to stratified allocation is $1 - \frac{1}{4n}$. Even for n as small as 10, the expected relative efficiency is nearly 100%. Meier (1981) and Mantel (1984) also discussed the issue of pre-stratification as an aspect of design opposed to poststratification as a part of the analysis. Armitage and Gehan (1974) discussed statistical methods for the identification and use of

prognostic factors. In this research it is assumed subjects were randomized using pre-stratification.

2.1.2. *Stratified log-rank test (SLRT)*

The extension of Mantel's (1966) log-rank test for homogeneous populations to heterogeneous study groups is the stratified log-rank test. Following the notation given for the LRT with subscript j as the stratum indicator (j=1, $\cdots$, n_g) and n_{s_j} as the size of stratum j, the SLRT is given by:

$$SLRT = \frac{\left(\sum_{j=1}^{n_g} \sum_{i=1}^{n_{s_j}} \delta_{ij} \left(x_{ij} - E_o(x_{ij}) \right) \right)^2}{\sum_{j=1}^{n_g} \sum_{i} \delta_{ij} V_o(x_{ij})}$$

$$= \frac{\left(\sum_{j} \sum_{i} \left(a_{ij} - \frac{n_{Aij}}{n_{ij}} \right) \right)^2}{\sum_{j} \sum_{i} \frac{n_{Aij} n_{Bij}}{n_{ij}}}. \qquad (2.1.1)$$

SLRT has an asymptotic χ^2 distribution under the hypothesis $F_{Aj} = F_{Bj}$, where F_j (j = 1, 2,$\cdots$, n_g) refers to the distribution function of each stratum. Assuming no ties, the SLRT can be shown to be equivalent to the score statistic under Cox's stratified proportional hazards model. Define the hazard function of the j^{th} stratum (j = 1, 2,$\cdots$, n_g) as:

$$\lambda_j(t \mid x, \underset{\sim}{z}) = \lambda_j(t) \exp(\alpha x + \underset{\sim}{\beta z}) \qquad (2.1.2)$$

The (conditional, marginal or partial) likelihood for the stratified PHM with one indicator variable for treatment(0,1) is given by:

$$L = \prod_j L_j$$

where,

$$L_j = \prod_i \left(\frac{e^{\alpha x_{ij}}}{\sum_{\Re_j(t_{(i)})} e^{\alpha x_{mj}}} \right),$$

and,

$$\log L_j = \sum_{i=1}^{n} \delta_{ij} \left(\alpha x_{ij} - \log \left(\sum_{\Re_j(t_{(i)})} e^{\alpha x_{mj}} \right) \right)$$

Taking the first and second derivative of the log likelihood equation and evaluating at $\alpha = 0$, gives:

$$U_j(0) = \frac{\partial}{\partial \alpha} \log L_j(0) = \sum_{i=1}^{n_{s_j}} \delta_{ij} \left(x_{ij} - \frac{\sum_{l \in \Re_j} x_{lj}}{n_{ij}} \right) = \sum_i \left(a_{ij} - \frac{n_{Aij}}{n_{ij}} \right),$$

in which n_{ij} = number of individuals in $\Re_j(t_{(i)})$, and

$$i_j(0) = \sum_i \delta_{ij} \left(\frac{1}{n_{ij}} \sum_{l \in \Re_j(t_{(i)})} x_{lj}^2 - \left(\frac{\sum_{\Re_j} x_{lj}}{n_{ij}} \right)^2 \right)$$

$$= \sum_i \frac{n_{Aij}}{n_{ij}} \left(1 - \frac{n_{Aij}}{n_{ij}} \right) = \sum_i \frac{n_{Aij} n_{Bij}}{n_{ij}}.$$

The score statistic for the test of $\alpha = 0$, or equivalently, $f_{Aj} = f_{Bj}$, $F_{Aj} = F_{Bj}$ or $\overline{F}_{Aj} = \overline{F}_{Bj}$ ($j = 1, \cdots, n_g$) is given by:

$$\sum_j U_j(0) = \sum_j \sum_i \left(a_{ij} - \frac{n_{Aij}}{n_{ij}} \right) .$$

The score test for a stratified PHM is given by:

$$S_{SPH} = \frac{\left(\sum_j U_j(0) \right)^2}{\sum_j i_j(0)} = \frac{\left(\sum_j \sum_i \left(a_{ij} - \frac{n_{Aij}}{n_{ij}} \right) \right)^2}{\sum_j \sum_i \frac{n_{Aij} n_{Bij}}{n_{ij}}}, \qquad (2.1.3)$$

from equation (2.1.1) and (2.1.3), it is clearly that S_{SPH} coincides with the SLRT. Extension can be made to the case in which there are more than two treatment groups.

2.1.3. *Efficiency of the LRT vs SLRT*

The LRT is fully efficient under a proportional hazards assumption when there is no stratum effect. In testing for no treatment effect, the use of the log-rank test is conservative when treatments were assigned by a stratified design (Green & Byar, 1978). The asymptotic relative efficiency (ARE) of the log-rank test vs the SLRT is 1 when there are no strata effects and a constant treatment effect. The ARE of SLRT vs. LRT when there are strata effects and a constant treatment effect can be derived using the results of Lagakos and Schoenfeld (1984) and Morgan (1986).

If the PHM with (indicator) covariates is the correct model with the hazard specified by $\lambda_j(t|\ x, \underset{\sim}{z}) = \lambda(t)\ e^{\alpha x + \underset{\sim}{\beta z}}$ and α is of order $n^{-1/2}$ as $n \to \infty$, the score statistic is given by:

$$S_{PH} = \sum_j \left(x_j - \frac{\sum_{\Re_j} x_m e^{\underset{\sim}{\beta z}}}{\sum_{\Re_j} e^{\underset{\sim}{\beta z}}} \right) .$$

S_{PH} has asymptotic mean, $\alpha \int P(1 - P)\, dF_y(t)$, and variance, $\int P(1 - P)\, dF_y(t)$, where $P = Pr(X=1)$, $F_y(t)$ is the probability of observing a death by time t, i.e. $F_y(t) = F_t(t)F_c(t)$. Let S_j be the score statistic for stratum j. The test statistic for a stratified PHM, assuming the λ_j's are proportional to each other with $\lambda_j(t \mid x) = \lambda_j(t)\, e^{\alpha x}$ is

$$S_{SPH} = \sum_{j=1}^{n_g} S_j.$$

For a PHM, the asymptotic mean of S_j becomes:

$$\alpha\ P(1 - P) \sum_{j=1}^{n_g} \int d\ F_{y_j}(t) \quad \to \quad \alpha\ P(1 - P) \int d\ F_y(t),$$

and the asymptotic variance:

$$P(1 - P) \sum_{j=1}^{n_g} \int d\ F_{y_j}(t) \quad \to \quad P(1 - P) \int d\ F_y(t).$$

The ARE of SLRT(or S_{SPH}) vs. S_{PH} is 1.

Hence, for a stratified PHM, if we further assume that the hazards of the stratum are proportional, then this is essentially the PH model with indicator covariates. Several results in the literature concerning omission of covariates from the PHM can then be applied. In simple randomized trials, Lagakos and Schoenfeld (1984) and Morgan (1986) found that the omission of an important covariate from the Cox PH model still gave tests of "no treatment effect" with nominal size. However, there is substantial efficiency loss. With stratified randomization, Gail (1988) showed that the omission of stratum effects from Cox's PH model led to a test with subnominal size, while in simple randomization the test from a PH model with omitted stratum effect achieved nominal size. When a covariate is omitted from the PHM, the hazards for the two treatment groups are no longer proportional. The derivation for an unstratified model obtained by omitting n_g-1 indicator variables for the n_g strata, is shown below. When the n_g-1 binary covariates are omitted, the hazard function is:

$$\lambda(t \mid x) = \frac{f(t \mid x)}{\overline{F}(t \mid x)} = \frac{\int f(t|\, x, \underset{\sim}{z})\, d\, F_{\underset{\sim}{z}}(\underset{\sim}{z})}{\int \overline{F}(t|\, x, \underset{\sim}{z})\, d\, F_{\underset{\sim}{z}}(\underset{\sim}{z})} = \frac{\int \lambda(t|\, x, \underset{\sim}{z})\, \overline{F}(t|\, x, \underset{\sim}{z})\, d\, F_{\underset{\sim}{z}}(\underset{\sim}{z})}{\int \overline{F}(t, x, \underset{\sim}{z})\, d\, F_{\underset{\sim}{z}}(\underset{\sim}{z})},$$

where α is the parameter for treatment effect, $\underset{\sim}{\beta}$ is the parameter vector for the n_g-1 binary covariates, with $\lambda(t|x, \underset{\sim}{z}) = \lambda(t)\, e^{\alpha x + \underset{\sim}{\beta}\underset{\sim}{z}}$ and $\overline{F}(t,x,\underset{\sim}{z}) = \exp\{-\Lambda(t) e^{\alpha x + \underset{\sim}{\beta}\underset{\sim}{z}}\}$, the hazard function may then be written as:

$$\lambda(t|x) = \lambda(t) e^{\alpha x} A(x,t)$$

where $$A(x,t) = \frac{\int e^{\underset{\sim}{\beta}\underset{\sim}{z}} e^{-\Lambda(t)\, e^{\alpha x + \underset{\sim}{\beta}\underset{\sim}{z}}}\, dF_{\underset{\sim}{z}}(\underset{\sim}{z})}{\int e^{-\Lambda(t)\, e^{\alpha x + \underset{\sim}{\beta}\underset{\sim}{z}}}\, dF_{\underset{\sim}{z}}(\underset{\sim}{z})}.$$

The hazard ratio is,

$$\frac{\lambda(t|x=1)}{\lambda(t|x=0)} = \phi(t)e^{\alpha x} , \tag{2.1.2}$$

where,

$$\phi(t) = \frac{\int e^{\underset{\sim}{\beta z}} B_1 dF_z(\underset{\sim}{z}) \int B_2 dF_z(\underset{\sim}{z})}{\int e^{\underset{\sim}{\beta z}} B_2 dF_z(\underset{\sim}{z}) \int B_1 dF_z(\underset{\sim}{z})}$$

with

$$B_1 = e^{-\Lambda(t)e^{\alpha x+\underset{\sim}{\beta z}}} \quad \text{and} \quad B_2 = e^{-\Lambda(t)e^{\underset{\sim}{\beta z}}} .$$

For non-PHM (i.e. $\lambda(t|x,\underset{\sim}{z}) = \lambda(t)e^{\alpha g(t)x+\underset{\sim}{\beta z}}$),

$$\frac{\lambda(t|x=1)}{\lambda(t|x=0)} = e^{\alpha g(t)} . \tag{2.1.3}$$

From (2.1.2) and (2.1.3),

$$\phi(t)e^{\alpha x} = e^{\alpha g(t)} ,$$

or equivalently,

$$g(t) = 1 + \frac{\log\phi(t)}{\alpha},$$

as $\alpha \rightarrow 0$, $\phi(t) \equiv 1$, $\log\phi(t) = 0$.

Using L'Hospital's rule,

$$\lim_{\alpha \rightarrow 0} \frac{\log\phi(t)}{\alpha} = \lim_{\alpha \rightarrow 0} \frac{1}{\phi(t)} \frac{\partial\phi(t)}{\partial\alpha}$$

hence,

$$g(t) \rightarrow 1 + \lim_{\alpha \rightarrow 0} \frac{\partial\phi(t)}{\partial\alpha} ,$$

where g(t) (Morgan, 1986) is

$$1 + \frac{\Lambda(t)\ B(t, \beta z)}{B(t, 0)} - \frac{\Lambda(t)\ B(t, 2\beta z)}{B(t, \beta z)} . \qquad (2.1.4)$$

In the above, $\Lambda(t) = \int \lambda(u)\, du$ and

$$B(t, u) = \int e^{u\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta z}}}}\, d\, F_{\underset{\sim}{z}}(\underset{\sim}{z})$$

The asymptotic relative efficiency of omitting covariates from the PHM relative to including the covariates is given by:

$$ARE\left(\frac{LRT}{S_{PH}}\right) = \frac{\left\{\int g(t)\, d\, F_y(t)\right\}^2}{\left\{\int d\, F_y(t)\right\}^2}$$

Therefore, the ARE of LRT vs SLRT becomes:

$$ARE\left(\frac{LRT}{SLRT}\right) = \frac{\left\{\int g(t)\, d\, F_y(t)\right\}^2}{\left\{\int d\, F_y(t)\right\}^2} .$$

Asymptotic distribution theory for the log-rank test assuming proportional hazards with and without adjustment for covariates is discussed in Schoenfeld (1981, 1983). When important covariates are omitted from the proportional hazards model the estimate of treatment effect is unbiased only under the hypothesis of no treatment effect. Based on simple randomization, Gail, Wieand and Piantadosi gave the exact asymptotic bias in the estimate of treatment effect using Cox's model when needed covariates are omitted. Lagakos and Schoenfeld (1984) investigated the effects of misspecifying a PH regression model on the associated partial-likelihood score test for comparing two randomized treatments in the presence of covariates. Based on this, Morgan (1986) derived the asymptotic relative efficiency of the PH model when all covariates are omitted from the model. All of these results on omitting covariates are analogous to not taking stratum effect into account in the analysis of treatment results. If there are no stratum effects, then $g(t) = 1$ and the ARE of LRT vs SLRT will be equal to 1. If there is no censoring, then $\int dF_y(t) = 1$ and the equations can be simplified.

In summary, the LRT is the locally most powerful rank statistic for testing of treatment effect and is fully efficient if there are no strata effects. However, as the strata effect gets large, the mean of the LRT is not consistent and becomes more conservative and inefficient. The SLRT is efficient when the number of subjects within stratum is large, however, as the number of subjects within stratum gets small, the

SLRT becomes very inefficient as will be detailed in section 2.2.

§ 2.2 *Stratification (Fixed Number within stratum)*

Comparison of two treatments based on censored paired survival data is becoming more popular. Studies such as matched pair experiments or twin studies allow better control over the extraneous factors when examining the relationship between outcome measurements and treatments. Some twin studies have been used in the medical area to investigate the relation between suspected risk factors and the mortality or morbidity experience of human populations (Cederlof *et al.* 1971; W.H.O., 1966). Hammond (1964) used a matched pair analysis to study smoking in relation to mortality and morbidity in the United States.

Sen (1968) showed that the asymptotically efficient test based on the method of n rankings is the rank sum test for the completely random design and is Friedman's test for the randomized complete block design. However, uniformly efficient test statistics are not available for the stratified design in which the within stratum sizes are fixed and the number of stratum, n_g, are increasing. The SLRT proposed by Mantel (1966) for a fixed number of stratum can be used in this situation but as will be shown in section 2.2.2, it is very inefficient when the number of subjects within each stratum is small.

Mantel and Ciminera (1979) proposed a test for matched pair data based on the use of log-rank scores for use in the analysis of litter-matched data on time to tumor appearance and computed the mean and variance of the sum of the log-rank scores by using the stratified permutation distribution. In their study, each litter provided a paired cohort of animals where the animals receiving treatment A are compared with their littermates receiving treatment B. This method of assigning scores is analogous to combining stratum before assigning ranks rather than assigning ranks within stratum in

comparing two treatment groups.

Wei (1980) proposed a censored paired-data test by modifying the variance of a test for paired observations (Gehan, 1965; Gilbert, 1962). He performed a Monte Carlo power study that allowed the censoring distributions to vary within strata and showed that his asymptotically distribution-free test is more powerful than the sign test under a Block-Basu (1974) bivariate exponential model. However, the Monte Carlo simulation results given by Cheng (1984) indicated that some tests are more powerful than Wei's test. O'Brien and Fleming (1987) also showed that Prentice-Wilcoxon scores are generally better than Gehan scores upon which Wei's test was based.

Woolson and Lachenbruch (1980) proposed a generalized signed rank test procedure for the matched pairs testing problem with randomly right censored observations. They also presented censored data extensions of a class of rank tests based on the intrablock rankings (e.g. logistic scoring, extreme-value scoring) for randomized block designs (Woolson and Lachenbruch, 1981). Some comparisons were made among studies of two sets of matched data using three types of analysis (Woolson and Lachenbruch, 1982), namely, analyses which ignore the pairing (i.e. the LRT), analyses which take the pairing into account by covariate adjustment (i.e. a parametric Weibull PHM with pairing covariates) and analyses designed for paired data (i.e. signed rank tests). They suggested the use of the logistic scores incorporated into the generalized signed rank test. O'Brien and Fleming (1987) proposed a paired Prentice-Wilcoxon statistic for censored paired data. In their simulation studies, they compared Prentice-Wilcoxon, Woolson-Lachenbruch generalized signed rank and sign tests and found that the Prentice-Wilcoxon test is most powerful against all but the exponential scale alternative (where the Woolson-Lachenbruch generalized signed rank is more powerful). However, if outlier pairs are introduced, the power of the Woolson-Lachenbruch generalized signed rank is decreased.

In the case of no censoring, survival data can be thought of as multivariate outcomes or longitudinal data with repeated measures. Liang & Zeger (1986) have proposed an extension of generalized linear models to the analysis of longitudinal data. Zeger & Liang (1986) further proposed an approach to the analysis of repeated measures longitudinal data for discrete and continuous outcomes.

One area of interest is the application of the PH model to the regression analysis of survival times when there are many strata (e.g. matched pairs). Holt and Prentice (1974) used the PH model along with exponential and Weibull specializations, and assume that the covariates act multiplicatively on the hazard function while the underlying survival distribution may vary from pair to pair. Cuzick and Clayton (1985) proposed a multivariate generalization of the PH model in an attempt to model dependent survival times and apply these methods to censored matched pairs. Gill (1985) noted that there are computational difficulties in applying Cuzick and Clayton's methods. Hougaard (1986) proposed a class of continuous multivariate lifetime distributions. He suggested using an iterative method to find the estimate of α in Cox's model without the assumption of parametric hazard, and yet he reported that the statistical properties of the procedure are still unknown.

2.2.1. *The LRT - Log-Rank Test*

The associated hazard function is given by: $\lambda(t \mid x) = \lambda(t)\ e^{\alpha x}$, where x is the treatment indicator. Under the proportional hazards model, we have shown that the LRT is equivalent to the score test for the hypothesis of no treatment differences (i.e. $\alpha = 0$). The numerator of the LRT can be written as:

$$LRT_{NUM} = \sum_i \delta_i \Big(x_i - \frac{\sum_{j \in \Re_i} x_j}{\sum_{j \in \Re_i} 1} \Big) = \sum_{i=1}^{n} \delta_i \Big(x_i - \frac{n_{Ai}}{n_i} \Big),$$

and the denominator as:

$$LRT_{DEN} = \sum_i \delta_i \Big(\frac{\sum_{i \in \Re_i} x_i}{\sum_{i \in \Re_i} 1} \Big(1 - \frac{\sum_{i \in \Re_i} x_i}{\sum_{i \in \Re_i} 1} \Big) \Big) = \sum_{i=1}^{n} \delta_i \frac{n_{Ai} n_{Bi}}{n_i}.$$

Let $\alpha = \log\{ \lambda_A(t) / \lambda_B(t) \}$. (2.2.1)

Assume α is $O(n^{-1/2})$; then the asymptotic distribution of the LRT_{NUM} is given by Shoenfeld (1981) as:

Asy. mean: $\alpha \, P(1-P) \int d \, F_y(t)$

Asy. Variance: $P(1-P) \int d \, F_y(t)$

where $F_y(t)$ is the probability of observing a death by time t.

However, if the proportional hazards model with stratum covariates is correct and α is of order $n^{-1/2}$, the asymptotic mean and variance of LRT_{NUM} if stratum effect exists are given by (Lagakos & Schoenfeld, 1984 and Morgan, 1986) :

Asy. mean: $\alpha \, P(1-P) \int g(t) \, d \, F_y(t) = \alpha^* P(1-P) \int d \, F_y(t)$ (2.2.2)

Asy. Variance: $P(1-P) \int d \, F_y(t)$

where g(t) is defined in equation (2.1.4).

Under the null hypothesis, $\alpha = 0$, this statistic will have mean zero. If there are no stratum effects, then g(t) = 1 and the statistic will be unbiased and fully efficient. If there is no censoring, then $\int dF_Y(t) = 1$ and the equations can be simplified. In summary, the LRT is the locally most powerful rank statistic for testing no treatment effect and is fully efficient if there are no strata effects. However, as the stratum effect gets large, the mean is not consistent (with $|\alpha^*| < |\alpha|$), and becomes more conservative and inefficient.

2.2.2. *The SLRT - Uncensored Case*

For stratification with a fixed number of patients within stratum, but allowing the number of stratum to increase as n increases, the mean and variance of the SLRT can be derived using the hypergeometric distribution. Under the proportional hazards model, assume $\alpha = \log\{ \lambda_A(t)/ \lambda_B(t)\}$ is $O(n^{-1/2})$, let $x_{ij} = 1$ if the death at time i in stratum j occurred in treatment group A, $x_{ij} = 0$ otherwise. Let n_{ij} be the number of patients at risk at time i in stratum j, n_{Aij}, the number of patients in group A at risk at time i in stratum j, n_{s_j}, the number of patients in stratum j and $n_{s_{Aj}}$, the number of patients in group A in stratum j. Conditional on n_{Aij} and n_{ij}, the $\{x_{ij}\}$ can be treated as if they were a sequence of independent Bernoulli random variables with means:

$$E(x_{ij}|\, n_{Aij}\,,\, n_{ij}) = \mu_{ij} = \frac{n_{Aij}\, e^{\alpha}}{n_{Aij}\, e^{\alpha} + n_{ij} - n_{Aij}}\,.$$

Expanding in a Taylor series about zero we get,

$$\mu_{ij} = \frac{n_{Aij}}{n_{ij}} + \alpha \frac{n_{Aij}}{n_{ij}} \left(1 - \frac{n_{Aij}}{n_{ij}} \right) = e_{ij} + \alpha\, e_{ij} (1 - e_{ij}),$$

where $e_{ij} = \frac{n_{Aij}}{n_{ij}}$. The numerator of the SLRT can be written as:

$$\sum (x_{ij} - e_{ij}) = \sum (x_{ij} - \mu_{ij}) + \alpha \sum e_{ij} (1 - e_{ij})$$

The expected value of the numerator of the SLRT converges to:

$$E(SLRT_{NUM}) = \alpha\, E \sum_{j}^{n_g} \sum_{i}^{n_{s_j}} e_{ij} (1 - e_{ij})$$

and the denominator :

$$SLRT_{DEN} \rightarrow E(\sum_{j}^{n_g} \sum_{i}^{n_{s_j}} e_{ij} (1 - e_{ij})) .$$

It is clear that $n_{Aij} \sim$ Hypergeometric (n_{s_j}, $n_{s_{Aj}}$, n_{ij}). Let $n_{s_{Aj}} = p_j n_{s_j}$ and $n_{s_{Bj}} = (1-p_j) n_{s_j}$, where $p_j = Pr(x_{ij}=1)$, $j=1, \cdots, n_g$. For simplicity, we will assume balanced treatment allocation, (i.e. $p_j = p$, $\forall$ j), and

$$E(n_{Aij}) = p\, n_{ij} , \qquad E(e_{ij}) = p$$

and

$$Var(n_{Aij}) = n_{ij}\, p(1 - p)\left(1 - \frac{n_{ij} - 1}{n_{s_j} - 1} \right) = n_{ij}\, p(1 - p)\left(\frac{n_{s_j} - n_{ij}}{n_{s_j} - 1} \right)$$

$$E(n_{Aij}^2) = n_{ij}\, p(1 - p)\left(\frac{n_{s_j} - n_{ij}}{n_{s_j} - 1} \right) + n_{ij}^2\, p(1 - p)$$

then,

$$\mathbb{E}(\, e_{ij}^2 \,) = \frac{p(1-p)}{n_{ij}}\left(\frac{n_{s_j} - n_{ij}}{n_{s_j} - 1}\right) + p(1-p)$$

$$\mathbb{E}(\, e_{ij}(1 - e_{ij})) = p(1-p)\left(1 - \frac{1}{n_{ij}}\left(\frac{n_{s_j} - n_{ij}}{n_{s_j} - 1}\right)\right.$$

$$\mathbb{E}\sum(\, e_{ij}(1 - e_{ij})) = p(1-p)\sum_{j=1}^{n_g}\sum_{i=1}^{n_{s_j}}\left(1 - \frac{1}{n_{ij}}\left(\frac{n_{s_j} - n_{ij}}{n_{s_j} - 1}\right)\right).$$

Assume $\alpha = \log\{\, h_A(t)/\, h_B(t)\}$ is $O(\, n^{-1/2})$ and $n_{Aj} = pn_{s_j}$, the SLRT is asymptotically normally distributed with mean and variance given by :

$$\text{Asy. mean:} \qquad \alpha\, p(1-p)\sum_{j}^{ng}\sum_{i=1}^{n_{s_j}}\left(1 - \frac{1}{i}\left(\frac{n_{s_j} - i}{n_{s_j} - 1}\right)\right) = \alpha\, p(1-p)\, \nu \qquad (2.2.3)$$

$$\text{Asy. variance:} \qquad p(1-p)\, \nu \qquad (2.2.4)$$

The mean and variance of the SLRT depend only on the stratum size (n_{s_j}) and are independent of the stratum effect. The SLRT can be very inefficient if there is no stratum effect and a small number of subjects within stratum. For instance, if we assume proportional hazards, equal balanced design, no censoring and no stratum effect, and if each stratum has 2 (matched pair design), 4, 6, or 10 subjects, the efficiency of the SLRT is 50%, 64%, 71%, and 78% respectively when compared to the LRT under H_o. Survival time from clinical trials with a stratified design will frequently involve some censoring. It is then important to examine the extent to which the results of this section can be extended in the presence of censoring. These results will be extended to allow for censoring in chapter 3.

2.2.3 The MLRT

A recent method proposed for analyzing highly stratified data is a modified log-rank test (MLRT) given by Schoenfeld and Tsiatis (1987). Basically, these authors used the numerator of the LRT and derived a new valid denominator. Let the survival data be represented by (Y, δ, X, $\underset{\sim}{Z}$), where Y denotes the observed survival time, Y=min(T,C), where T is the survival time and C is the censoring time, δ is the failure indicator such that δ=1 if Y=T, X is a treatment indicator with X=1 if the subject receives treatment A (X=0, if treatment B), and $\underset{\sim}{Z}$ is the column vector of n_g-1 strata indicators (or binary covariates). Denote the underlying hazard function $\lambda(t|x) = \lambda_j(t)\, e^{\alpha x}$(j= 1, 2,$\cdots$, n_g), and assume proportional stratum effect. The hazard function for this situation can be written as: $\lambda(t|x, \underset{\sim}{z}) = \lambda(t)\, e^{\alpha x+\underset{\sim}{\beta}\underset{\sim}{z}}$, where α is the parameter of treatment effect, and $\underset{\sim}{\beta}$ is the parameter row vector of stratum effect. Define p_j as the proportion of patients in stratum j with X_{ji}=1 and q_j= 1 − p_j, similarly, denote p the proportion of patients with X_{ji} = 1 and q = 1 − p. Let $C_{j\gamma}$ denote the censoring distribution of stratum j, γ = A, B, treatment indicator. Two kinds of censoring are considered in Schoenfeld and Tsiatis (1987) paper, namely : (i) Censoring depends on treatment but not on stratum. (ii) Censoring depends on stratum but not on treatment. As is customary, assume that, conditional on stratum and treatment, censoring is noninformative on survival. The numerator of the MLRT can be written as:

$$T= \sum_j \sum_i \delta_{ji} \left\{ q_j\, X_{ji} S_{ji2} - p_j\, (1 - X_{ji})\, S_{ji1} \right\} \qquad (2.2.9)$$

where,

$$S_{ji1} = \frac{\sum_m \sum_l \left(q_m\, X_{ml}\right) I\left\{ Y_{ml} \geq Y_{ji} \right\}}{p \sum_m \sum_l \left(q_m\, X_{ml}\right) I\left\{ Y_{ml} \geq Y_{ji} \right\} + q \sum_m \sum_l \left(p_m (1 - X_{ml})\right) I\left\{ Y_{ml} \geq Y_{ji} \right\}}$$

which is a weighted sum of counting processes across stratum with different hazard functions. If the proportion of patients in each treatment is the same for each stratum, i.e. $p_j = p$ and $q_j = q \quad \forall\ j$ (balanced randomization), then T is essentially the numerator of the LRT. Note that the numerator of the LRT is unbiased if there are no strata effects and $p_j = p$, $q_j = q$ $\forall j$, otherwise biased. The numerator used by the MLRT is an adjusted version of the LRT allowing it to be unbiased even with unequal allocation and unequal censoring. The denominator of the MLRT which is the square root of the estimated variance is given below. The estimated variance proposed by Schoenfeld & Tsiatis (1987) is

$$V = \sum_{j=1}^{n_g}\Bigg\{p_j^2 \sum_{i=1}^{n_{s_j}}(1-X_{ji})\, d_{ji1}^2 + q_j^2 \sum_{i=1}^{n_{s_j}} X_{ji} d_{ji2}^2 - \frac{1}{n_{s_j}}\Big(\sum_{i=1}^{n_{s_j}}(1-X_{ji})d_{ji1}\Big)\Big(\sum_{i=1}^{n_{s_j}} X_{ji} d_{ji2}^2\Big)\Bigg\}, \tag{2.2.10}$$

where

$$d_{ji1} = \delta_{ji} S_{ji1} - \sum_{m,l}\left\{\frac{\Big(pq_m X_{ml} + qp_m(1-X_{ml})\Big)\delta_{ml}\, I\left\{Y_{ml} \le Y_{ji}\right\} S_{ml1}}{\sum_{m',l'}\Big(pq_m X_{m'l'} + qp_m(1-X_{m'l'})\Big) I\left\{Y_{m'l'} \ge Y_{ji}\right\}}\right\}$$

and d_{ji2} is similarly defined with S_{ji1} replaced by S_{ji2}.

In order to establish asymptotic normality, Shoenfeld and Tsiatis approximate the statistic T by a sum of independent random variables T^* with mean zero, where $T^* - T \to 0$ in probability. Let $N_{A_j}(t)$ denote the number of patients at risk at time t in stratum j from treatment A.

Define

$$\hat{\pi}_1(t) = \frac{\sum q_j N_{Aj}(t)}{\sum \left\{ pq_j N_{Aj}(t) + qp_j N_{Bj}(t) \right\}} .$$

$\hat{\pi}_2(t)$ is similarly defined by replacing $\sum q_j N_{Aj}$ with $\sum p_j N_{Bj}$. Note that $\hat{\pi}_1(t) = S_{ji1}$ and $\hat{\pi}_2(t) = S_{ji2}$. Then, T^* can be expressed as a sum of independent random variables, namely,

$$T^* = \sum^{B} p_j D_{ji1} - \sum^{A} q_j D_{ji2}, \qquad (2.2.11$$

where $\sum^{\gamma}$ for $\gamma =$ A, B is defined as the sum over j, i; j=1, $\cdots$, n_g, i=1, $\cdots$, n_{s_j}, for treatment A or B. Two censoring conditions considered in Schoenfeld and Tsiatis's paper are as follows:

(i) Censoring depends on treatment but not on the stratum; that is $C_{jA}=C_A$ and $C_{jB}=C_B$, for j=1, 2, $\cdots$, n_g.

(ii) Censoring depends on the stratum but not on treatment; that is $C_{jA}=C_{jB}$, for j=1, 2, $\cdots$, n_g.

For censoring condition (i), the hazard function for the entire population, $\lambda^*(t)$, is given by:

$$\lambda^*(u) = \frac{\sum_{j=1}^{n_g} w_j \ \lambda_j(u)\, e^{-\Lambda_j(u)}}{\sum_{j=1}^{n_g} w_j\, e^{-\Lambda_j(u)}} . \qquad (2.2.12)$$

For case (ii), the $\lambda^*(t)$ is given by:

$$\lambda^*(u) = \frac{\sum_{j=1}^{n_g} w_j \ \lambda_j(u)\, e^{-\Lambda_j(u)} C_j(u)}{\sum_{j=1}^{n_g} w_j\, e^{-\Lambda_j(u)} C_j(u)}\ , \tag{2.2.13}$$

where

$$w_l = n_{s_j} p_j q_j \tag{2.2.14}$$

is the weight chosen for censoring pattern (i) and (ii), and

$$D_{ji} \ = \delta_{ji}\pi_1(y_{ji}) - \int_0^{y_{ji}} \lambda^*(u)\, du \ = \ \delta_{ji}\pi_1(y_{ji}) - \Lambda^*(y_{ji})\ .$$

It is clear that, $\mathbb{E}(T^*) = \sum^B p_j \mathbb{E}(D_{ji1}) - \sum^A q_j \mathbb{E}(D_{ji2}) = 0$. Therefore T^* can be expressed as a sum of independent mean zero random variables, namely,

$$T^* = \ \sum^B p_j\Big(D_{ji1} - \mathbb{E}(D_{ji1})\Big) - \sum^A q_j\Big(D_{ji2} - \mathbb{E}(D_{ji2})\Big)\ .$$

The variance of T^* is,

$$\mathrm{Var}(T^*) = \sum^B p_j^2 \mathbb{E}(D_{ji1}^2) + \sum^A q_j^2 \mathbb{E}(D^2{}_{ji2}) - \sum^B p_j^2 \mathbb{E}^2(D_{ji1}) + \sum^A q_j^2 \mathbb{E}^2(D_{ji2}) \tag{2.2.13}$$

Under suitable regularity conditions, the variable $\frac{T^*}{\sqrt{\mathrm{Var}(T^*)}}$ converges to a standard normal distribution. If $\mathrm{Var}(T^*)$ is $O(N)$, then,

$$\frac{T^* - T}{\sqrt{\mathrm{Var}(T^*)}} \ \to 0 \qquad \text{in prob.}$$

Hence, $\frac{T}{\sqrt{\mathrm{Var}(T^*)}}$ converges to a standard normal.

The denominator of MLRT, defined in (2.2.2), is considered a consistent estimate for Var(T*). That is,

$$\frac{V}{Var(T^*)} \rightarrow 1 \qquad \text{in prob.}$$

Therefore by Slutsky's theorem, the MLRT $= \frac{T}{\sqrt{V}}$ would converge to a standard normal.

To evaluate the properties of the LRT, the SLRT and the MLRT, Schoenfeld and Tsiatis did a computer simulation. The results are shown in Table 2.1 (Schoenfeld & Tsiatis, 1987, Table 1).

In the table, in the second column from the right, the first entry is 1.1, which can be shown to be 1.0 in theory. Note that the ARE of MLRT versus LRT is not given because the authors reports that it is not valid. This statement is misleading; the use of the LRT when there is a stratum effect results in an inefficient test which is nevertheless still valid. In section 3.3, the asymptotic mean and variance of the MLRT will be derived and the theoretical asymptotic relative efficiencies will be presented.

Table 2.1

Simulation Results from Shoenfeld & Tsiatis (1987, Table 1)

Stratum size	Strata effect	Size MLRT	Size LRT	Size SLRT	Power MLRT	Power LRT	Power SLRT	ARE versus LRT	ARE versus SLRT
(a) *No censoring: balanced design*									
2	None	0·06	0·05	0·06	0·93	0·91	0·71	1·1	2·3
2	5×	0·06	0·03	0·05	0·91	0·87	0·71	—	2·0
2	10×	0·06	0·03	0·07	0·87	0·77	0·71	—	1·7
20	None	0·05	0·05	0·05	0·94	0·93	0·89	—	1·2
20	5×	0·05	0·03	0·05	0·89	0·80	0·90	—	0·98
50	None	0·06	0·05	0·06	0·94	0·92	0·92	—	1·1
50	5×	0·06	0·02	0·05	0·83	0·75	0·88	—	0·88
(b) *Censoring: balanced design*									
2	None	0·05	0·05	0·05	0·72	0·73	0·59	0·92	1·37
2	10×	0·05	0·03	0·05	0·82	0·58	0·68	—	1·36
(c) *No censoring: unbalanced design*									
50*	5×	0·06	0·24	0·05	0·84	0·98	0·89	—	0·89
50†	None	0·05	0·05	0·05	0·90	0·94	0·88	0·99	0·88
50†	5×	0·05	0·65	0·05	0·84	1·0	0·88	—	0·95

* 20 in one group; 30 in the other. † 16 in one group; 34 in the other.
MLRT, modified log rank test; LRT, ordinary log rank test; SLRT, stratified log rank test; ARE, asymptotic relative efficiency.

Chapter III

ASYMPTOTIC PROPERTIES OF TESTS FOR HIGHLY STRATIFIED DATA

In Chapter I and II the literature on the tests of no treatment effect for highly stratified survival data were reviewed. The most popular tests used, the LRT, the SLRT and the most recently developed test, the MLRT, were discussed. In this chapter the asymptotic mean and variance of the SLRT will be derived to allow censoring and the variance of the MLRT will be derived in the case of a balanced design. Then the asymptotic properties of these three tests will be numerically evaluated when the distribution of β's across strata are binary, uniformly or normally distributed. The numerical results will also be used to compare with Schoenfeld and Tsiatis's simulation results.

§ 3.1 *The LRT*

As discussed in section 2.1.3, the LRT is fully efficient under proportional hazards when there is no stratum effect. However, the LRT does not take into account strata effects and is an inefficient test when they exists. The asymptotic mean and variance of the LRT when the true model is PHM with stratum covariates are:

$$\text{Asy. mean: } \alpha\, p(1-p)\int g(t)\, d\, F_y(t) = \alpha^* p(1-p) \tag{3.1.1}$$

$$\text{Asy. Variance: } p(1-p)\int d\, F_y(t) \tag{3.1.2}$$

where α^* is defined as:

$$\alpha^* = \alpha \int g(t)\, d\, F_y(t) \tag{3.1.3}$$

where g(t) (Morgan, 1986) as given in equation (2.1.4) is,

$$1 + \frac{\Lambda(t)\, B(t, \beta z)}{B(t, 0)} - \frac{\Lambda(t)\, B(t, 2\beta z)}{B(t, \beta z)},$$

with $\Lambda(t) = \int \lambda(u)\, du$ and,

$$B(t, u) = \int e^{u}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta z}}}\, d\, F_z(\underset{\sim}{z}).$$

Again assume that,

$$\alpha = \log\{ \lambda_A(t)/ \lambda_B(t)\} \text{ is } O(n^{-1/2}). \tag{3.1.4}$$

If there is no censoring, then $\int dF_y(t) = 1$ and the equations can be simplified. Notice that $|\alpha^*| < |\alpha|$, as the stratum effect gets large, the mean is shrinking and becomes very inefficient.

§ 3.2 *The SLRT*

The SLRT is often used when there is stratum effect. However it is very inefficient when the stratum size is small. The asymptotic mean and variance of the SLRT when there is no censoring and the proportional hazards model holds are,

Asy. mean: $$\alpha\, p(1-p)\sum_{j}^{ng}\sum_{i=1}^{n_{s_j}}\left(1-\frac{1}{i}\left(\frac{n_{s_j}-i}{n_{s_j}-1}\right)\right)=\alpha p(1-p)\nu \qquad (3.2.1)$$

Asy. variance: $$p(1-p)\nu \qquad (3.2.2)$$

where $$\nu=\sum_{j}^{ng}\sum_{i=1}^{n_{s_j}}\left(1-\frac{1}{i}\left(\frac{n_{s_j}-i}{n_{s_j}-1}\right)\right). \qquad (3.2.3)$$

Assume there is fixed type I censoring and the same proportion of censoring in each treatment group, say, P_C. Let $\delta_{ij}= 0$ if the subject is censored, $\delta_{ij}= 1$, otherwise. Let n_{d_j} denote the number of uncensored subjects (deaths) in the j^{th} stratum. For simplicity, assume equal balanced treatment allocation, i.e. $n_{s_{Aj}}= n_{s_{Bj}}= n_{s_j}/2$.

Then,

$$E(n_{Aij})=n_{ij}\frac{n_{s_j}}{2n_{s_j}}=\frac{n_{ij}}{2}\quad,\qquad E(e_{ij})=\frac{1}{2}$$

and

$$Var(n_{Aij})=n_{ij}\left(\frac{n_{s_j}}{2n_{s_j}}\right)\left(1-\frac{n_{s_j}}{2n_{s_j}}\right)\left(1-\frac{n_{ij}-1}{n_{s_j}-1}\right)=\frac{n_{ij}}{4}\left(\frac{n_{s_j}-n_{ij}}{n_{s_j}-1}\right),$$

with

$$E(n_{Aij}^2)=\frac{n_{ij}}{4}\left(\frac{n_{s_j}-n_{ij}}{n_{s_j}-1}\right)+\frac{n_{ij}^2}{4},$$

then,

$$E(e_{ij}^2)=\frac{1}{4n_{ij}}\left(\frac{n_{s_j}-n_{ij}}{n_{s_j}-1}\right)+\frac{1}{4}$$

and

$$E(e_{ij}(1-e_{ij}))=\frac{1}{4}\left(1-\frac{1}{n_{ij}}\left(\frac{n_{s_j}-n_{ij}}{n_{s_j}-1}\right)\right).$$

Using the same notation as for the uncensored case (section 2.2.2), and conditioning on n_{Aij} and n_{ij}, the distribution of x_{ij}'s are the same as in the uncensored case. Their means are

$$\mathbb{E}(x_{ij} | \, n_{Aij} \, , n_{ij}) = \mu_{ij} = \frac{n_{Aij} \, e^{\alpha}}{n_{Aij} \, e^{\alpha} + n_{ij} - n_{Aij}} \, .$$

Expanding in a Taylor series about zero, then,

$$\mu_{ij} \doteq \frac{n_{Aij}}{n_{ij}} + \alpha \frac{n_{Aij}}{n_{ij}} \left(1 - \frac{n_{Aij}}{n_{ij}} \right) = e_{ij} + \alpha \, e_{ij} \, (1 - e_{ij}),$$

where $e_{ij} = \frac{n_{Aij}}{n_{ij}}$.

Note that with censoring, the numerator of the SLRT can be written as:

$$\sum \delta_{ij} (x_{ij} - e_{ij}) = \sum \delta_{ij} (x_{ij} - \mu_{ij}) + \alpha \sum \delta_{ij} \, e_{ij} \, (1 - e_{ij}).$$

Thus, the expected values of the numerator of the SLRT converges to:

$$\mathbb{E}(SLRT_{NUM}) = \alpha \, \mathbb{E} \sum_{j}^{n_g} \sum_{i}^{n_{s_j}} \delta_{ij} \, e_{ij} \, (1 - e_{ij})$$

$$= \alpha \sum_{j}^{n_g} \sum_{i}^{n_{s_j}} \mathbb{E} \, \delta_{ij} \, e_{ij} \, (1 - e_{ij})$$

$$= \alpha \sum_{j}^{n_g} \mathbb{E} \left\{ \sum_{i}^{n_{d_j}} e_{ij} \, (1 - e_{ij}) \right\}$$

$$= \frac{\alpha}{4} \sum_{j=1}^{n_g} \underset{n_{d_j}}{E} \left\{ \sum_{i}^{n_{d_j}} \left(1 - \frac{1}{i} \cdot \frac{n_{s_j} - i}{n_{s_j} - 1} \right) \right\}$$

$$= \frac{\alpha}{4} \sum_{j=1}^{n_g} \left\{ \sum_{m=1}^{n_{s_j}} \binom{n_{s_j}}{m} (1-\pi_j^m)\, \pi_j^{n_{s_j} - m} \sum_{i=1}^{m} \left(1 - \frac{1}{i} \cdot \frac{n_{s_j} - i}{n_{s_j} - 1} \right) \right\}$$

$$= \frac{\alpha\, \nu^c}{4},$$

where,

$$\nu^c = \sum_{j=1}^{n_g} \left\{ \sum_{m=1}^{n_{s_j}} \binom{n_{s_j}}{m} (1-\pi_j^m)\, \pi_j^{n_{s_j} - m} \sum_{i=1}^{m} \left(1 - \frac{1}{i} \cdot \frac{n_{s_j} - i}{n_{s_j} - 1} \right) \right\}, \tag{3.2.4}$$

and $\pi_j = e^{-\Lambda(t)\, e^{\beta_j}}$, is the probability of been censored.

For the denominator, it can be shown that,

$$SLRT_{DEN} \rightarrow E \sum_{j}^{n_g} \sum_{i}^{n_{s_j}} \delta_{ij}\, e_{ij}\, (1 - e_{ij})$$

$$= \frac{\nu^c}{4},$$

where ν^c is defined in equation (3.2.4).

In summary, for equally balanced design with Type I fixed censoring, if the treatment effect, α, is of order $O(n^{-1/2})$, the SLRT is asymptotically normally distributed with mean and variance

Asy. mean: $\frac{\alpha\, \nu^c}{4}$ (3.2.5)

Asy. variance: $\frac{\nu^c}{4}$. (3.2.6)

§ 3.3 *The MLRT - Derivation of the Variance of MLRT*

Following the notation defined in section 2.2.3, for the case of no stratum effect and balanced design, the numerator of the MLRT is the same as the numerator of LRT. That is,

$$T = \sum_j \sum_i \delta_{ji} \left\{ q_j \, X_{ji} S_{ji2} - p_j \, (1 - X_{ji}) \, S_{ji1} \right\},$$

where, p_j is the proportion of patients in stratum j with treatment A (i.e. $X_{ji}=1$) and $q_j = 1 - p_j$, $\delta_{ij} = 0$ if the subject is censored, $\delta_{ij} = 1$, otherwise, and,

$$S_{ji1} = \frac{\sum_m \sum_l \left(q_m \, X_{ml} \right) I \left\{ Y_{ml} \geq Y_{ji} \right\}}{p \sum_m \sum_l \left(q_m \, X_{ml} \right) I \left\{ Y_{ml} \geq Y_{ji} \right\} + q \sum_m \sum_l \left(p_m (1 - X_{ml}) \right) I \left\{ Y_{ml} \geq Y_{ji} \right\}}.$$

Let $\lambda_j(t)$ be the hazard function for stratum j and $\lambda^*(t)$ be the hazard function for the entire population. Two censoring conditions considered in section 2.2.3 were, (i) Censoring depends on treatment but not on stratum; and (ii) Censoring depends on stratum but not on treatment. The hazard function for the entire population, $\lambda^*(t)$, for censoring pattern (i) is given by:

$$\lambda^*(u) = \frac{\sum_{j=1}^{n_g} w_j \; \lambda_j(u) \, e^{-\Lambda_j(u)}}{\sum_{j=1}^{n_g} w_j \, e^{-\Lambda_j(u)}} \quad , \tag{3.3.1}$$

where $w_j = n_j p_j q_j$, as defined in section (2.2.14). For case (ii), the $\lambda^*(t)$ is

$$\lambda^*(u) = \frac{\sum_{j=1}^{n_g} w_j \; \lambda_j(u) \, e^{-\Lambda_j(u)} C_j(u)}{\sum_{j=1}^{n_g} w_j \, e^{-\Lambda_j(u)} C_j(u)} \, . \tag{3.3.2}$$

3.3.1 *The MLRT - For the case of no censoring*

In this section, the theoretical variance of T^* is derived and evaluated by numerical integration in section 3.4. The variance of T^* will be derived in this section for the case of no censoring and balanced design. In the case of no censoring,

$$D_{ji} = 1 - \int_0^{t_{ji}} \lambda^*(u)\, du = 1 - \Lambda^*(t_{ji}) ,$$

and from equation (2.2.13), the variance of T^* can be written as:

$$Var(T^*) = \sum^B p_j^2 E(D_{ji1}^2) + \sum^A q_j^2 E(D^2{}_{ji2}) - \sum^B p_j^2 E^2(D_{ji1}) + \sum^A q_j^2 E^2(D_{ji2}).$$

Schoenfeld & Tsiatis relied on a simulation study to compare the ARE of the MLRT with the SLRT and the LRT rather than deriving the variance as a function of stratum effect. Since $p_j = p$, $q_j = q$, $\forall\, j$ and $D_{ji1} = D_{ji2}$,

$$Var(T^*) = p^2 \sum^2 \left\{ E(D_{ji}^2) - \left(E(D_{ji})\right)^2 \right\} + q^2 \sum^A \left\{ E(D_{ji}^2) - \left(E(D_{ji})\right)^2 \right\} \qquad (3.3.3)$$

where,

$$E_j(D_{ji}^2) = 1 - 2 \cdot E_j\left(\Lambda^*(x_{ji})\right) + E_j\left(\Lambda^*(x_{ji})\right)^2 \qquad (3.3.4)$$

$$E_j\left(\Lambda^*(t)\right) = \int \Lambda^*(t)\, d\, F_j(t)\, dt$$

$$= -\int \Lambda^*(t)\, d\, e^{-\Lambda_j(t)}$$

$$= -\left\{ \Lambda^*(t)\, e^{-\Lambda_j(t)} \right\}_0^{\infty} + \int \lambda^*(u)\, e^{-\Lambda_j(t)} dt$$

For this case, $\Lambda^*(t) \rightarrow \infty$ as $t \rightarrow \infty$ and the first term goes to zero, therefore,

$$E_j\left(\Lambda^*(t)\right) = \int \lambda^*(u)\, e^{-\Lambda_j(t)} dt. \tag{3.3.5}$$

and,

$$\begin{aligned} E_j\left(\Lambda^*(t)\right)^2 &= \int \Lambda^{*2}(t)\, d\, F_j(t) \\ &= -\int \Lambda^{*2}(t)\, d\, e^{-\Lambda_j(t)} \\ &= -\left\{ [\Lambda^*(t)]^2\, e^{-\Lambda_j(t)} \right\}_0^\infty + 2 \cdot \int \lambda^*(t)\, \Lambda^*(t)\, e^{-\Lambda_j(t)} dt \\ &= 2 \cdot \int \Lambda^*(t)\, \lambda^*(t)\, e^{-\Lambda_j(t)} dt \end{aligned} \tag{3.3.6}$$

Combining equations (3.3.4)-(3.3.6), then,

$$E_j\,(\,D_{ji}^2\,) = 1 - 2\int_0^\infty \lambda^*(t)\, e^{-\Lambda_j(u)} dt + 2\int_0^\infty \lambda^*(t)\, \Lambda^*(t)\, e^{-\Lambda_j(u)} dt$$

$$= 1 - 2 \cdot E_j\left(\Lambda^*(t)\,\right) + E_j\left(\Lambda^*(t)\right)^2$$

and

$$\left(E_j(D_{ji})\right)^2 = \left(1 - E_j(\Lambda^*(t)\right)^2 = 1 - 2 \cdot E_j\left(\Lambda^*(t)\right) + \left\{E_j\left(\Lambda^*(t)\right)\right\}^2$$

therefore,

$$V_j(D_{ji}) = E_j(D_{ji}^2) - \left(E_j(D_{ji})\right)^2 = E_j\left(\Lambda^*(t)\right)^2 + \left\{E_j\left(\Lambda^*(t)\right)\right\}^2$$

$$= 2\int_0^\infty \Lambda^*(t)\, \lambda^*(t)\, e^{-\Lambda_j(t)} dt - \left(\int_0^\infty \lambda^*(t)\, e^{-\Lambda_j(t)} dt\right)^2.$$

Since,

$$\Lambda^*(t) = \int_0^t \lambda^*(u)\, du = -\log \frac{1}{n_g} \sum e^{-\Lambda_j(t)}$$

and

$$S^*(t) = e^{-\Lambda^*(t)},$$

from (3.3.3),

$$\mathrm{Var}(T^*) = p\,q \sum_{j=1}^{n_g} V_j(D_{ji})$$

$$= p\,(1-p) \sum \left\{ 2 \int_0^\infty \left(-\log \tfrac{1}{n_g} \sum e^{-\Lambda_i(t)} \right) \frac{\sum w_i \lambda_i(t) e^{-\Lambda_i(t)}}{\sum w_i e^{-\Lambda_i(t)}} e^{-\Lambda_j(t)} dt \right.$$

$$\left. - \left(\int_0^\infty \frac{\sum w_i \lambda_i(t)\, e^{-\Lambda_i(t)}}{\sum w_i e^{-\Lambda_i(t)}} e^{-\Lambda_j(t)} dt \right)^2 \right\}$$

$$= p(1-p)\,\eta \qquad (3.3.7)$$

where,

$$\eta = \sum_j \left\{ 2 \int_0^\infty \left(-\log \tfrac{1}{n_g} \sum e^{-\Lambda_i(t)} \right) \frac{\sum w_i \lambda_i(t) e^{-\Lambda_i(t)}}{\sum w_i e^{-\Lambda_i(t)}} e^{-\Lambda_j(t)} dt \right.$$

$$\left. - \left(\int_0^\infty \frac{\sum w_i \lambda_i(t)\, e^{-\Lambda_i(t)}}{\sum w_i e^{-\Lambda_i(t)}} e^{-\Lambda_j(t)} dt \right)^2 \right\} \qquad (3.3.8)$$

In summary, the MLRT is asymptotically normally distributed with mean

$$\alpha^*\, p(1-p) \qquad (3.3.9)$$

and variance given by equation (3.3.7).

3.3.2 *The MLRT - For the case of censoring*

Recall that Y=min(T,C) is the observed survival time, where T is the survival time and C is the censoring time. For censoring pattern (i), that is when censoring depends on treatment but not on the stratum, it can be shown that,

$$\pi_1(u) = \frac{C_1(u)}{pC_1(u)+qC_2(u)} \quad \text{and} \quad \pi_2(u) = \frac{C_2(u)}{pC_1(u)+qC_2(u)} \ , \tag{3.3.9}$$

for censoring pattern (ii), that is when censoring depends on the stratum but not on treatment, then, $\pi_1(u)=\pi_2(u)=1$. It is of more interest to assume censoring pattern (ii) in this dissertation as the generalization for pattern (i) can be easily derived. Note that for censoring pattern (ii), $D_{ji\mathrm{A}}=D_{ji\mathrm{B}}=D_{ji}$, and,

$$D_{ji} \ = \delta_{ji} - \int_0^{y_{ji}} \lambda^*(u)\, du \ = \ \delta_{ji} - \Lambda^*(y_{ji}) \ , \tag{3.3.10}$$

where $\lambda^*(t)$ is defined in equation (3.3.2), and,

$$\Lambda^*(y) = \int_0^y \lambda^*(u)du \ = \ -\log \tfrac{1}{n_g} \sum e^{-\Lambda_j(y)} C_j(y).$$

The probability density function of Y=min(T,C) for balanced design and case (ii) censoring can be written as:

$$f_y(y) = [1-F_t(y)]\, f_c(y) + [1-F_c(y)] f_t(y)$$

$$= -e^{-\Lambda_j(y)} dC_j(y) + C_j(y)\lambda_j(y)\, e^{-\Lambda_j(y)}.$$

From (2.2.13), the variance of T^* can be written as:

$$\mathrm{Var}(T^*) = \sum^B p_j^2 \mathbb{E}(D_{ji1}^2) + \sum^A q_j^2 \mathbb{E}(D^2{}_{ji2}) - \sum^B p_j^2 \mathbb{E}^2(D_{ji1}) + \sum^A q_j^2 \mathbb{E}^2(D_{ji2}).$$

Since $p_j = p$, $q_j = q$, $\forall\, j$ and $D_{ji1} = D_{ji2}$, from equation (3.3.3),

$$\mathrm{Var}(T^*) = p^2 \sum^2 \left\{ \mathbb{E}(D_{ji}^2) - \left(\mathbb{E}(D_{ji}) \right)^2 \right\} + q^2 \sum^A \left\{ \mathbb{E}(D_{ji}^2) - \left(\mathbb{E}(D_{ji}) \right)^2 \right\},$$

where from equation (3.3.10),

$$D_{ji}^2 = \delta_{ji} - 2 \cdot \delta_{ji} \Lambda^*(y_{ji}) + (\Lambda^*(y_{ji}))^2.$$

It can be shown that,

$$\mathbb{E}_j(D_{ji}^2) = \mathbb{E}\delta_{ji} - 2 \cdot \mathbb{E}_j\left(\delta_{ji} \Lambda^*(y_{ji}) \right) + \mathbb{E}_j\left(\Lambda^*(y_{ji}) \right)^2,$$

with $\mathbb{E}\delta_{ji} = P_{cj}$ the proportion of censoring in stratum j.

$$\mathbb{E}_j\left(\delta_{ji} \Lambda^*(y) \right) = \int \delta_{ji} \Lambda^*(y)\, \mathrm{d}\, F_j(y)\, \mathrm{d}y$$

$$= \int \Lambda^*(y)\, C_j(y) \lambda_j(y) e^{-\Lambda_j(y)} \mathrm{d}y$$

$$= -\int \Lambda^*(y)\, C_j(y)\, \mathrm{d}\, e^{-\Lambda_j(y)}$$

$$= -\left\{ \Lambda^*(y)\, C_j(y)\, e^{-\Lambda_j(y)} \right\}_0^\infty + \int \lambda^*(u) C_j(y)\, e^{-\Lambda_j(y)} \mathrm{d}y. \qquad (3.3.11)$$

Since $\Lambda_j^*(y) \to \infty$ as $y \to \infty$ and $\Lambda_j(0) = 0$ by definition, the first term goes to zero, therefore,

$$\mathbb{E}_j\left(\delta_{ji} \Lambda^*(y) \right) = \int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} \mathrm{d}y\,.$$

Now,

$$E_j\left(\Lambda^*(y)\right)^2 = \int \Lambda^{*2}(y)\, d\, F_j(y)$$

$$= -\int \Lambda^{*2}(y) C_j(y)\, d\, e^{-\Lambda_j(y)}$$

$$= -\left\{ [\Lambda^*(y)]^2 C_j(y)\, e^{-\Lambda_j(y)} \right\}_0^\infty + 2 \cdot \int \lambda^*(y)\, \Lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy$$

$$= 2 \cdot \int \Lambda^*(y)\, \lambda^*(y)\, C_j(y)\, e^{-\Lambda_j(y)} dy. \qquad (3.3.12)$$

Combining equations (3.3.11)-(3.3.12), then,

$$E_j\,(\,D_{ji}^2\,) = P_{cj} - 2\int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy + 2\int_0^\infty \lambda^*(y)\, \Lambda^*(y)\, C_j(y) e^{-\Lambda_j(y)} dy$$

and

$$\left(E_j(D_{ji})\right)^2 = P_{cj}^2 - 2 \cdot P_{cj} \int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy + \left\{ \int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy \right\}^2,$$

therefore,

$$V_j(D_{ji}) = E_j(D_{ji}^2) - \left(E_j(D_{ji})\right)^2$$

$$= P_{cj} - P_{cj}^2 + 2(P_{cj} - 1) \int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy$$

$$+ 2\int_0^\infty \lambda^*(y)\, \Lambda^*(y)\, C_j(y)\, e^{-\Lambda_j(y)} dy - \left\{ \int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy \right\}^2.$$

Since,

$$\Lambda^*(y) = \int_0^y \lambda^*(u)\, du = -\log \frac{1}{n_g} \sum e^{-\Lambda_j(y)}$$

and $S^*(y) = e^{-\Lambda^*(y)}$, from equation (3.3.3),

$$\mathrm{Var}(T^*) = p\,q \sum_{j=1}^{n_g} V_j(D_{ji})$$

$$= p\,(1-p)\sum P_{c_j} - P_{c_j}^2 + 2(P_{c_j}-1)\int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy$$

$$+ 2\int_0^\infty \lambda^*(y)\,\Lambda^*(y)\,C_j(y)\, e^{-\Lambda_j(y)} dy - \left\{\int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy\right\}^2$$

$$= p(1-p)\,\eta^c, \qquad (3.3.13)$$

where,

$$\eta^c = \sum P_{c_j} - P_{c_j}^2 + 2(P_{c_j}-1)\int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy$$

$$+ 2\int_0^\infty \lambda^*(y)\,\Lambda^*(y)\,C_j(y)\, e^{-\Lambda_j(y)} dy - \left\{\int \lambda^*(y) C_j(y)\, e^{-\Lambda_j(y)} dy\right\}^2. \qquad (3.3.14)$$

In summary, the MLRT is asymptotically normally distributed with mean $\alpha^* p (1-p)$, and variance given by equation (3.3.13).

§ 3.4 *Evaluation of Asymptotic Relative Efficiencies -*

Comparisons of LRT, SLRT and MLRT as functions of strata effects and strata sizes

The asymptotic distribution of these three existing tests are summarized in Table 3.1 and Table 3.2 for the case of no censoring. Table 3.1 is for the case of fixed number of strata with the sample size within each stratum asymptotically approaching infinity. Table 3.2 is for the case of highly stratified data where the number of strata increases with the sample size and the number within each stratum stays fixed.

When there is no stratum effect these three tests are asymptotically identically distributed as the stratum size approaches infinity while the number of strata is

(need revision)

Table 3.1

Summary of Asymptotic Distributions of the LRT, the SLRT and the MLRT

Assume†: $n_{s_j} \to \infty$, $\forall\ j$, n_g *fixed*, $p = \frac{1}{2}$ *and no censoring*

	no stratum effect	stratum effect
LRT	$N(\frac{\alpha}{4}, \frac{1}{4})$	$N(\frac{\alpha^*}{4}, \frac{1}{4})$
SLRT	$N(\frac{\alpha}{4}, \frac{1}{4})$	$N(\frac{\alpha}{4}, \frac{1}{4})$
MLRT	$N(\frac{\alpha}{4}, \frac{1}{4})$	$N(\frac{\alpha^*}{4}, \frac{\eta}{4})$

where α and α^* are defined in equation (3.1.3) and (3.1.4)
† n_{s_j}: stratum size of stratum j; n_g: number of strata

Table 3.2

Summary of Asymptotic Distributions of the LRT, the SLRT and the MLRT
Assume: $n_g \to \infty$, n_{s_j} *fixed*, $\forall\ j$, $p = \frac{1}{2}$ *and no censoring*

	no stratum effect	stratum effect
LRT	$N(\frac{\alpha}{4}, \frac{1}{4})$	$N(\frac{\alpha^*}{4}, \frac{1}{4})$
SLRT	$N(\frac{\alpha\nu}{4}, \frac{\nu}{4})$	$N(\frac{\alpha\nu}{4}, \frac{\nu}{4})$
MLRT	$N(\frac{\alpha}{4}, \frac{1}{4})$	$N(\frac{\alpha^*}{4}, \frac{\eta}{4})$

where η is defined in equation (3.3.8) and ν is in equation (3.2.3)

fixed. When there is a stratum effect, the asymptotic distribution of the SLRT is the same as when there is no stratum effect and is fully efficient, but the asymptotic mean of the LRT and the MLRT are smaller than when there is no stratum effect and the asymptotic variance of the MLRT is smaller than where there is no stratum effect.

When the stratum size is considered fixed and the number of stratum approaches infinity so that the sample size goes to infinity, the asymptotic distributions of the LRT and the MLRT are the same as in Table 3.1. However, the SLRT is less efficient with its mean and variance decreased. Again, the asymptotic distribution of the SLRT is the same whether there is a stratum effect or not.

The results above summarized the asymptotic distribution of these three tests for the case of a fixed number of strata with the sample size within each stratum asymptotically approaching infinity in which the data were highly stratified. We consider the comparison of these three test procedures on the basis of their asymptotic distributional properties. The asymptotic relative efficiency (ARE), or Pitman efficiency, is used as a measure of the amount of precision lost by the use of the test instead of the theoretically optimal test. The ARE represents the relative efficiency of the test under conditions in which the sample size is infinite and the alternative hypothesis is close to the null hypothesis (i.e. consider only local alternatives). Let S and T be two tests such that each test statistic is asymptotically normally distributed and each tests the same H_o and the same class of alternative hypotheses. The ARE of S with respect to T is given by:

$$\mathrm{ARE}\left(\frac{S}{T}\right) = \frac{(E'(S))^2}{V(S)} \cdot \frac{V(T)}{(E'(T))^2}$$

$$= \frac{\left(\frac{\partial E(S)}{\partial \alpha}\right)^2}{\left(\frac{\partial E(T)}{\partial \alpha}\right)^2} \cdot \frac{Var(T)}{Var(S)}\Bigg|_{\alpha=0}$$

where $\mathbb{E}'(S)$ is the derivative of the expected value of S as H_a approaches H_o. Since at the null hypothesis, H_o:$\alpha = 0$,

$$\mathbb{E}(S) = 0 \text{ and } \mathbb{E}(T) = 0,$$

as $\alpha \rightarrow 0$. Using a Taylor series expansion, it can be shown that,

$$\frac{\partial \mathbb{E}(S)}{\partial \alpha} \approx \frac{\mathbb{E}(S)}{\alpha}, \text{ for } \alpha \text{ near } 0.$$

Thus, the ARE of test S vs test T can be written as:

$$\frac{(\mathbb{E}(S))^2}{(\mathbb{E}(T))^2} \cdot \frac{\text{Var}(T)}{\text{Var}(S)}, \text{ for } \alpha \text{ near } 0.$$

Therefore the AREs of these three tests can be summarized as follows:

$$\text{ARE(LRT/ SLRT)} = \frac{(\alpha^*)^2}{(\alpha\nu)^2} \cdot \nu = \frac{(\alpha^*)^2}{\alpha^2\nu},$$

$$\text{ARE(LRT/ MLRT)} = \frac{(\alpha^*)^2}{(\alpha^*)^2} \cdot \eta = \eta,$$

$$\text{ARE(SLRT/ MLRT)} = \frac{(\alpha\nu)^2}{(\alpha^*)^2} \cdot \frac{\eta}{\nu} = \frac{\alpha^2\nu\eta}{(\alpha^*)^2},$$

where α and α^* are defined in equation (3.1.3) and (3.1.4), η is defined in equation (3.3.8) and ν is in equation (3.2.3).

Since the asymptotic means and variances, and the ARE's can be explicitly expressed as functions of the degree of heterogeneity between strata (β), the ARE's can be numerically evaluated. Three distributions of the strata effects, β, were considered as

examples, namely, β is dichotomously distributed (i.e. assume half of the strata have hazard $\lambda(t)$, and the other half have hazard $\lambda(t)\cdot e^{\beta}$), β is uniformly distributed, and β is normally distributed. The ARE's are evaluated by numerical integrations which were based upon the use of equally spaced points or Gaussian Quadrature to approximate the integral. The FORTRAN programs (all in double-precision arithmetic) are given in APPENDIX A.

3.4.1 *When beta is dichotomous distributed*

This section will consider the case where half of the strata have hazard $\lambda(t)$ and half of them have hazard $\lambda(t)\cdot e^{\beta}$ and the number of stratum $n_g \rightarrow \infty$ as the sample size $n \rightarrow \infty$. Figures 3.1 to 3.10 show the comparison of the asymptotic means, variances and efficiencies for the LRT, the SLRT and the MLRT as a function of stratum effect (parameter β) and stratum size calculated by numerical integration.

Figures 3.1 and 3.2, show the asymptotic mean and variance of the numerator of the LRT, the statistics are given in equations (3.1.1) and (3.1.2). The asymptotic mean is $0.2577\alpha\sqrt{n}$ when there is no stratum effect (β=0); however as the stratum effect gets larger, the mean becomes smaller (converges to $0.125\alpha\sqrt{n}$ as the stratum effect, $\beta \rightarrow \infty$), while the asymptotic variance stays constant at 0.25.

Figures 3.3 and 3.4, give the asymptotic mean and variance of the numerator of the SLRT, the statistics are given in equations (3.2.1) and (3.2.2). As it is shown in the figures, the asymptotic mean and variance of SLRT are independent of β - the stratum effect; however, both the mean and variance depend on the number of subjects within each stratum.

Figures 3.5 and 3.6 indicate the relationship of the asymptotic mean and variance of the MLRT and the stratum effect -β. The asymptotic mean of the MLRT is

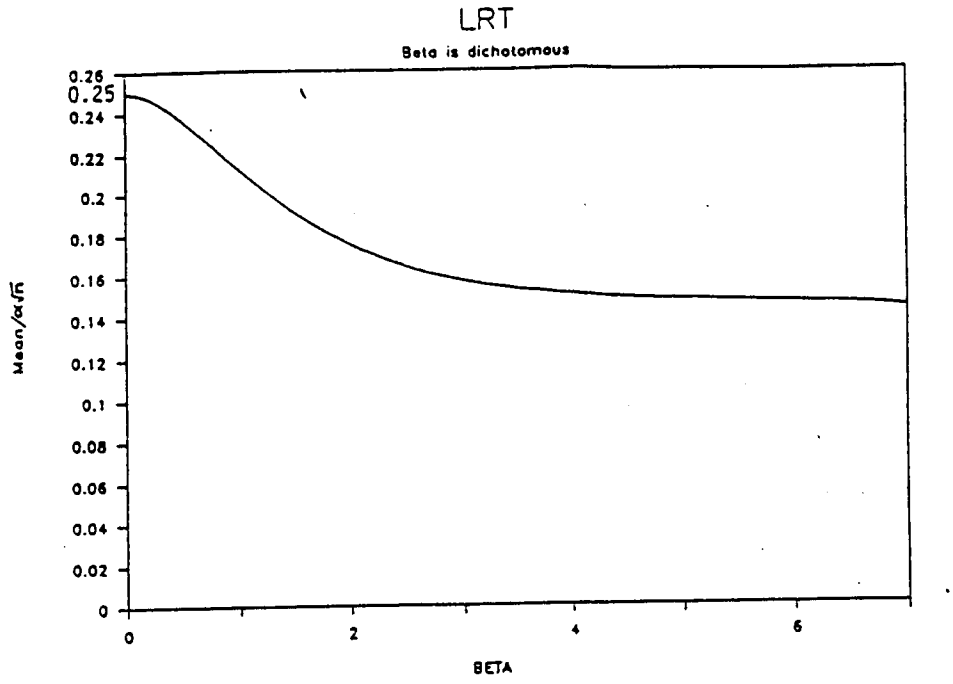

Figure 3.1: The Asymptotic Mean of the LRT as a Function of Stratum Effect, β

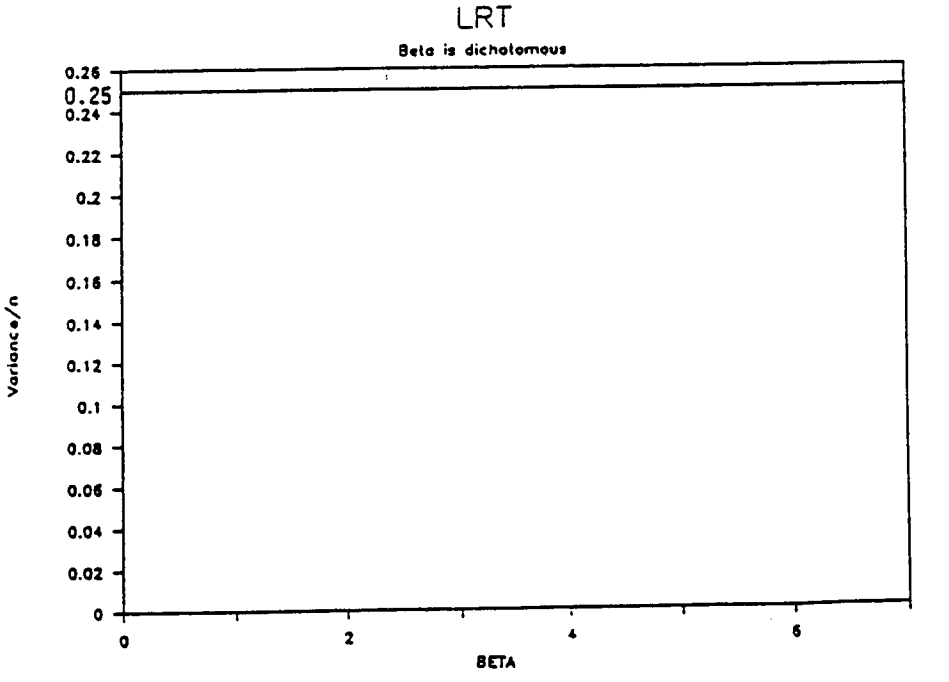

Figure 3.2: The Asymptotic Variance of the LRT as Function of Stratum Effect, β

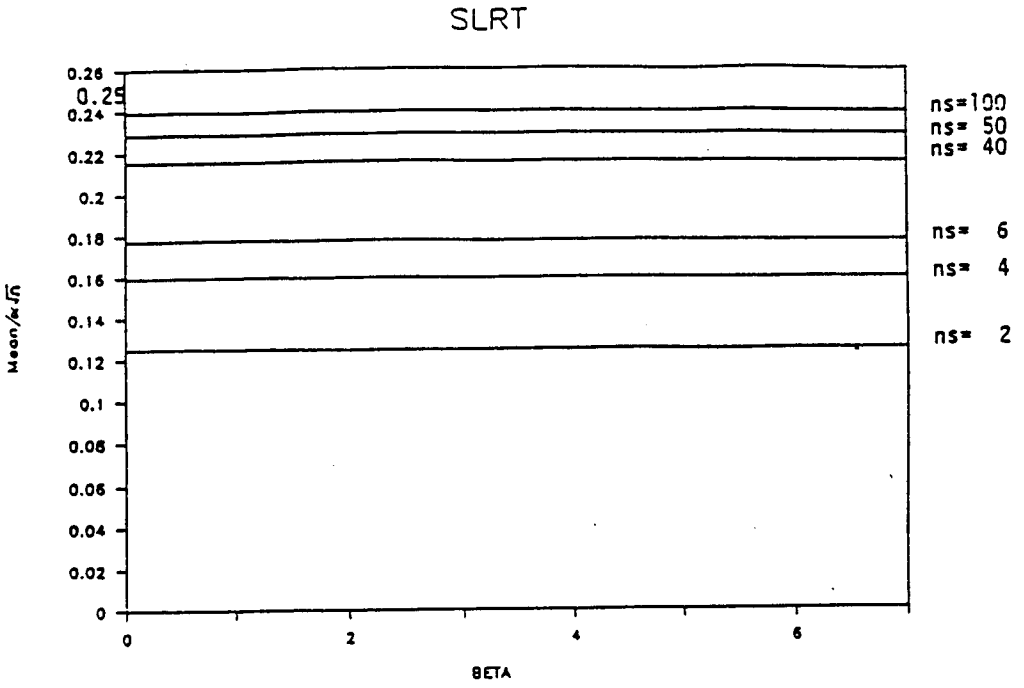

Figure 3.3: The Asymptotic Mean of the SLRT as a Function of Stratum Effect, β, and Stratum Size, ns

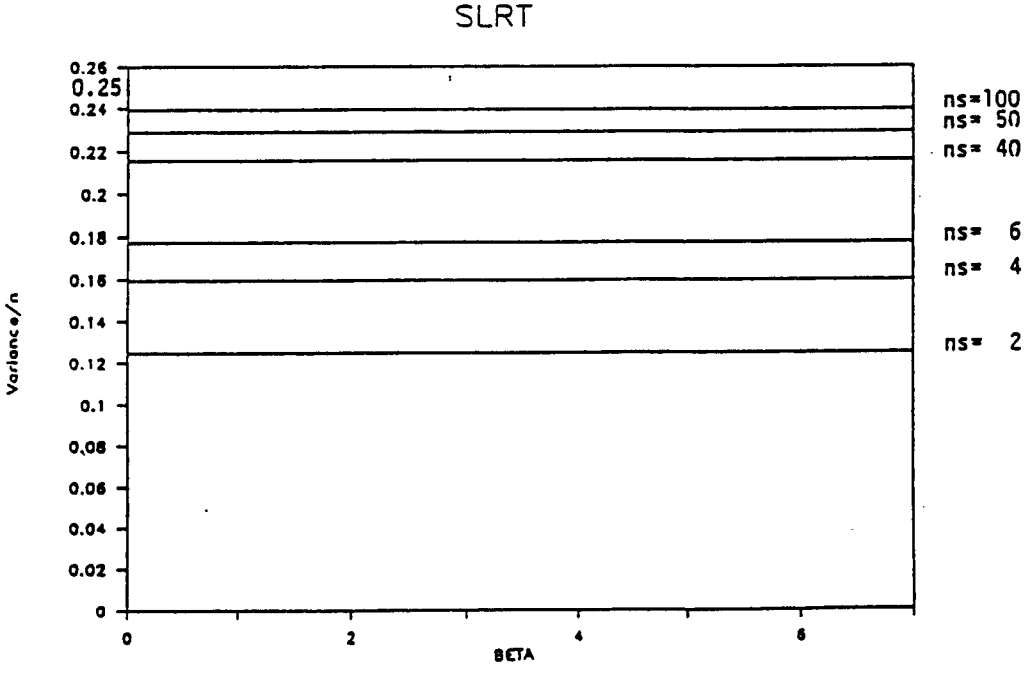

Figure 3.4: The Asymptotic Variance of the SLRT as a Function of Stratum Effect, β, and Stratum Size, ns

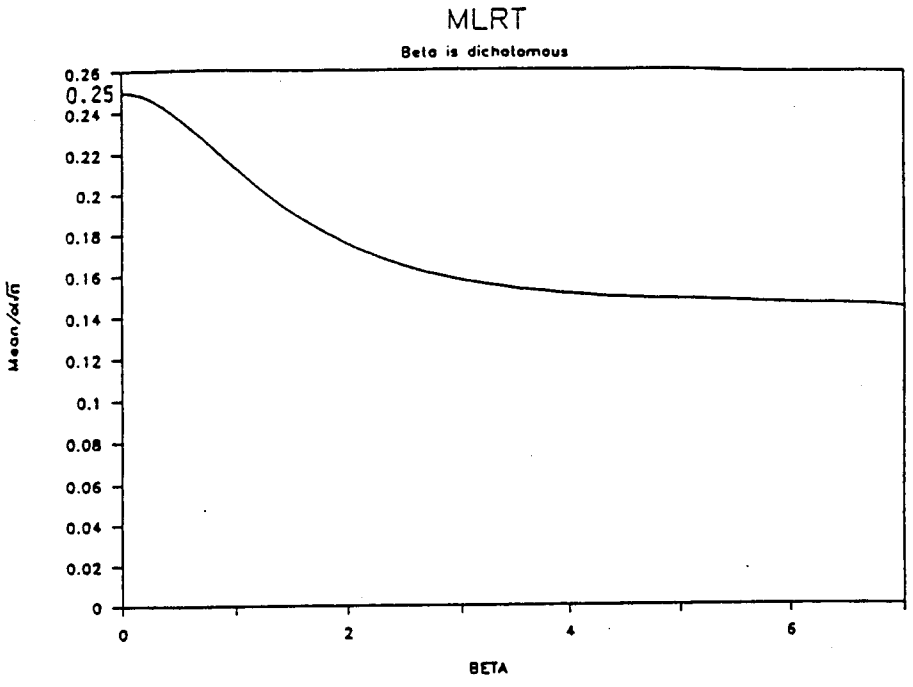

Figure 3.5: The Asymptotic Mean of the MLRT as a Function of Stratum Effect, β

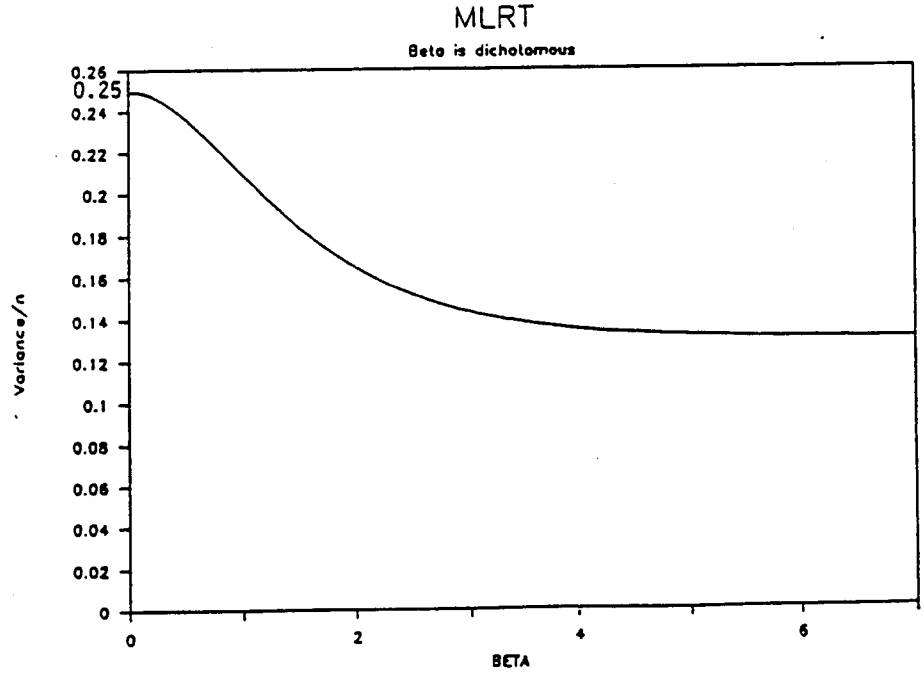

Figure 3.6: The Asymptotic Variance of the MLRT as a Function of Stratum Effect, β

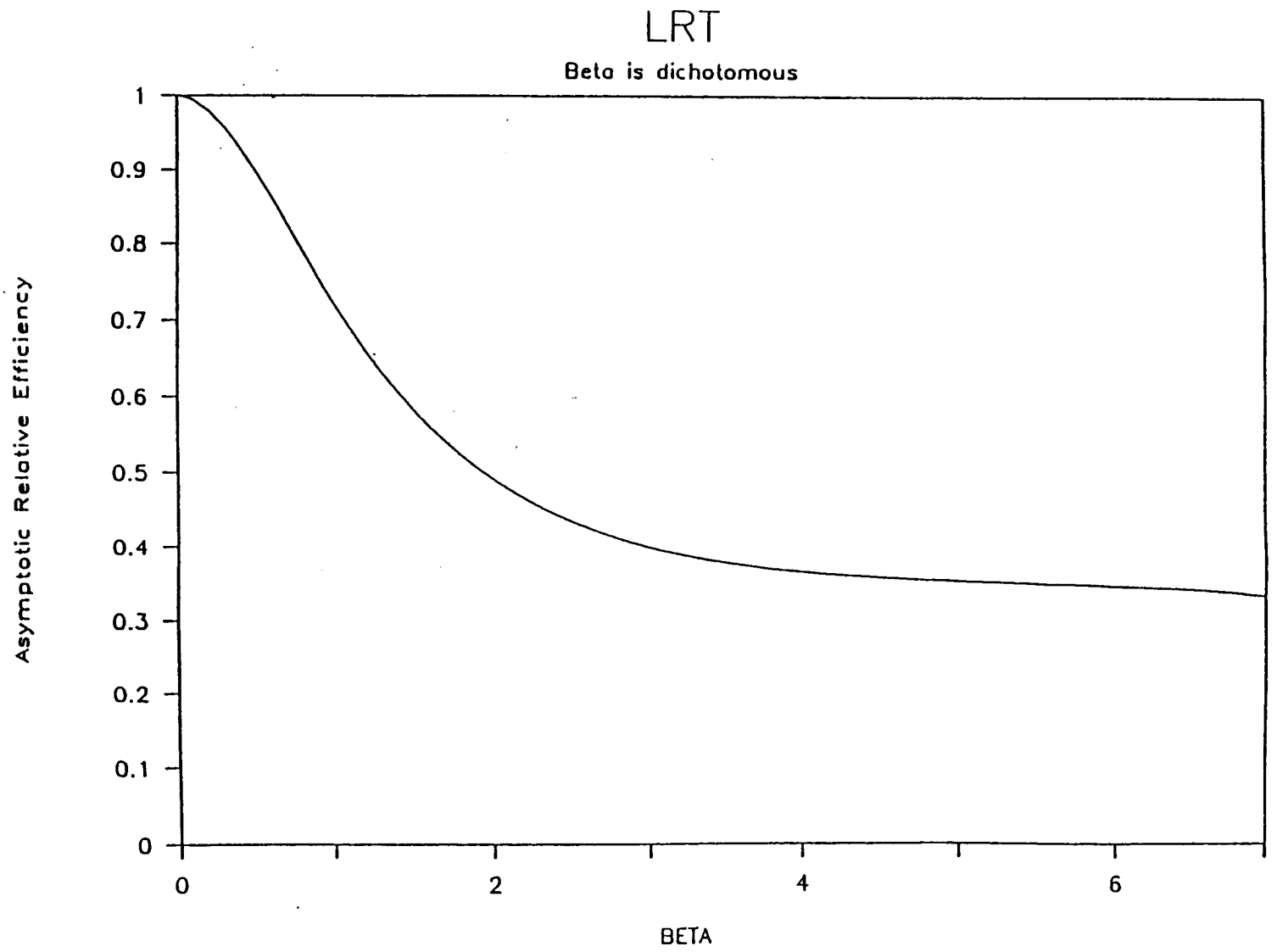

Figure 3.7: The Asymptotic Relative Efficiency of the LRT as a Function of β

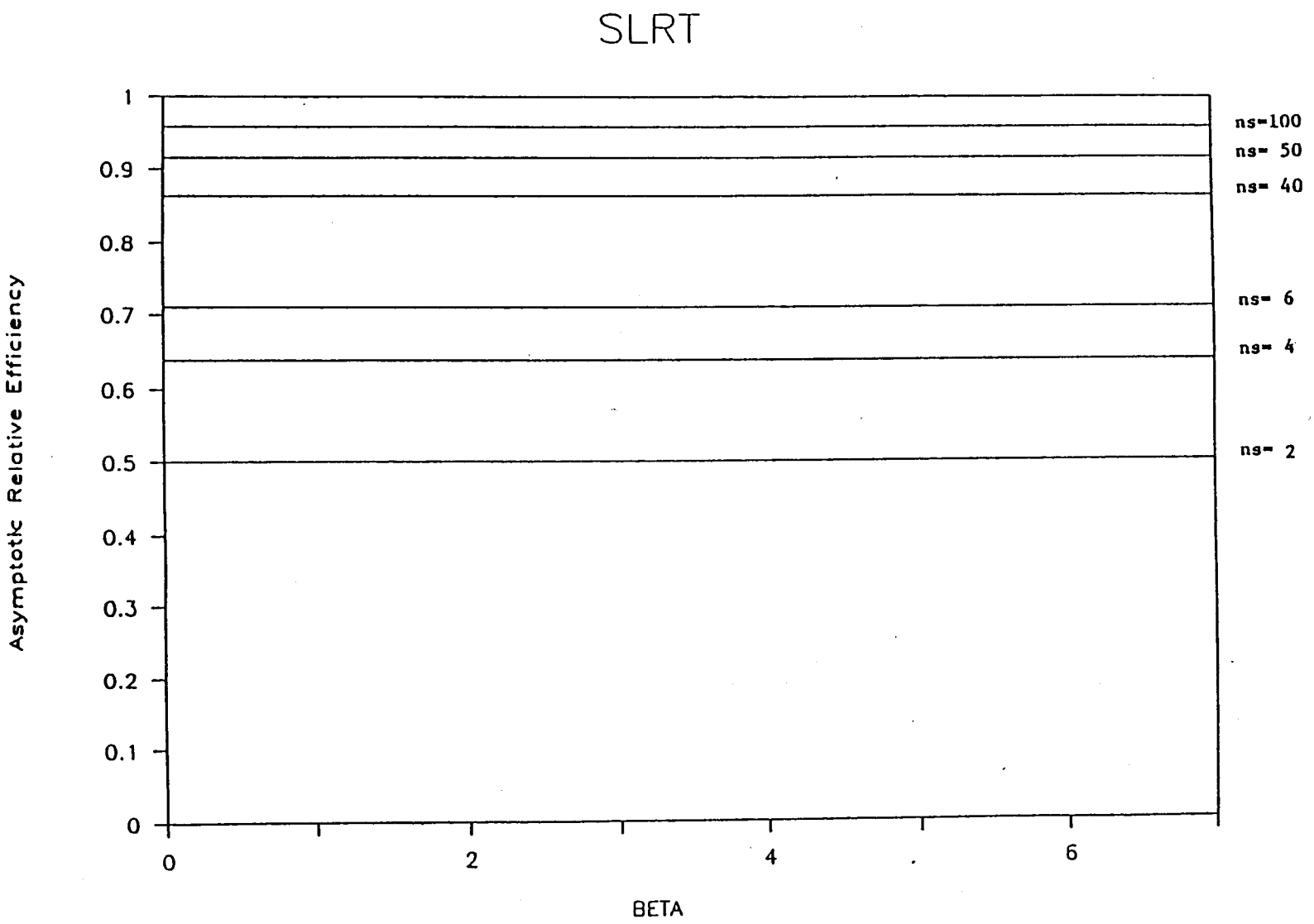

Figure 3.8: The ARE of the SLRT as a Function of β and ns
β: stratum effect; ns: stratum size

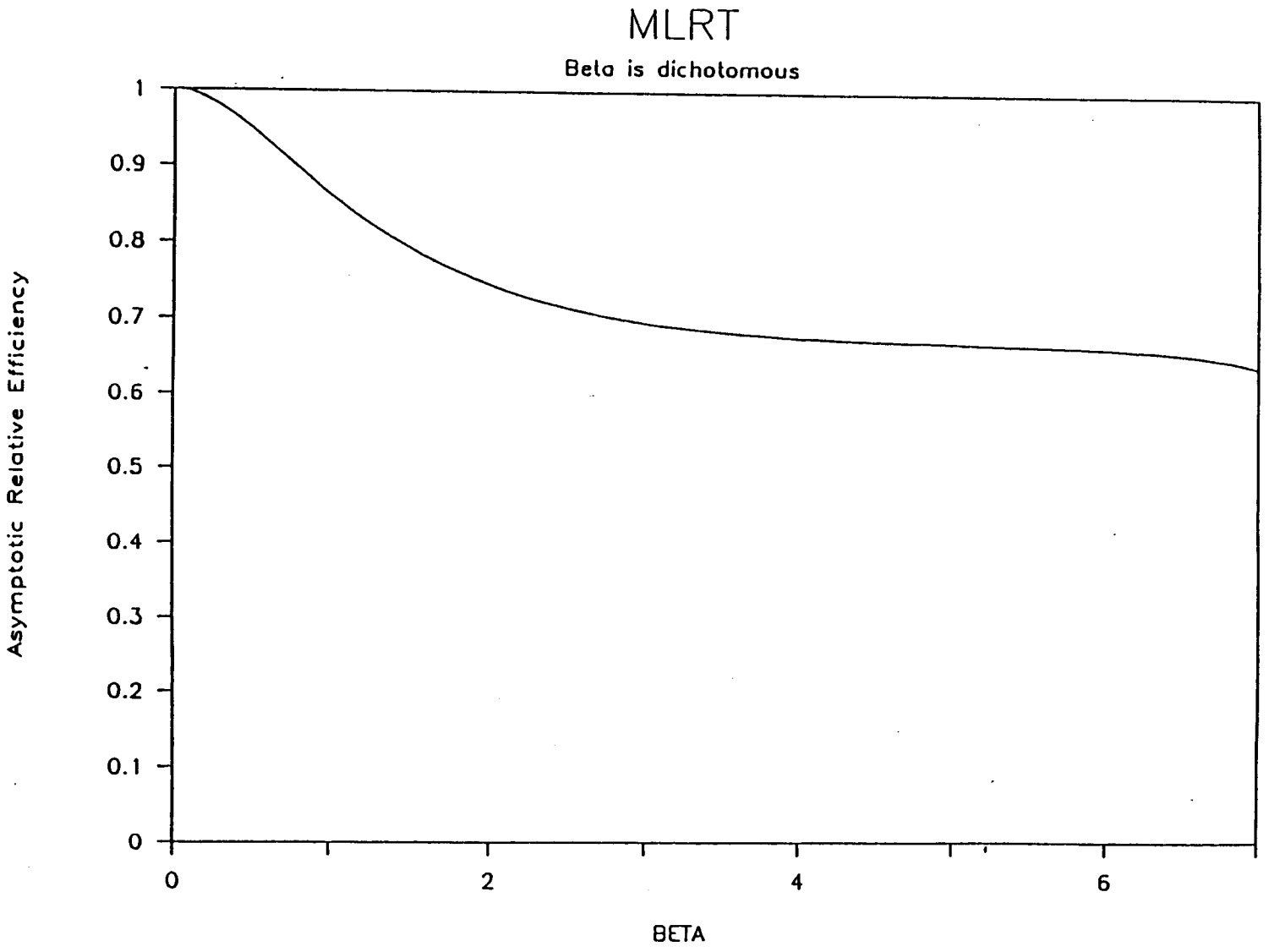

Figure 3.9: The Asymptotic Relative Efficiency of the MLRT as a Function of β

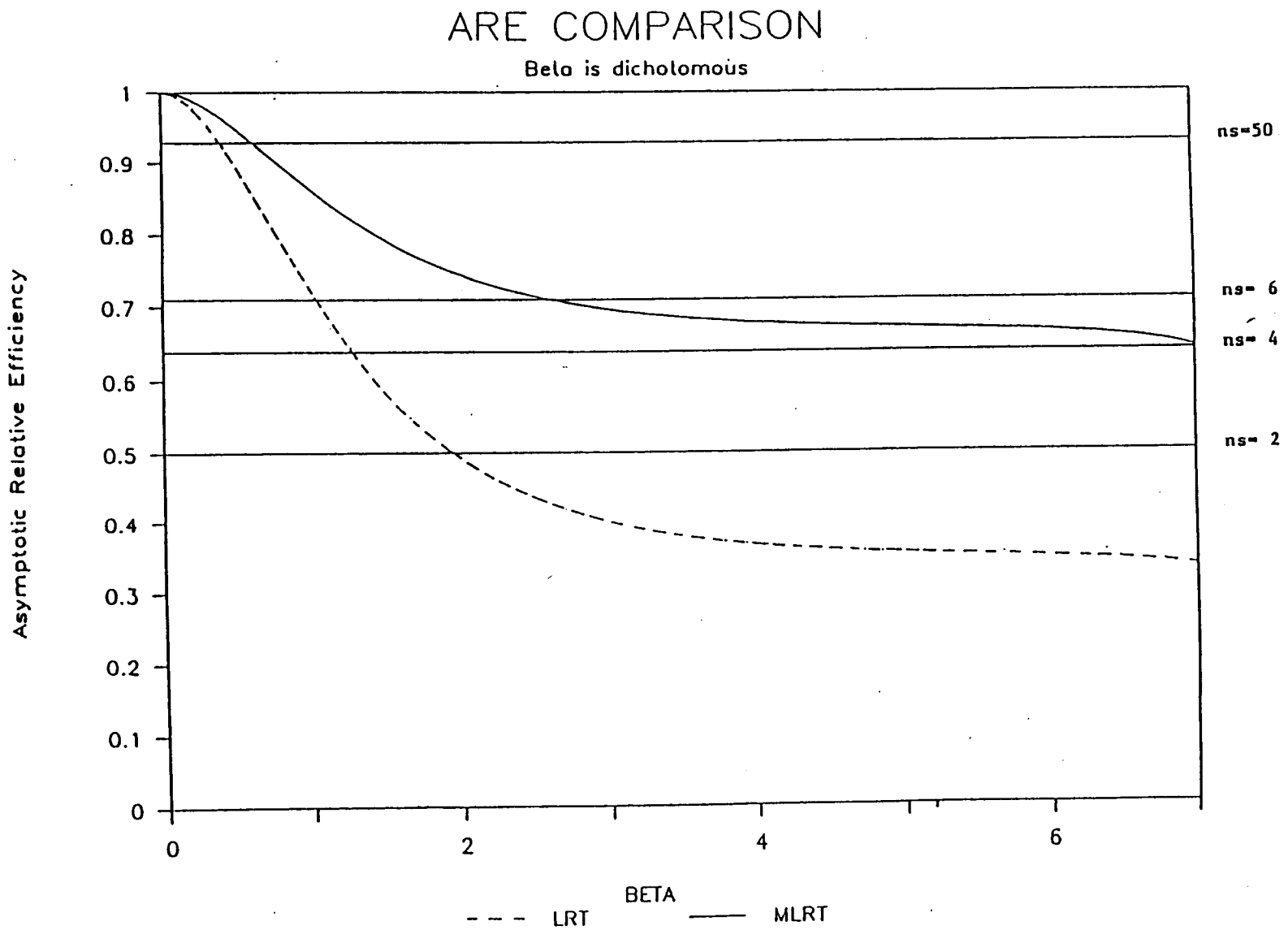

Figure 3.10: The ARE of the LRT, the SLRT and the MLRT as a Function of Stratum Effect, β, and Stratum Size, ns; β is dichotomously distributed

the same as the LRT as can be shown in equation (3.3.6). The variance of MLRT as in equation (3.3.5) can be shown to be 0.25 when β=0 (no stratum effect), but shrinks as the stratum effect gets larger.

As compared to the fully efficient model (the true model where the β's are known), the asymptotic efficiency is given by $\frac{4\cdot(\text{mean})^2}{\text{Var}}$. Figures 3.7, 3.8 and 3.9 give the ARE of each test compared to the correct model with known β's. The relative comparison of the three tests is shown in Figure 3.10. If there is no stratum effect, both the LRT and the MLRT are fully efficient, while the efficiency of the SLRT depends on the stratum size. When the stratum size is as large as 100, 80 and 50, the efficiencies are 0.96, 0.95 and 0.93 respectively. However as the stratum size gets smaller, the efficiency becomes smaller. When the stratum size is 2, the efficiency of the SLRT is only 0.5. As the stratum effect gets larger, the LRT becomes less efficient (e.g. for $\beta >$ 2, the LRT is less efficient than the SLRT even when the stratum size is only 2). The MLRT is always better than the LRT and better than the SLRT when the stratum size is 2 (i.e. matched pair design). When the stratum size is 4, the MLRT becomes less efficient than the SLRT as β, the stratum effect, is larger than 7. When the stratum size is 6, the MLRT is less efficient than the SLRT when the stratum effect, β, is greater than 3. When the stratum size is 50, the MLRT is less efficient than the SLRT when β is only 0.7.

3.4.2 *When beta is uniform(0,A)*

This section will consider the case as before with the number of stratum, $n_g \to \infty$, as the sample size, $n \to \infty$, but with the stratum parameter, β, uniformly distributed with range A. WLOG, assume $\beta \sim$ Uniform(0,A). In performing the numerical integrations to approximate the efficiencies, the density of the β's was approximated by $f(\beta)$=1/50, β=i*A/51, i=1 to 50.

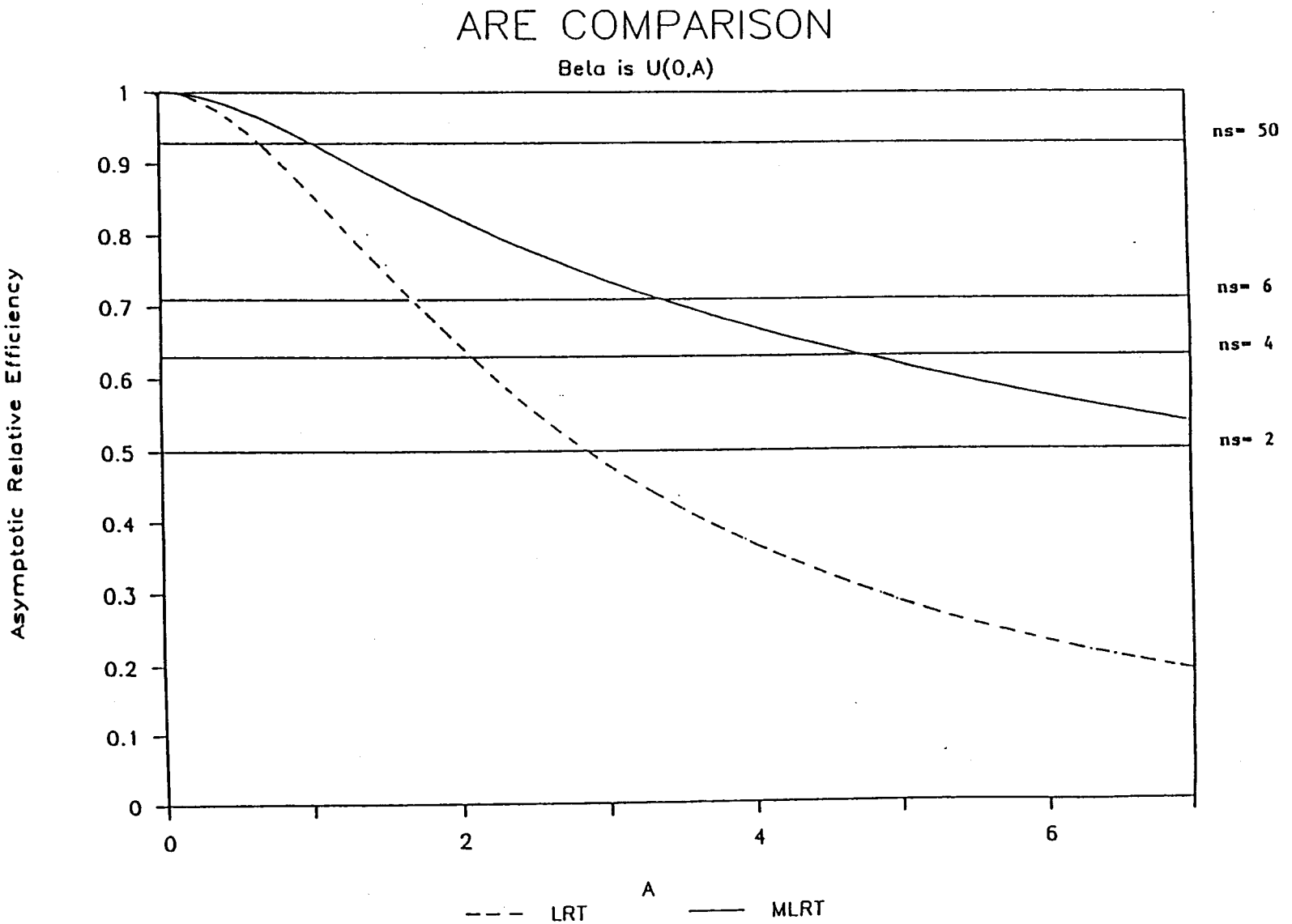

Figure 3.11: The Asymptotic Relative Efficiencies of the LRT, the SLRT and the MLRT as a Function of β, and Stratum Size, ns; β is Uniform (0, A)

To avoid redundancy, separate graphs for each of the tests are not presented, only the figure of the ARE comparisons is shown. Figure 3.11 shows the comparisons of the three tests based on the ARE's as functions of A when β is Uniform(0,A). Note that the ARE's of the SLRT are the same as when β is dichotomously distributed. Again, if there is no stratum effect (corresponding to when A=0), both the LRT and the MLRT are fully efficient, however as the heterogeneity of the stratum effect gets larger, these two tests become very inefficient. For matched pair design (i.e. n_s=2), the LRT is less efficient than the SLRT for A>2.9, for stratum size 4, the SLRT will be more efficient than the LRT for stratum effect, β, distributed as small as Uniform(0, 2.1). When the stratum size is 4, the MLRT is less efficient than the SLRT if A is greater than 4.8, etc. As can be seen from Figure 3.11 and in comparison with the graph given in Figure 3.10, the ARE of the LRT is larger for uniformly distributed β than dichotomous β, when β is less than 4. However when the stratum effect, β, is greater than 4, Figure 3.11 indicates the opposite. The ARE's of the MLRT also similarly crossed. In summary the rate of efficiency loss is faster when β is uniformly distributed than when β is dichotomous distributed.

3.4.3 *When beta is Normal(0, σ^2)*

As an example where the stratum effects have a continuous distribution with tails, consider the case where the distribution of β among strata is normally distributed. WLOG, assume β is normal(0, σ^2). Again, consider the number of strata $n_g \rightarrow \infty$ as the sample size $n \rightarrow \infty$, while the stratum size is fixed. Followed by the same notation given in section 3.1, the asymptotic mean and variance of the LRT become,

$$\text{Asy. mean: } \alpha\ p(1-p)\int g(t)\ d\ F_y(t) \qquad (3.4.1)$$

$$\text{Asy. Variance: } p(1-p)\int d\ F_y(t) \qquad (3.4.2)$$

where g(t) is defined in equation (2.1.4), and

$$d\,F_z(z) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{ - \frac{z^2}{2\sigma^2} \}. \tag{3.4.3}$$

In performing the numerical integrations to approximate the variance, the integral was restricted to $z \in (-3.5, +3.5)$, and the density $d\,F_z(z)$ is normalized and discretized at values $> \frac{j-1}{nk-1} - 3.5$, $(j=1, \cdots, nk)$, by

$$f_j = \frac{1}{\sqrt{2\pi}} \exp\{ - \frac{\{(j-1)/(nk-1)*7.0-3.5\}^2}{2} \},$$

where nk is the number of stratum used in this numerical evaluation. To numerically approximate the density of β, β is given by

$$\beta_j = \frac{(j-1)\cdot 7\cdot\sigma}{(nk-1)} - 3.5\sigma, \quad j=1 \text{ to } nk.$$

Since the SLRT does not depend on strata effects, the ARE of the SLRT is the same as before. For the MLRT, the asymptotic mean is the same as the LRT as given in equation (3.4.1) and the asymptotic variance is given by,

$$p\,(1-p)\sum_j \left\{ 2\int_0^\infty \left(- \log \sum_i f_i\, e^{-\Lambda_i(t)}\right) \frac{\sum_i f_i\, \lambda_i(t) e^{-\Lambda_i(t)}}{\sum_i f_i\, e^{-\Lambda_i(t)}}\, e^{-\Lambda_j(t)} dt \right.$$

$$\left. - \left[\int_0^\infty \frac{\sum_i f_i\, \lambda_i(t)\, e^{-\Lambda_i(t)}}{\sum_i f_i\, e^{-\Lambda_i(t)}}\, e^{-\Lambda_j(t)} dt \right]^2 \right\}. \tag{3.4.4}$$

where f_j is given in equation (3.4.3), $\lambda_i(t) = e^{\beta_i}$ and $\Lambda_j(t) = t\cdot e^{\beta_j}$.

Figure 3.12 illustrates the ARE's of the three tests when the stratum parameter, β, is normally distributed with mean zero and variance σ^2. As shown in the figure, the

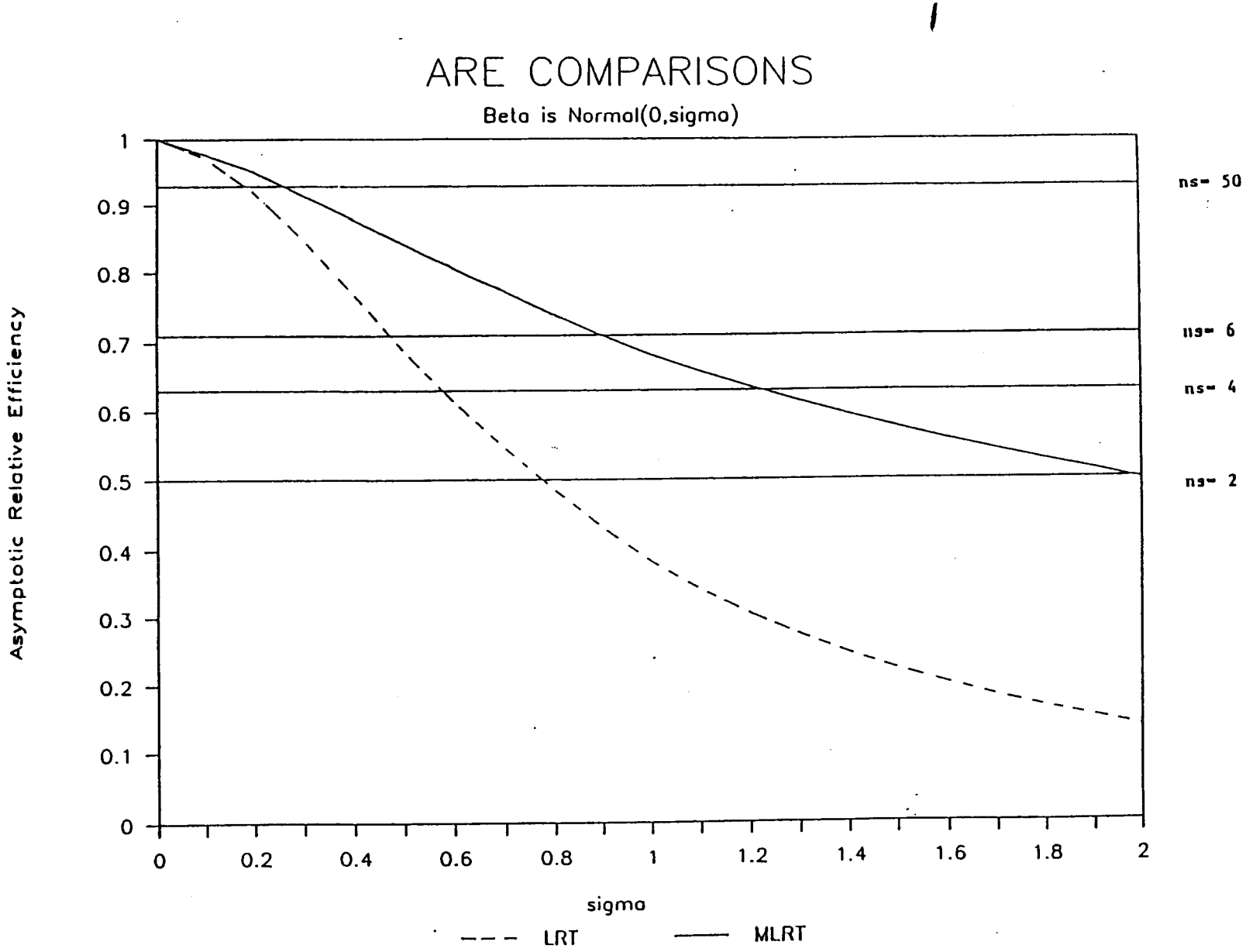

Figure 3.12: The Asymptotic Relative Efficiencies of the LRT, the SLRT and the MLRT as a Function of β, and Stratum Size, ns; β is Normal $(0, \sigma)$

LRT is less efficient than the SLRT when stratum size is 50 and σ is greater than 0.2. When the stratum size is 4 and 8, the LRT is less efficient than the SLRT when σ is greater than 0.5, and greater than 0.78 respectively. The MLRT is as efficient as the LRT when there is no stratum effect ($\sigma = 0$), and always more efficient than the LRT when there is a stratum effect. The MLRT is less efficient than the SLRT for σ is greater than 0.3, 1.25, and 2.0 for stratum sizes are 50, 4 and 2 respectively.

3.4.4 *Comparison with Schoenfeld and Tsiatis's simulation results*

The comparisons of the MLRT with the LRT and the SLRT by Schoenfeld and Tsiatis (1987) were done by simulation. In their simulation, the hazard rates were assumed distributed uniformly among the strata. That assumption corresponds to β being dichotomously distributed when n_g=2 or β uniformly distributed for other n_g's. With no censoring and a balanced design, the same table was generated using numerical integration to compare with their simulation results. Table 3.4 shows the comparison of their simulation results with the numerical evaluations.

The first column of Schoenfeld and Tsiatis's simulation results only gave the AREs of the MLRT vs the LRT when there is no stratum effect and the stratum size is 2. The value from simulation is 1.1, while theoretically the value is known to be 1.0 as was given in our numerical results column. The AREs of the MLRT vs the LRT is not given in their table since they argued that the use of the LRT is not valid. However, this argument is not true. The use of the LRT when there is known stratum effect will result in an inefficient test, nevertheless, it is still a "valid" test. Therefore, numerically those ARE's were computed and shown in column (3). As it was shown in column (3), the MLRT is always better than the LRT.

Comparing column (2) with column (4), it can be seen that the simulation

results are in general larger than the numerical results but with the same trend. In order to know more about the MLRT as opposed to the SLRT, the ARE's when strata effect is 10-fold and stratum sizes are 20 and 50 were also computed. For matched pair designs, the MLRT is better than the SLRT as was also shown in Figure(3.10). However, as the stratum size increases the gain of the MLRT over the SLRT is smaller. When the strata effect gets large, the MLRT becomes less efficient than the SLRT as was shown in column (4).

Table 3.3

No censoring: balanced design

		Simulation results		*Numerical results*	
Stratum size	Strata† effect	ARE vs LRT (1)	ARE vs SLRT (2)	ARE vs LRT (3)	ARE vs SLRT (4)
2	None	1.1	2.3	1.0	2.0
2	5X	-	2.0	1.2	1.73
2	10X	-	1.7	1.4	1.61
20	None	-	1.2	1.0	1.16
20	5X	-	0.98	1.32	0.95
20	10X	-	-	1.65	0.87
50	None	-	1.1	1.0	1.08
50	5X	-	0.88	1.4	0.84
50	10X	-	-	1.6	0.78

- indicate the values not given in Schoenfeld and Tsiatis's (1987) table.

† A 5X strata effect means that the ratio of the largest strata hazard to that of the smallest strata hazard rate is 5. Similarly for 10X.

Chapter IV

A MODIFIED SCORE TEST

For the testing of treatment effects on highly stratified survival data, three tests appear in the literature were studied, the LRT, the SLRT and the MLRT. The asymptotic distributions and efficiencies of these tests have been examined in the previous chapters. As discussed in Chapter 3, the LRT is inefficient in the case of heterogeneous populations. The SLRT is often used to control this heterogeneity. However, if the number of subjects in each stratum is small, such as in matched designs, the SLRT can be very inefficient. The recently proposed MLRT is fully efficient when there are no strata effects (like the LRT) but is less efficient than the SLRT for large strata effects. In this chapter, a new stratified test is proposed and the properties of this test will be explored. Section 4.1 presents the proposed test. Section 4.2 studies the properties of the proposed statistic when there are large strata effects. Section 4.3 investigates the properties of the test conditioned on the estimated strata effects. After deriving the properties of the test in Section 4.3, the unconditional distribution of the proposed test is given in Section 4.4, when there are no strata effects. An estimate for the variance of the test statistic is proposed in Section 4.5. Simulation results are given in Section 4.6. The distribution of the parameter estimates for strata effects is derived in Sections 4.7 - 4.9. The distribution of stratum effect estimates is derived in Section 4.7 for the case of large stratum sizes. In order to derive the distribution of stratum effect estimates for small stratum sizes, the expected value of these estimates conditioned on the death order of one of the member in the stratum is required and derived in Section 4.8. The distribution of stratum effect estimates is given in Section

4.9 using numerical results obtained from Section 4.8. All results will be summarized in Section 4.10.

§ 4. 1 *The Proposed Test*

The numerator of the proposed test is the score statistic of the proportional hazards model (PHM) for testing that the parameter of treatment effect, α, is equal to zero. Recall that the PHM has an indicator variable for treatment, x, and a set of indicator variables, $\underset{\sim}{z}$ (vector of n_g binary variables, with one constraint), for strata effects. The PHM is specified by the hazard relationship:

$$\lambda(t|x, \underset{\sim}{z}) = \lambda(t)\, e^{\alpha x+\underset{\sim}{\beta z}},$$

where $\lambda(t)$ is an arbitrary and unspecified base-line hazard function. The (conditional, marginal or partial) log likelihood is given by:

$$\mathcal{L} = \log L = \sum_{i=1}^{n} \delta_i \left(\alpha x+\underset{\sim}{\beta z} - \log \left(\sum_{j \in \Re_i} \exp(\alpha x_j+\underset{\sim}{\beta z}_j) \right) \right),$$

where δ_i=1 if the subject is uncensored, δ_i=0 otherwise. The first derivative of this log likelihood with respect to α and evaluated at α=0 is the score statistic:

$$\hat{S} = n^{-1/2} \frac{\partial \mathcal{L}}{\partial \alpha}\Big|_{\alpha=0} = n^{-1/2} \sum_i \delta_i \left[x_i - \frac{\sum_{\Re_i} x_j\, e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}}{\sum_{\Re_i} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}} \right], \qquad (4.1.1)$$

where $\underset{\sim}{\beta}$ is the parameter vector of strata effects. WLOG, assume all strata have same number of subjects, i.e. $n_{s_1} = n_{s_2} = \cdots = n_s$. If there is no censoring, $\hat{\underset{\sim}{\beta}}$ is the solution

to:

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n}\left[z_{1i} - \frac{\sum_{j\in \Re_i} z_{1j}\, e^{\hat{\beta} z_j}}{\sum_{j\in \Re_i} e^{\hat{\beta} z_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{1i}\, e^{\hat{\beta}_1}}{\sum_{j=1}^{n_g} n_{ji}\, e^{\hat{\beta}_j}} = 0$$

$$\frac{\partial \ell}{\partial \beta_2} = \sum_{i=1}^{n}\left[z_{2i} - \frac{\sum_{j\in \Re_i} z_{2j}\, e^{\hat{\beta} z_j}}{\sum_{j\in \Re_i} e^{\hat{\beta} z_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{2i}\, e^{\hat{\beta}_2}}{\sum_{j=1}^{n_g} n_{ji}\, e^{\hat{\beta}_j}} = 0$$

$$\vdots$$

$$\frac{\partial \ell}{\partial \beta_{n_g}} = \sum_{i=1}^{n}\left[z_{n_g i} - \frac{\sum_{\Re_i} z_{n_g j}\, e^{\hat{\beta} z_j}}{\sum_{\Re_i} e^{\hat{\beta} z_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{n_g i}\, e^{\hat{\beta}_{n_g}}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0.$$

The denominator of the proposed test is a consistent estimates of the variance of $\hat{S}$ which will be discussed later in Section 4.5. The proposed test - modified score test (MST) is given by

$$\text{MST} = \frac{\hat{S}^2}{\hat{\text{V}}\text{ar}(\hat{S})}\,. \tag{4.1.2}$$

§ 4.2 *Properties of the PH Score Statistic for "Large" Strata Effects*

This section will study the numerator of the MST, $\hat{S}$, when there are large strata effects. The definition of "large" strata effects is as follows. Let R_{ji} be the hazard ratio of stratum j to stratum i, in which R_{ji} can be written as

$$R_{ji} = \frac{\lambda_j(t)}{\lambda_i(t)} = e^{\beta_{j,i}}. \tag{4.2.1}$$

Large strata effects can be expressed as the limit as $R_{ji} \to \infty$, $\forall$ $j < i$. The following lemma is an illustration of the above definition in detail.

Lemma 4.1. Let $t_1, t_2, t_3, t_4, \cdots, t_n$ be the ordered death times, and let $g_1, g_2, g_3, \cdots, g_n$ denote the individual stratum indicators. WLOG, assume a matched pair design, i.e. $n_s=\frac{n}{n_g}=2$. If there are "large strata effects", then the order of death times will follow sequence described below with probability = 1.

death order (t_i):	t_1	t_2	t_3	t_4	t_5	t_6	$\cdots$	t_{n-1}	t_n	
strata (g_i):	1	1	2	2	3	3	$\cdots$	n_g	n_g	(4.2.2)

Proof: Let n_{ij} denote the number of patients at risk in stratum j at the i^{th} death time. Based on the relationship between the marginal likelihood and probability of rank order of death times,

Pr(member of stratum $\mathfrak{g}$ at risk will be the next to die | those at risk)

$$= \frac{\sum_{l \in \mathfrak{g}} \lambda_l(t)}{\sum_{l \in \mathfrak{R}_i} \lambda_l(t)},$$

and

Pr(member of stratum j will be the next to die at time i | those at risk)

$$= \frac{n_{ij}\, e^{\beta_j}}{\sum_l n_{il}\, e^{\beta_l}} = \frac{n_{ij}}{\sum_l n_{il}\, e^{\beta_l - \beta_j}} = \frac{n_{ij}}{\sum_l n_{il}\, R_{lj}}$$

$$= \begin{cases} \frac{n_{ij}}{n_{ij}} = 1 & \text{if } n_{il} = 0,\ \forall\ l<j \\ 0 & \text{if there is an } l<j \text{ such that } n_{il} > 0\ . \end{cases}$$

Therefore, the next subject to die will be in the smallest numbered stratum among those subjects at risk.

□

Lemma 4.2. If the death times are ordered by strata as in (4.2.2), then the likelihood is maximized by $\hat{\beta}_{j,j+1} \to \infty$, $\forall\ j$.

Proof: The first derivative of the log likelihood with respect to the stratum parameters, $\underset{\sim}{\beta}$, at $\alpha=0$, are,

$$\frac{\partial}{\partial \beta_1} \log L = \sum_{i=1}^{n} \left[z_{1i} - \frac{\sum_{j \in \Re_i} z_{1j}\, e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{1i}\, e^{\beta_1}}{\sum_{j=1}^{n_g} n_{ji}\, e^{\beta_j}}$$

$$\frac{\partial}{\partial \beta_2} \log L = \sum_{i=1}^{n} \left[z_{2i} - \frac{\sum_{j \in \Re_i} z_{2j}\, e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{2i}\, e^{\beta_2}}{\sum_{j=1}^{n_g} n_{ji}\, e^{\beta_j}}$$

$$\vdots$$

$$\frac{\partial}{\partial \beta_{n_g}} \log L = \sum_{i=1}^{n} \left[z_{n_g} \frac{\sum_{j \in \Re_i} z_{n_g j}\, e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}} \right] = n_s - \sum_{i=1}^{n} \frac{n_{n_g i}\, e^{\beta_{n_g}}}{\sum_{j=1}^{n_g} n_{ji} e^{\beta_j}}$$

and the score statistic is

$$\frac{\partial}{\partial \alpha} \log L = \sum_i \left[x_i - \frac{\sum_{\Re_i} x_j\, e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}}{\sum_{\Re_i} e^{\underset{\sim}{\beta} \underset{\sim}{z}_j}} \right].$$

From equation (4.2.1), $\hat{\beta}_{j,j+1}$ is the estimated hazard ratio of stratum j and j+1. If the death times are ordered by strata as in equation (4.2.2) and if $\hat{\beta}_{j,j+1} \to \infty$

(i.e. $e^{\hat{\beta}_{j,i}} \rightarrow \infty$, $\forall\, j < i$), the first derivative of the log likelihood will be 0, with second derivative $\leq$ 0, the log likelihood is maximized. □

Lemma 4.1 and Lemma 4.2 are used to derive the following Theorem 4.1 which shows the property of the proposed test when there are large strata effects.

Theorem 4.1. If there are large strata effects as shown in equation (4.2.2), the score statistic is equivalent to the stratified log rank test (SLRT).

Proof: Using Lemma 4.1 and Lemma 4.2, large strata effects imply the strict ordering of death times as in equation (4.2.2) with probability 1. In turn, this implies $\hat{\beta}_{j,j+1} \rightarrow \infty$. As $\hat{\beta}_{j,j+1} \rightarrow \infty$, the score statistic can be written as:

$$\hat{S} = n^{-1/2}\sum_i \left[x_i - \frac{\sum_{\Re_i} x_j\, e^{\hat{\beta}_j}}{\sum_{\Re_i} e^{\hat{\beta}_j}} \right] \rightarrow n^{-1/2}\sum_i \left[x_i - \frac{\sum_{\Re_i} x_j\, u_{ij}}{\sum_{\Re_i} u_{ij}} \right]$$

$$\rightarrow\ n^{-1/2}\sum_j\sum_i (a_{ij} - \frac{n_{Aij}}{n_{ij}}) \ \equiv \ SLRT_{NUM}$$

where,

$$u_{ij} = \begin{cases} 1 & \text{if } i^{th} \text{ and } j^{th} \text{ patients belong to the same stratum} \\ 0 & \text{otherwise} \end{cases}$$

n_{ij} = # of patients at risk at time i in stratum j

n_{Aij}= # of patients, group A, at risk at time i in stratum j

$$a_{ij} = \begin{cases} 1 & \text{death at time i, stratum j, group A} \\ 0 & \text{death at time i, stratum j, group B .} \end{cases}$$

□

As the strata effects get larger, $\hat{S}$ converges to the numerator of the SLRT. For the extreme case of "large strata effects", the numerator of the proposed test, $\hat{S}$, is equivalent to the numerator of the SLRT.

§ 4.3 *Properties of PH Score Statistic Conditioned on Estimated Stratum Effects*

This section considers the properties of the proposed statistic conditioned on the values of the estimated stratum effects. The derivation of the distribution of $\hat{S}$ based on equation (4.1.1) conditioned on the observed $\hat{\beta}$'s is given below. Under H_0: $\alpha=0$, the score statistic of PHM with the true indicator stratum parameters can be written as:

$$S = S(\underset{\sim}{\beta}\underset{\sim}{z}).$$

The mean and variance of $S(\hat{\underset{\sim}{\beta}}\underset{\sim}{z})$, when $\hat{\beta}$ is considered a fixed constant, can be obtained by extending results on the mean and variance of $S(\underset{\sim}{\beta}\underset{\sim}{z})$ when omitting a subset of covariates. Assume the proportion of patients in each treatment is the same for each stratum (i.e. $p_j = p$, $\forall$ j or balanced design) and $p = \frac{1}{2}$. Let the treatment variable take value $\pm\frac{1}{2}$, then,

$$n^{-1/2} S \sim N(0, \tfrac{1}{4}).$$

Let

$$U(t_i, y) = \frac{\sum_{\Re_i} x_j e^{y_j}}{\sum_{\Re_i} e^{y_j}} \qquad i = 1, \cdots, n. \tag{4.3.1a}$$

and

$$V_0(t, y) = \int e^{y} e^{\{-\Lambda(t) e^{\underset{\sim}{\beta}\underset{\sim}{z}}\}} d F_Z(z) \tag{4.3.1b}$$

$$V_1(t, y) = \int e^y \, e^{\{-\Lambda(t)\, e^{\alpha/2+\underset{\sim}{\beta}\underset{\sim}{z}}\}} \, d\, F_Z(z) \qquad (4.3.1c)$$

$$V_2(t, y) = \int e^y \, e^{\{-\Lambda(t)\, e^{-\alpha/2+\underset{\sim}{\beta}\underset{\sim}{z}}\}} \, d\, F_Z(z)\,. \qquad (4.3.1d)$$

Let α be the true parameter for treatment effect, $\hat{\alpha}$, the parameter estimate. When $\hat{\underset{\sim}{\beta}}\underset{\sim}{z}$ is used in the model rather than $\underset{\sim}{\beta}\underset{\sim}{z}$, the parameter estimate for x will be $\hat{\alpha}^*$ which is the solution of the (incorrectly specified) log likelihood equation:

$$0 = \sum_i \left[x_i - \frac{\sum_{\Re_i} x_j \, e^{\hat{\alpha}^* x_j + \hat{\underset{\sim}{\beta}}\underset{\sim}{z}_j}}{\sum_{\Re_i} e^{\hat{\alpha}^* x_j + \hat{\underset{\sim}{\beta}}\underset{\sim}{z}_j}} \right] \qquad (4.3.2)$$

Using methods similar to Morgan (1988, Appendix) on omitting a subset of covariates, $\hat{S}$ can be re-expressed as:

$$n^{-1/2}\, \hat{S} = n^{-1/2} \sum \left[x_i - U(t_i, \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) \right] = S_1 + S_2 + S_3$$

where,

$$S_1 = n^{-1/2} \sum \left[x_i - U(t_i, \alpha x + \underset{\sim}{\beta}\underset{\sim}{z}) \right]$$

$$S_2 = n^{-1/2} \sum \left[U(t_i, \alpha x + \underset{\sim}{\beta}\underset{\sim}{z}) - U(t_i, \alpha^* x + \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) \right]$$

$$S_3 = n^{-1/2} \sum \left[U(t_i, \alpha^* x + \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) - U(t_i, \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) \right].$$

S_1 is the log likelihood equation for α under the true model. Furthermore, assuming the β's are bounded, S_1 is asymptotically normally distributed with mean 0 and variance p(1-p). If $p=\frac{1}{2}$, then, $S_1 \sim N(0, \frac{1}{4})$. S_2 can be re-expressed as $S_2 = S_{21}+S_{22}+S_{23}$,

where,

$$S_{21} = n^{-1/2}\sum\left[U(t_i, \alpha x+\underset{\sim}{\beta}\underset{\sim}{z}) - U(t_i, \hat{\alpha} x+\underset{\sim}{\beta}\underset{\sim}{z})\right]$$

$$S_{22} = n^{-1/2}\sum\left[U(t_i, \hat{\alpha} x+\underset{\sim}{\beta}\underset{\sim}{z}) - U(t_i, \hat{\alpha}^* x+\underset{\sim}{\hat{\beta}}\underset{\sim}{z})\right]$$

$$S_{23} = n^{-1/2}\sum\left[U(t_i, \hat{\alpha}^* x+\underset{\sim}{\hat{\beta}}\underset{\sim}{z}) - U(t_i, \alpha^* x+\underset{\sim}{\hat{\beta}}\underset{\sim}{z})\right].$$

The second term, S_{22}, equals 0 by definition of $\hat{\alpha}$ and $\hat{\alpha}^*$ being the solutions to the log likelihood equations ($\sum x_i = \sum U(t_i, \hat{\alpha} x+\underset{\sim}{\beta}\underset{\sim}{z})$ and $\sum x_i = \sum U(t_i, \hat{\alpha}^* x+\underset{\sim}{\hat{\beta}}\underset{\sim}{z})$). The first term, S_{21}, converges in probability to zero (Schoenfeld, 1983). The third term, S_{23}, also converges in probability to zero (Morgan, 1988). S_3 can be expanded in a first order Taylor series about $\alpha = 0$. Note that $\alpha^* = 0$ when $\alpha = 0$ ($\alpha^* = \alpha^*(\alpha)$, $\alpha^*(0) = 0$). Then,

$$S_3(\alpha) = S_3(0) + \alpha S_3'(0),$$

where

$$S_3(0) = 0$$

$$S_3'(0) = n^{-1/2}\sum\left[\frac{\sum_{\Re_i} x_j x_j \left(\frac{d\alpha^*}{d\alpha}\right) e^{\underset{\sim}{\hat{\beta}}\underset{\sim}{z}}}{\sum_{\Re_i} e^{\underset{\sim}{\hat{\beta}}\underset{\sim}{z}}}\right]\Bigg|_{\alpha=0}.$$

If $x = \pm\frac{1}{2}$, S_3 can be written as,

$$S_3 = n^{-1/2}\alpha\sum \tfrac{1}{4}\left(\frac{d\alpha^*}{d\alpha}\right)\Big|_{\alpha=0}.$$

For $\alpha \to n^{-1/2}$, as $n \to \infty$ (i.e. $\alpha = O(n^{-1/2})$),

$$\sum_{i=1}^{n} n^{-1/2}\,\alpha \to O(1) \quad \text{and} \quad S_3 \to \alpha\, n^{1/2}\left(\frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}\right)\frac{1}{4}.$$

Note that the hazard at time T=t is given by

$$\lambda(t|\, x, z) = \lambda(t)\, e^{\alpha x + \underset{\sim}{\beta} \underset{\sim}{z}}.$$

Assume X is independent of Z (e.g. all patients are randomly allocated to two treatment groups with equal probability), then

$$f_X(x) = \tfrac{1}{2} \qquad x = \tfrac{1}{2}, -\tfrac{1}{2}$$

$$d\, F_{XZ} = \tfrac{1}{2}\, d\, F_Z$$

$$d\, F_{T|XZ}(t \mid x, z) = \lambda(t)\, e^{\alpha x + \underset{\sim}{\beta} \underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\alpha x + \beta z}}\, dt$$

$$\begin{aligned} d\, F_{TXZ}(t, x, z) &= d\, F_{T|XZ}(t \mid x, z) \cdot d\, F_{XZ} \\ &= \tfrac{1}{2}\, \lambda(t)\, e^{\alpha x + \underset{\sim}{\beta} \underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\alpha x + \beta z}}\, d\, F_Z\,. \end{aligned}$$

In order to evaluate the asymptotic limit of $\left.\frac{d\alpha^*}{d\alpha}\right|_{\alpha=0}$ conditioned on $\hat{\underset{\sim}{\beta}}\underset{\sim}{z}$ and z, the distribution of $\hat{\underset{\sim}{\beta}}$ and the conditional distribution of T given x, z and $\hat{\underset{\sim}{\beta}}$ is required. As the number within strata, ns, becomes large, $\hat{\underset{\sim}{\beta}}$ converges to $\underset{\sim}{\beta}$ and the conditional distribution of T give X, Z and $\hat{\underset{\sim}{\beta}}$ converges to $d\, F_{T|XZ}$ given above. However, for small values of ns, the distribution of T conditioned on $\hat{\underset{\sim}{\beta}}$ is dependent on the value of $\hat{\underset{\sim}{\beta}}$. For example, in the very extreme case of ns=1, $\hat{\underset{\sim}{\beta}}$ will be composed of estimates that are different orders of infinity and $\hat{\underset{\sim}{\beta}}$ uniquely determines the rank order of the t_i's. The distribution of T given $\hat{\underset{\sim}{\beta}}$, which is a complex implicit function of the observed t_i's, can not be explicitely obtained; therefore, the efficiency of the score test for small ns will be determined by numberical simulation in Section 4.6. The following lemma is used to evaluate $\left.\frac{d\alpha^*}{d\alpha}\right|_{\alpha=0}$ for large strata effects.

Lemma 4.3. Let $\dot{\alpha}^* = \frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}$ (4.3.3)

then as $ns \to \infty$

$$\dot{\alpha}^* = \int_t \frac{\lambda(t)\Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z})}\, dt.$$

Proof: By definition, α^*, the parameter value, is the solution to the asymptotic limit of the log likelihood equation (4.3.2) where the asymptotic limit is given by:

$$0 = \int_t\int_x\int_z \{ x\, d\, F_{TXZ}(t,x,z)\} - \int_t\int_x\int_z \{ d\, F_{T|XZ}(t,x,z)\, dF_{XZ}(x,z)\}$$

$$\cdot \frac{\int_u\int_x\int_z \{ x\, e^{\alpha^* x+\hat{\underset{\sim}{\beta}}\underset{\sim}{z}} d\, F_{TXZ}(-u,x,z)\}\, I(u>t)\}}{\int_u\int_x\int_z \{ e^{\alpha^* x+\hat{\underset{\sim}{\beta}}\underset{\sim}{z}} d\, F_{TXZ}(-u,x,z)\}\, I(u>t)\}},$$

when $x = \frac{1}{2}, -\frac{1}{2}$ this can be written as:

$$0 = \int_t \tfrac{1}{2}\lambda(t)\, \{ V_1(t, \tfrac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z}) - V_2(t, -\tfrac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z})\}$$

$$- \lambda(t)\, \{V_1(t, \tfrac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z}) + V_2(t, -\tfrac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z})\} \cdot \frac{\tfrac{1}{2}\{ V_1(t, \tfrac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z}) - V_2(t, -\tfrac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})\}}{\{V_1(t, \tfrac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z}) + V_2(t, -\tfrac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})\}},$$

where $V_0(t, y)$, $V_1(t, y)$ and $V_2(t, y)$ are defined in equations (4.3.1b, c, d). Differentiating w.r.t. α and noting that $\alpha^* = 0$ when $\alpha = 0$, thus,

$$\frac{\partial\, V_1(t, \frac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z})}{\partial\alpha}\Big|_{\alpha=0} = \tfrac{1}{2}\{V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z})\} = -\frac{\partial\, V_2(t, -\frac{\alpha}{2}+\underset{\sim}{\beta}\underset{\sim}{z})}{\partial\alpha}\Big|_{\alpha=0},$$

$$\frac{\partial\, V_1(t, \frac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})}{\partial\alpha}\Big|_{\alpha=0} = \tfrac{1}{2}\{\frac{\partial\alpha^*}{\partial\alpha}\, V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})\} = -\frac{\partial\, V_2(t, -\frac{\alpha^*}{2}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})}{\partial\alpha}\Big|_{\alpha=0},$$

using equation (4.3.3) then,

$$0 = \int_t \Big\{ \tfrac{1}{4}\lambda(t)\Big(V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z}) + V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z})\Big)$$

$$- \tfrac{1}{2}\lambda(t)\Big(V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z}) - V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) + \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z})\Big) \cdot 0$$

$$- \tfrac{1}{2}\,\lambda(t)\,\frac{V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z})} \cdot \Big(\dot{\alpha}^*\, V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z} + \hat{\underset{\sim}{\beta}}\underset{\sim}{z})\Big)$$

$$+ \lambda(t)\,\frac{V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{4V_0^2(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z})} \cdot 0 \cdot 0 \Big\}\, dt$$

$$= \int_t \tfrac{1}{2}\,\lambda(t)\,\{\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) - \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z}) + \frac{\Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z} + \hat{\underset{\sim}{\beta}}\underset{\sim}{z})\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z})} \,\}\, dt$$

$$- \tfrac{1}{2}\,\dot{\alpha}^* \int_t \lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})\, dt.$$

Note that,

$$\begin{aligned}
\int_t \lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})\, dt &= \int_t \lambda(t) \int e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, d\, F_Z(z)\, dt \\
&= -\int_z \int_t \{-\lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, dt\, d\, F_Z(z) \\
&= -\int_z [\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\,]\,|_0^\infty\, d\, F_Z(z) \\
&= 1 - V_0(\infty, 0) \\
&= 1\,,
\end{aligned}$$

and

$$\int_t \lambda(t)\,\Lambda(t)\,V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z}) = \int_t \lambda(t)\,\Lambda(t) \int e^{2\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, d\,F_Z(z)\, dt$$

$$= \int_z \int_t \{\,\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\} \, d\, \{-\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\} \, d\, F_Z(z)$$

$$= \int_z \{ \int_t \lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}} - \Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}} \} \,|_0^\infty dF_Z(z)$$

$$= \left[\, V_0(t, 0) - \Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}) \,\right] |_0^\infty$$

$$= 1\,.$$

From the above,

$$0 = \tfrac{1}{2} \int_t \{ \frac{\lambda(t)\,\Lambda(t)\,V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}+\underset{\sim}{\hat{\beta}}\underset{\sim}{z})\,V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \underset{\sim}{\hat{\beta}}\underset{\sim}{z})} \}\, dt - \tfrac{1}{2}\,\dot{\alpha}^*$$

$\Rightarrow$

$$\dot{\alpha}^* = \frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0} = \int_t \frac{\lambda(t)\Lambda(t)\,V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}+\underset{\sim}{\hat{\beta}}\underset{\sim}{z})\,V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \underset{\sim}{\hat{\beta}}\underset{\sim}{z})}\, dt. \qquad (4.3.4)$$

Several special cases of Lemma 4.5 deserve consideration. Consider the case of no censoring. Since $n^{-1/2}\,\hat{S} = S_1+S_2+S_3$, where S_1 is asymptotically normal with mean 0 and variance $\frac{1}{4}$, S_2 converges to 0, and S_3 converges to $\dot{\alpha}^*\frac{n^{1/2}\alpha}{4}$. Thus, $\hat{S}$ is asymptotically normal with mean $\dot{\alpha}^*\frac{n^{1/2}\alpha}{4}$ and variance $\frac{1}{4}$. That is, the score statistic, as shown in equation (4.1.1), when conditioned on the $\hat{\beta}$'s is asymptotically normally distributed with mean $\dot{\alpha}^*\frac{n^{1/2}\alpha}{4}$ and variance $\frac{1}{4}$, where $\dot{\alpha}^*$ is given in equation (4.3.4). Note that the above derivation was under the assumption of the weighted CLT, which is according to Liapunov (Rao, 1973, Proposition 1.4.3). This assumption may be written formally as follows. Let $S_n = Y_1 + Y_2 + \cdots + Y_n$, and $s_n^2 = \sigma_1{}^2 + \sigma_2{}^2 + \cdots + \sigma_n{}^2$, where $\sigma_n{}^2 = E(X_n - \mu_n)^2 \neq 0$. If $EX_n = \mu_n$, and $Y_n = X_n - \mu_n$, and if

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} E\,|Y_j|^{2+\delta} \to 0$$

for some $\delta > 0$, then

$$\frac{S_n}{s_n} \rightarrow N(0, 1).$$

Here, $X_1 = n_{1i}\hat{K}_1, \cdots, X_n = n_{n_g i}\hat{K}_{n_g}$, n_{j_i} is fixed and $\frac{n_{j_i}}{\sum_j n_{ji}} \rightarrow 0$, $\forall$ j, therefore, CLT holds if none of the $\hat{K}_i$ are dominating, for $Y_j = n_{ji}(\hat{K}_j - \bar{K})$, that is

$$\text{if } \frac{1}{s_{n_g}^{2+\delta}} \sum_{j=1}^{n_g} E\,|Y_j|^{2+\delta} \rightarrow 0. \tag{4.3.5}$$

If some of the stratum effect estimates, $\hat{K}_i$, are dominating, which case the weighted CLT won't hold, then the asymptotic variance of $\hat{S}$, the proposed statistic, is no longer a constant. The asymptotic variance of $\hat{S}$ derived below is for more general cases. Let $l(i)$ be the stratum to which the i^{th} subjects belongs (i^{th} ordered death), and $\hat{K}_{(i)} = \hat{K}_{l(i)}$ be the stratum effect for the stratum to which the i^{th} subject belongs, then the score statistic, if conditioned on the estimated stratum effect, $\hat{\beta}$, is given by:

$$n^{-1/2}\,\hat{S}|\,\hat{\beta} = n^{-1/2}\sum_i \left(x_i - \frac{\sum_{\Re_i} x_j\, e^{\hat{\beta}_j}}{\sum_{\Re_i} e^{\hat{\beta}_j}}\right)$$

$$= n^{-1/2}\sum_i \left(x_i - \frac{\sum_{\Re_i} x_j\, \hat{K}_j}{\sum_{\Re_i} \hat{K}_j}\right)$$

$$= n^{-1/2}\left\{ x_1 - \frac{x_1\hat{K}_{(1)} + \cdots + x_n\hat{K}_{(n)}}{\hat{K}_{(1)} + \cdots + \hat{K}_{(n)}} \right.$$

$$+ x_2 - \frac{x_2\hat{K}_{(2)} + \cdots + x_n\hat{K}_{(n)}}{\hat{K}_{(2)} + \cdots + \hat{K}_{(n)}}$$

$$\vdots$$

$$+ x_i - \frac{x_i\hat{K}_{(i)} + \cdots + x_n\hat{K}_{(n)}}{\hat{K}_{(i)} + \cdots + \hat{K}_{(n)}}$$

$$\vdots$$

$$\left. + x_n - \frac{x_n\hat{K}_{(n)}}{\hat{K}_{(n)}} \right\}$$

$$= n^{-1/2}\Big\{ x_1 - \frac{x_1\hat{K}_{(1)}}{\hat{K}_{(1)}+\cdots+\hat{K}_{(n)}}$$

$$+ x_2 - x_2\hat{K}_{(2)}\Big(\frac{1}{\hat{K}_{(1)}+\cdots+\hat{K}_{(n)}} + \frac{1}{\hat{K}_{(2)}+\cdots+\hat{K}_{(n)}}\Big)$$

$$\vdots$$

$$+ x_i - x_i\hat{K}_{(i)}\Big(\frac{1}{\hat{K}_{(1)}+\cdots+\hat{K}_{(n)}} + \cdots + \frac{1}{\hat{K}_{(i)}+\cdots+\hat{K}_{(n)}}\Big)$$

$$+ x_n - x_n\hat{K}_{(n)}\Big(\frac{1}{\sum_{l=1}^{n}\hat{K}_{(l)}} + \cdots + \frac{1}{\sum_{l=i}^{n}\hat{K}_{(l)}} + \cdots + \frac{1}{\hat{K}_{(n)}}\Big)\Big\}$$

$$= n^{-1/2}\sum_{i=1}^{n} x_i\Big(1 - \hat{K}_{(i)}\sum_{j=1}^{i}\frac{1}{\sum_{l\in\Re_j}\hat{K}_{(l)}}\Big)$$

$$= n^{-1/2}\sum_{i=1}^{n} x_i\,\mathcal{W}_i\,,$$

where

$$\mathcal{W}_i = \Big(1 - \hat{K}_{(i)}\sum_{j=1}^{i}\frac{1}{\sum_{l\in\Re_j}\hat{K}_{(l)}}\Big).$$

Note that $\mathcal{W}_i$ are not all positive, and $\sum_{i=1}^{n}\mathcal{W}_i = 0$. The variance of $\hat{S}|\hat{\beta}$ is given by

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = \sum_{i=1}^{n}\mathcal{W}_i^2\,\mathrm{Var}(x_i) + \sum_{i}\sum_{i\neq j}\sum_{j}\mathcal{W}_i\mathcal{W}_j\,\mathrm{Cov}\,(x_i, x_j)$$

If the trial is balanced within stratum as is the case in matched pair design, it can be shown that

$\mathrm{Cov}\,(x_i, x_j) = 0$ if subject i, j are not in the same stratum,

$\mathrm{Cov}\,(x_i, x_j) = -\frac{p(1-p)}{(n_s-1)}$ if subject i and j are in the same stratum.

From above,

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = p(1-p)\sum_{i=1}^{n}\mathcal{W}_i^2 \; - \frac{p(1-p)}{(n_s-1)}\sum_{i^*}\sum_{j^*\neq i^*}\mathcal{W}_{i^*}\mathcal{W}_{j^*}, \qquad (4.3.6a)$$

where i^* and j^* are in the same stratum.

If the treatments are unbalanced within stratum, it can be shown that

$$\mathrm{Cov}(x_i, x_j) = 0, \ \forall \ i, j.$$

Thus,

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = \frac{p(1-p)}{n} \sum_{i=1}^{n} w_i^2 \, . \tag{4.3.6b}$$

In general, the statistic, $\hat{S}$, as given in equation (4.1.1), when conditioned upon estimated stratum effect, will be normally distributed with mean $\mathring{\alpha}^* \frac{n^{1/2}\alpha}{4}$, where $\mathring{\alpha}^*$ is given by equation (4.3.4) and variance given in equations (4.3.6a,b). If none of the $\hat{K}_i$ are dominating, that is, equation (4.3.5) holds, then equation (4.3.6a,b), ignoring o(1) terms, can be shown to converge to $\frac{1}{4}$ as $n \to \infty$.

§ 4.4 *Asymptotic Distribution of the Numerator of the Proposed Test*

The conditional distribution of the numerator of the proposed test, $\hat{S}$, given $\hat{\beta}$, the estimated stratum effect is given in Section 4.3. In this section, the unconditional distribution of $\hat{S}$ is discussed. Recall from Section 4.2, when there are large stratum effects, which corresponds to all $\hat{K}$'s diverging, the numerator of the proposed test $\hat{S}$ will converge to the numerator of the SLRT (Theorem 4.1). Thus, the situation of "large stratum effect" can be thought of as the limiting case of strata effects being uniformly distributed within infinite intervals.

In the case where the probability that any of the $\hat{K}$'s diverging is 0, i.e. the weighted CLT holds, the conditional distribution of $\hat{S}$ given $\hat{\beta}$ is normal with mean $\mathring{\alpha}^* \frac{n^{1/2}\alpha}{4}$ and variance $\frac{1}{4}$. If there are no strata effects, i.e. $\underset{\sim}{\beta} = 0$, it can be shown that $\mathring{\alpha}^* = 1$ and the distribution of $\hat{S}$ given $\hat{\beta}$ does not depend on the estimated strata effects. Thus, the distribution of $\hat{S}$ is normal with mean $= \frac{n^{1/2}\alpha}{4}$ and variance 1/4. Note that this is the same as the LRT with no strata effects. Under the null hypothesis of no treatment effects, the mean and variance of the unconditional distribution of $\hat{S}$ can be derived by:

$$\mathbf{E}\,\hat{S} = \underset{\hat{\beta}}{\mathbf{E}}\,\mathbf{E}(\,\hat{S}|\,\hat{\beta}) = 0, \tag{4.4.1}$$

and the variance of $\hat{S}$ can be derived by:

$$\hat{S} = \underset{\hat{\beta}}{\mathbf{E}}\,\mathrm{Var}(\hat{S}|\hat{\beta}) + \underset{\hat{\beta}}{\mathrm{Var}}\,\mathbf{E}(\,\hat{S}|\hat{\beta}) = \underset{\hat{\beta}}{\mathbf{E}}\,\mathrm{Var}(\hat{S}|\hat{\beta}), \tag{4.4.2}$$

using the distribution of the estimated stratum effect which will be discussed later.

As the stratum size gets larger, the probability that any $\hat{K}_j$, the stratum effect estimate, will diverge becomes smaller. This is the case when weighted CLT holds. In

fact, in most clinical trials, the strata effects is only moderate. For example, when $n_s = 2$,

$$\Pr(|\hat{K}_j| > M) \rightarrow n^{-1}, \ \forall\, j, \text{ for some positive constant } M<\infty, \text{ as } n\rightarrow\infty,$$

for $n_s=3$,

$$\Pr(|\hat{K}_j| > M) \rightarrow n^{-2}, \ \forall\, j, \text{ for some } M<\infty, \text{ as } n\rightarrow\infty,$$

and for $n_s=4$,

$$\Pr(|\hat{K}_j| > M) \rightarrow n^{-3}, \ \forall\, j, \text{ for some } M<\infty, \text{ as } n\rightarrow\infty.$$

As the stratum sizes get larger, the probability that any $\hat{K}$ will diverge is near 0. Therefore, for $n_s \geq 3$, the unconditional variance of $\hat{S}$ will converge to the conditional variance of $\hat{S}$, i.e. $\frac{1}{4}$.

§ 4.5 *Variance Estimate of the Proposed Test*

The asymptotic variance of the numerator of the proposed test conditioned on the estimated stratum effects when in general the weighted CLT does not apply is given in equations (4.3.6a,b):

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = p(1-p)\sum_{i=1}^{n} w_i^2 \text{ for unbalanced randomization}$$

and

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = p(1-p)\sum_{i=1}^{n} w_i^2 - \frac{p(1-p)}{(n_s-1)}\sum_{i^*\neq j^*}\sum w_{i^*}w_{j^*}$$

for balanced randomization within stratum, where i^* and j^* are in the same stratum

and,

$$\mathcal{W}_i = \left(1 - \hat{K}_{(i)} \sum_{j=1}^{i} \frac{1}{\sum_{l \in \mathcal{R}_j} \hat{K}_{(l)}} \right).$$

The proposed variance estimates is the same as the conditional variance:

$$\hat{V} = \mathrm{Var}(\hat{S} \mid \hat{\beta}) \tag{4.5.1}$$

The modified score test proposed is given by:

$$\mathrm{MST} = \frac{\left[n^{-1/2} \sum_{i=1}^{n} x_i \mathcal{W}_i \right]^2}{\hat{V}}. \tag{4.5.2}$$

§ 4.6 *Simulation Results*

A simulation study is conducted to examine properties of the MST vs. the LRT, the SLRT and the MLRT when there are moderate strata effects. The asymptotic relative efficiency (ARE), or Pitman efficiency, represents the relative efficiency of the test when the sample size is large and the alternative hypothesis is close to the null hypothesis, is used as a measure of the amount of precision lost by the use of the test instead of the theoretically optimal test. The comparisons were made based on the ARE of each test. The AREs are computed from the means and variances of the test statistics determined over the simulations. The definition of the ARE is given in Section3.4. The variances of the other tests were not significantly different from 1.0. The simulations were conducted as follows:

(1) Failure times were generated according to a PH model. Specifically an exponential PHM was used. However, since all test considered only used rank information, any other parametric PHM would give identical results.

(2) The slope of the $E(\hat{\alpha})$ was approximated by performing the simulations at

α=0 and $\alpha = \log(1.25)$.

(3) Efficiencies of the LRT were based on assuming the variance was 1.0 as was used in practice, rather than the true variance.

(4) Treatments are balanced within stratum. Total sample size is 200, with 100 patients on each treatment and no censoring. Censored samples will be considered in Section 5.3.

(5) For each data set, the tests were computed and the means, variances and the AREs of the test statistics were determined over the replications.

(6) Each experiment was simulated 1000 times.

The results of the simulation are presented in Table 4.1. The simulations were performed on a Micro VAX II computer, using FORTRAN language and its random number generator.

The stratum parameter, β, was generated to be uniformly distributed with range A, i.e. $\beta \sim$ Uniform(0, A). In the simulation, the values of A = 0, 1.5, 3 and 7 which correspond to hazard ratios of being in upper quartile versus being in lower quartile equal to 1, 2.1, 4.5, and 33.1, respectively.

Table 4.1 summarize the simulation results for stratum size equal 2 (corresponding to matched pair design) and for stratum size equal 10, with balanced design and no censoring. The values in parentheses are the theoretical ARE's which were used to plot Figure 3.11 in Chapter 3. As was shown in Table 4.1, the MST is nearly fully efficient (>95%) for no strata effects even when the number of subjects within strata are as small as 10. However, when stratum size is as small as 2 (matched pairs), the efficiency drops to about 70% efficient. The efficiency of MST is better than SLRT for all distributions of strata effects. The efficiencies of the LRT and the MLRT are 0.99 and 0.96, compared to the theoretical value, 1.0, when there are no stratum effects. However, as the stratum effects get large, the efficiencies of both the LRT and

the MLRT decline dramatically. The LRT is less efficient than the SLRT for A>3 (stratum size 2) and A>1.5 (stratum size 10), and the MLRT is less efficient than the SLRT for A>3 (stratum size 10). The values of A=1.5 and 3 correspond to hazard ratios of the interquartile ranges of 2.1 and 4.5. Notice that the LRT is consistently less efficient than the MLRT at different stratum sizes when the strata effects are large. As pointed out by Schoenfeld and Tsiatis (1987), when there are strata effects the variance estimate of the LRT (the commonly used denominator of the LRT) is not correct. The MLRT is actually the correct version of the LRT (with the correct variance estimate).

Figure 4.a and Figure 4.b gave graphical illustration of the simulation results for stratum size 2 and 10 respectively. As can be seen from Figures 4.a and 4.b, the MST is always better than the SLRT irregardless of the stratum effects. Note the flatness of the efficiencies as a function of A that shown in Figures 4.a and 4.b, as compared with the quick drop off of the LRT and the MLRT curves. As the stratum effects get larger, correspond to when approximately A>1.75 for the LRT, and A>3.25 for the MLRT when stratum size is 2 and when approximately A>0.5 for stratum size 10, both the LRT and the MLRT become less efficient than the MST. It can be predicted that for larger stratum sizes, the MST is fully efficient as the LRT and the MLRT when no stratum effects.

The true distribution of strata effects was never known in practice, however, these results give some indication of how sensitive the tests are to this distribution and how good these tests can be under a wide range of possibilities.

Table 4.1

Simulation results, no censoring, balanced design

Stratum size	Stratum† effect A	ARE of LRT	ARE of SLRT	ARE of MLRT	ARE of MST
2	0	0.9930 (1.0)	0.4923 (.5)	0.9606 (1.0)	0.7212
2	1.5	0.7545 (.7564)	0.4923 (.5)	0.8327 (.8813)	0.7127
2	3	0.4865 (.4843)	0.4923 (.5)	0.7102 (.7434)	0.6960
2	7	0.1939 (.1915)	0.4923 (.5)	0.5270 (.5455)	0.6791
10	0	0.9807 (1.0)	0.7670 (.7857)	0.9966 (1.0)	0.9634
10	1.5	0.7638 (.7564)	0.7670 (.7857)	0.8935 (.8813)	0.9653
10	3	0.4961 (.4843)	0.7670 (.7857)	0.7424 (.7434)	0.9554
10	7	0.2071 (.1915)	0.7670 (.7857)	0.5394 (.5455)	0.9242

† The stratum effect assumes that the stratum parameter, $\beta \sim U(0, A)$, the values of A = 0, 1.5, 3, and 7 correspond to hazard ratios of the upper quartile versus the lower quartile equal to 1, 2.1, 4.5 and 33.1 respectively. The values within parentheses are the theoretical AREs.

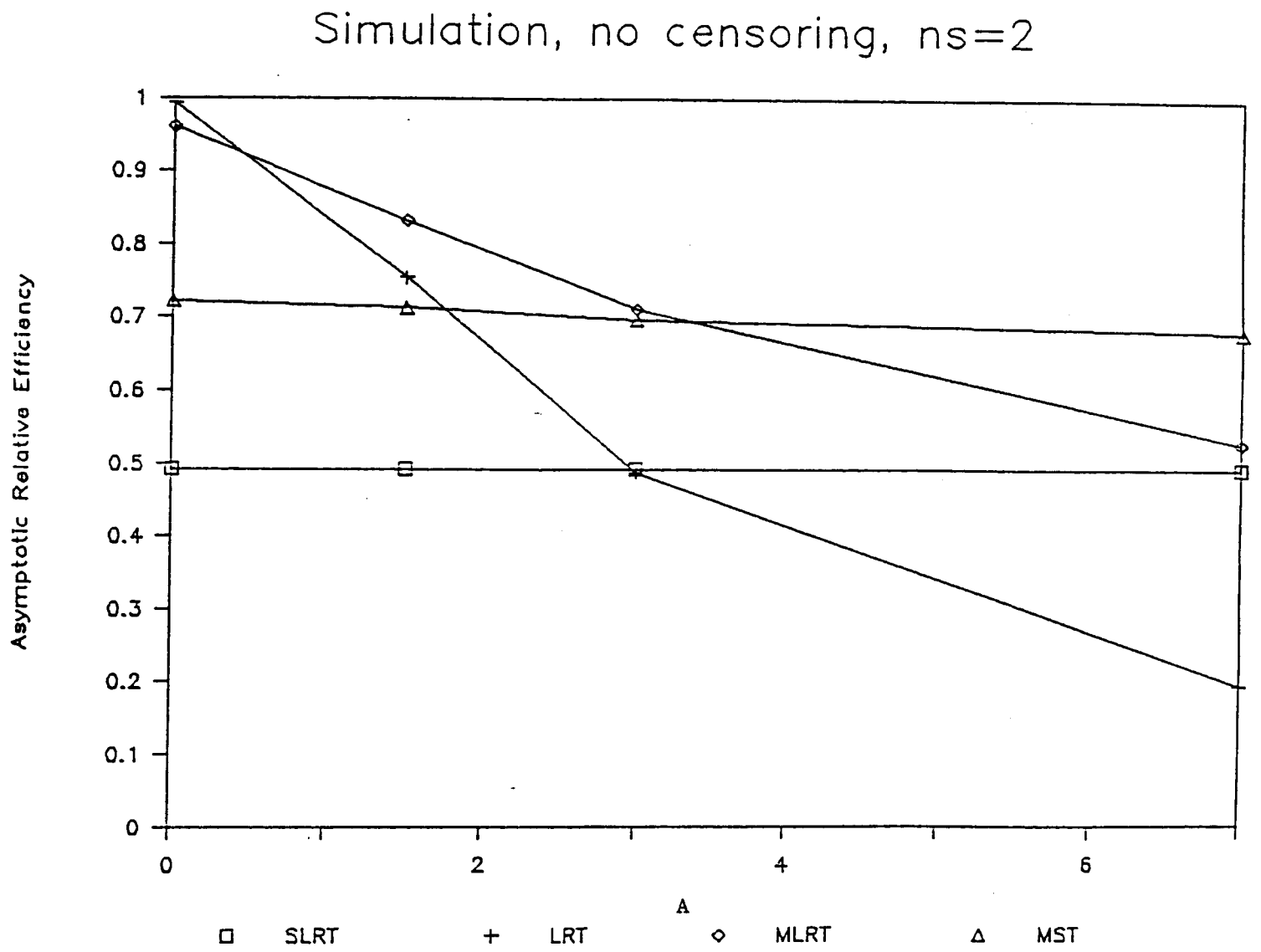

Fig. 4.a: Simulation results for stratum size 2 , Beta Uniform(0, A)

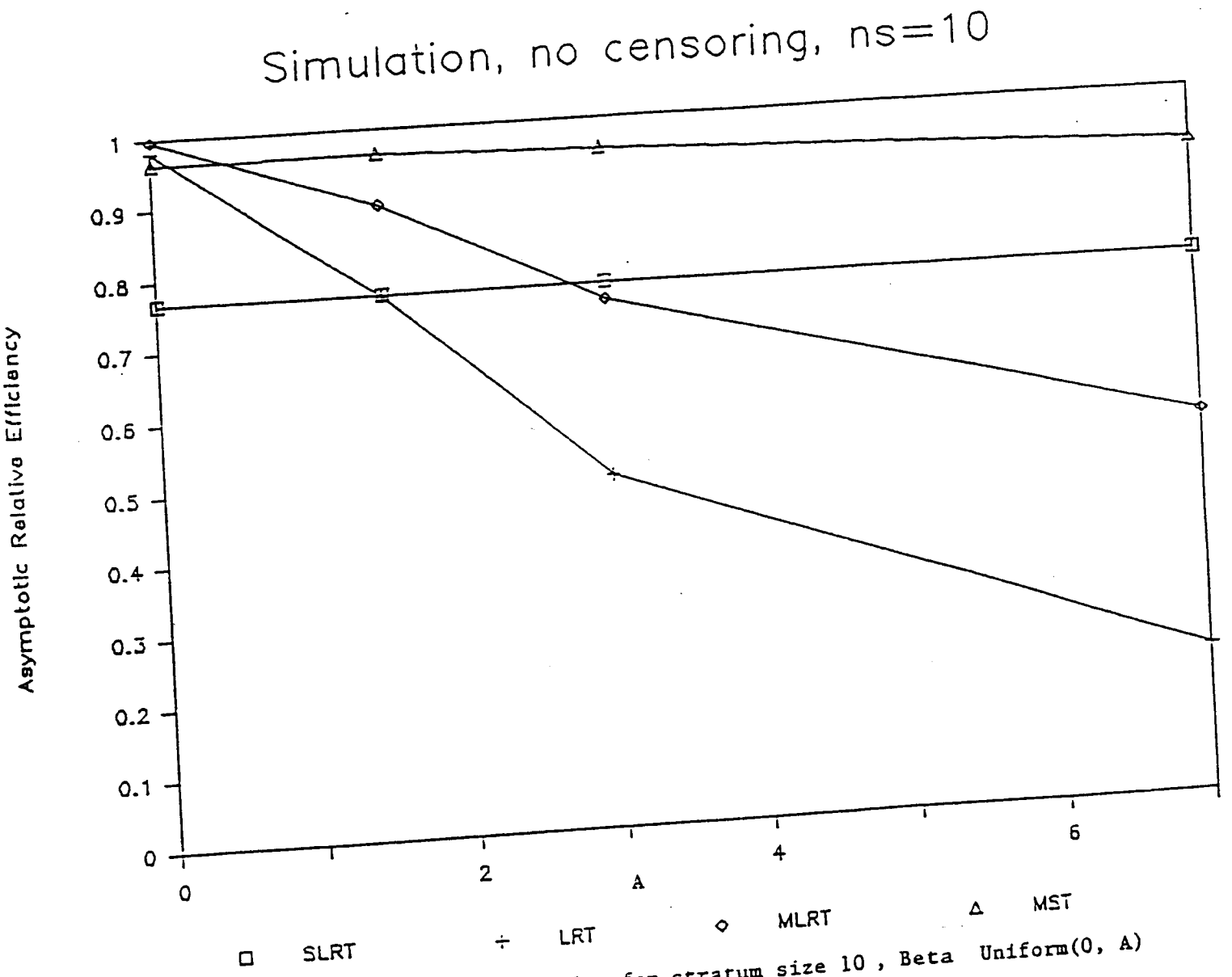

Fig. 4.b: Simulation results for stratum size 10 , Beta Uniform(0, A)

§ 4.7 *Distribution of Strata Effects for Large Stratum Sizes*

The distribution of the stratum effect, $\hat{K}_j$, for the case of large stratum sizes is studied in this section. Some additional definitions are:

n: total sample size

n_{s_j}: number within stratum j (n_{s_j} fixed)

n_g: number of stratum

$K_j = e^{\beta_j}$ $\quad j = 1, \cdots, n_g$ $\quad \beta$ is the parameter for indicator strata effects

n_{ji} = number of observations at risk at time i in stratum j

$E(\hat{K}_j) = \bar{K}$ (some constant)

Recall that the score statistic for the parameter for treatment effect, with indicator variables for strata effects, is given by:

$$\hat{S} = n^{-1/2} \sum_{i=1}^{n} \left[x_i - \frac{\sum_{\Re_i} x_j e^{\hat{\beta} z_j}}{\sum_{\Re_i} e^{\hat{\beta} z_j}} \right].$$

The parameter estimates for z, $\hat{\beta}$, are the solutions to the log likelihood equations subject to one constraint, say, $\bar{K}=1$:

$$0 = \sum_{i=1}^{n} z_{1i} - \sum_{i=1}^{n} \frac{\sum_{\Re_i} z_{1j} e^{\hat{\beta} z_j}}{\sum_{\Re_i} e^{\hat{\beta} z_j}}$$

$$= \sum_{i=1}^{n} z_{1i} - \sum_{i=1}^{n} \frac{n_{1i} \hat{K}_1}{\sum_{\Re_i} n_{ji} \hat{K}_j} \qquad (4.7.1)$$

$$0 = \sum_{i=1}^{n} z_{n_{gi}} - \sum_{i=1}^{n} \frac{\sum_{\Re_i} z_{n_{gj}} e^{\hat{\beta} z_j}}{\sum_{\Re_i} e^{\hat{\beta} z_j}}$$

$$= \sum_{i=1}^{n} z_{n_{gi}} - \sum_{i=1}^{n} \frac{n_{n_{gi}} \hat{K}_{n_g}}{\sum_{\Re_i} n_{ji} \hat{K}_j}$$

where $\Re_i$ indicates that the summation is over the risk set at time t_i.

Note that,

$$\sum_{j \in \Re_i} e^{\hat{\beta} z_j} = \sum_{j=1}^{n_g} n_{ji} \hat{K}_j \qquad \text{and} \qquad \sum_{j \in \Re_i} z_{1j} e^{\hat{\beta} z_j} = n_{1i} \hat{K}_1$$

The assumptions made here are as follows,

(a) no strata effects, i.e. $K_1 = K_2 = \cdots = K_{n_g}$

(b) $\sum_{j=1}^{n_g} n_{ji} \hat{K}_j$ converges to $\sum_{j=1}^{n_g} n_{ji} \bar{K} = \bar{K} \cdot n_i$, $\forall$ i, $\bar{K}$, some constant

(c) There exist an $M < \infty$, such that $Pr(|\hat{K}_j| > M) \rightarrow 0$, for all j, as $n \rightarrow \infty$.

To show that assumption (b) is true for large stratum sizes, i.e. $n_{s_j} \rightarrow \infty$, as $n \rightarrow \infty$, note that at time 1:

$\sum_{j=1}^{n_g} n_{j1} \hat{K}_j = \sum_{j=1}^{n_g} n_{s_j} \hat{K}_j$ will converge to $n \cdot \bar{K}$ by Central Limit Theorem (CLT),

and at time 2:

$$\sum_{j=1}^{n_g} n_{j2} \hat{K}_j = \sum_{j=1}^{n_g} n_{j1} \hat{K}_j - \hat{K}_l^*,$$

where $\hat{K}_l^*$ is the $\hat{K}$ in stratum l corresponding to the stratum of the person who died first. By the CLT, this will converge to:

$$\sum_{j=1}^{n_g} n_{j2}\, \mathbb{E}\hat{K}_j = \sum_{j=1}^{n_g} n_{j1}\, \overline{K} - \mathbb{E}\hat{K}_l^* .$$

Since $n_{s_j} \to \infty$ as $n \to \infty$,

$$\mathbb{E}\hat{K}_l^* = \frac{n_{s_l}-1}{n_{s_l}}\, \mathbb{E}\hat{K}_l \to \overline{K}.$$

Therefore, when $n_{s_j} \to \infty$, $\forall$ j, as $n \to \infty$, assumption (b) will hold. Some results from the following two lemmas are necessary before proceeding to the main theorem.

Lemma 4.4 If assumption (a) holds, as $n \to \infty$, n_{ji} and n_{li} ($j \neq l$, j, l, stratum indicator) are independently distributed as binomial(n_{s_j}, $\frac{n_i}{n}$), where n_{s_j} is fixed and $n_i = n-i+1$, i=1, 2, $\cdots$, n.

Proof: WLOG, let $(j, l) = (1, 2)$. Consider the random experiment with only two possible outcomes, a success or failure. The number at risk at time i in stratum j, n_{ji}, can be thought of as the sum of n_{s_j} independent Bernoulli random variables with common probability p=Pr(success)=$\frac{n_i}{n}$. Alternately, it may be thought of as the number of trials at time i in stratum j that yielded a success (survival) among the n_{s_j} trials. Therefore n_{ji} is binomial(n_{s_j}, $\frac{n_i}{n}$), $\forall$ j. To prove the independence of the n_{ji}, it is sufficient to show that the conditional distribution of $n_{1i}|\, n_{2i}$ is the same as the marginal distribution of n_{1i}. Specifically, it may be shown that given n_{2i}, $n_{1i}|\, n_{2i} \sim$ Binomial(n_{s_1}, $\frac{n_i-n_{2i}}{n-n_{s_2}}$). Since n_{s_2} is fixed, as $n \to \infty$,

$$\frac{n_i-n_{2i}}{n-n_{s_2}} = \frac{\frac{n_i}{n} - \frac{n_{2i}}{n}}{1 - \frac{n_{s_2}}{n}} \to \frac{n_i}{n}, \qquad \text{i.e.} \quad n_{1i}|\, n_{2i} \sim n_{1i}$$

Therefore, n_{1i} and n_{2i} are independent. □

Lemma 4.5 If $\lim_{n\to\infty} \Pr(|\hat{K}_j| > M) \to 0$, for some $M < \infty$, $\forall$ j, (i.e. $\hat{K}_j$ is bounded in probability) and $v_n = o(n^b)$, $b<0$, then $v_n \cdot \hat{K}_j \to 0$ as $n \to \infty$.

Proof: Let τ be some finite number, then,

$$\lim_{n\to\infty} \Pr(|v_n \cdot \hat{K}_j| > \tau) = \lim_{n\to\infty} \Big\{ \Pr(|v_n \cdot \hat{K}_j| > \tau, |\hat{K}_j| \le M) + \Pr(|v_n \cdot \hat{K}_j| > \tau, |\hat{K}_j| > M) \Big\}.$$

The first term of the RHS goes to zero since $\hat{K}_j$ is bounded and the second term of the RHS also goes to zero because

$$\Pr(|v_n \cdot \hat{K}_j| > \tau, |\hat{K}_j| > M) \le \Pr(|\hat{K}_j| > M) \to 0 \text{ as } n \to \infty. \qquad \square$$

Using the results of Lemma 4.4 and Lemma 4.5, the following theorem concerning the distribution of the $\hat{\beta}$'s can be established.

Theorem 4.2: Under assumptions (a) - (c), the $\hat{\beta}_j$'s are independent and identically with density function

$$f_Y(y) = \frac{n_{s_j}^{n_{s_j}}}{\Gamma(n_{s_j})} e^{-n_{s_j} y} e^{-n_{s_j} e^{-y}},$$

where j= 1, 2, $\cdots$, n_g.

Proof: Let $m_{j(i)}$ be the death order within the entire population for the person with death order i within the j^{th} stratum (i=1, 2, $\cdots$, n_{s_j} and j= 1, 2, $\cdots$, n_g). The following example illustrates a possible $m_{j(i)}$. For this example, let the trial have 3 strata (1, 2, 3) each stratum have 3 subjects ($n_{s_1} = n_{s_2} = n_{s_3} = 3$). If the death order with respect to stratum indicators is as follows

death order:	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	$(t_1<t_2<\cdots<t_8<t_9)$
w.r.t. *strata*:	1	1	3	3	2	2	1	2	1,	

then the death order within each stratum is:

Stratum 1 : 1 2 3

Stratum 2 : 1 2 3

Stratum 3 : 1 2 3 ,.

Thus, $m_{2(1)} = 2$, $m_{3(2)} = 4$ and $m_{1(3)} = 9$.

The following proof consists of three parts: (i) considers the distribution of $\hat{\beta}_1$, i.e. stratum 1. (ii) considers the distribution of $\hat{\beta}_2, \cdots, \hat{\beta}_{n_g}$. (iii) considers the independence of $\hat{\beta}_i$ and $\hat{\beta}_j$, $\forall\ i \neq j$, as $n \to \infty$.

(i) For stratum 1, using the definition of $m_{j(i)}$ and assumption (b), equation (4.7.1) can be written as:

$$0 = n_{s_1} - \sum_{i=1}^{n} \frac{n_{1i}\hat{K}_1}{\sum_{j=1}^{n_g} n_{ji}\hat{K}_j} = n_{s_1} - \sum_{i=1}^{n} \frac{n_{1i}\hat{K}_1}{n_i \bar{K}}$$

$$\Rightarrow \quad n_{s_1} = \frac{\hat{K}_1}{\bar{K}} \left(\sum_{i=1}^{n} \frac{n_{1i}}{n_i} \right)$$

$$= \frac{\hat{K}_1}{\bar{K}} \left(\sum_{i=1}^{m_{1(1)}} \frac{n_{s_1}}{n-i+1} + \sum_{i=m_{1(1)}+1}^{m_{1(2)}} \frac{n_{s_1}-1}{n-i+1} + \cdots + \sum_{i=m_{1(n_{s_1}-1)}+1}^{m_{1(n_{s_1})}} \frac{1}{n-i+1} \right.$$

$$= \frac{\hat{K}_1}{\bar{K}} \left(\sum_{i=1}^{m_{1(1)}} \frac{1}{n-i+1} + \sum_{i=1}^{m_{1(2)}} \frac{1}{n-i+1} + \cdots + \sum_{i=1}^{m_{1(n_{s_1})}} \frac{1}{n-i+1} \right)$$

$$\Rightarrow \quad \frac{\hat{K}_1}{\bar{K}} \left(\int_0^{\frac{m_{1(1)}}{n}} \frac{1}{1-u}\, du + \int_0^{\frac{m_{1(2)}}{n}} \frac{1}{1-u}\, du + \cdots + \int_0^{\frac{m_{1(n_{s_1})}}{n}} \frac{1}{1-u}\, du \right)$$

$$= \frac{\hat{K}_1}{\bar{K}} \left(F\left(\frac{m_{1(1)}}{n}\right) + F\left(\frac{m_{1(2)}}{n}\right) + \cdots + F\left(\frac{m_{1(n_{s_1})}}{n}\right) \right),$$

where

$$F(x) = \int_0^x \frac{1}{1-u}\,du \;=\; -\log(1-x).$$

From the above,

$$n_{s_1} = \frac{\hat{K}_1}{\bar{K}} \sum_{i=1}^{n}\Big(-\log(1-\frac{m_{1(i)}}{n})\Big). \tag{4.7.2}$$

Let $r_{1i} = \frac{m_{1i}}{n}$, for randomly chosen i, i=1, 2, $\cdots$, n_{s_1}, based on assumption (a), the r_{1i} are *i.i.d.* uniformly distributed. i.e.

$$r_{1i} \sim i.i.d.\ U(0,1) \qquad \text{and} \qquad (1 - r_{1i}) \sim i.i.d.\ U(0,1)$$

so that

$$-\log(1-r_{1i}) \sim i.i.d.\ \exp(1).$$

Since the sum of the ordered random variables is equivalent to the sum of the independent unordered random variables, the sum of the ordered random variables are distributed as the sum of n_{s_1} independent random exponential variables with mean 1. i.e.

$$\sum_i\Big(-\log(1-r_{1(i)})\Big) \;=\; \sum_i\Big(-\log(1-r_{1i})\Big) \;\sim\; \text{Gamma}(n_{s_1},1).$$

From equation (4.7.2),

$$\frac{n_{s_1}\,\bar{K}}{\hat{K}_1} \sim \text{Gamma}(n_{s_1},1), \text{ such that}$$

$$\frac{1}{\hat{K}_1} = e^{-\hat{\beta}_1} \sim \text{Gamma}(n_{s_1},\, n_{s_1}\,\bar{K}). \tag{4.7.3}$$

Setting $\bar{K} = 1$, $Y=-\log(\frac{1}{\hat{K}_1})=\hat{\beta}_1$ will have the following distribution:

$$f_Y(y) = \frac{n_{s_1}^{n_{s_1}}}{\Gamma(n_{s_1})} e^{-n_{s_1}y} e^{-n_{s_1}e^{-y}} . \qquad (4.7.4)$$

(ii) The distributions of $\hat{\beta}_2, \cdots, \hat{\beta}_{n_g}$ can be derived similarly and are the same as the distribution of $\hat{\beta}_1$.

(iii) We want to show that $\hat{\beta}_i$ and $\hat{\beta}_j$, $\forall\ i \neq j$ are independent as $n \to \infty$. Recall that $\hat{K}_1$ can be expressed as the solution to,

$$n_{s_1} = \sum_{i=1}^{n} \frac{n_{1i}\,\hat{K}_1}{\sum_{j=1}^{n_g} n_{ji}\hat{K}_j} \quad \to \quad \sum_{i=1}^{n} \frac{n_{1i}\,\hat{K}_1}{n_i\,\bar{K}} .$$

$\hat{K}_1$ conditioned on the value of $\hat{K}_2$ (i.e. $\hat{K}_1|\hat{K}_2$) is the solution to,

$$n_{s_1} = \sum_{i=1}^{n} \frac{n_{1i}\,\hat{K}_1}{\sum_{\substack{j=1 \\ j\neq 2}}^{n_g} n_{ji}\hat{K}_j + n_{2i}\hat{K}_2} \quad \to \quad \sum_{i=1}^{n} \frac{n_{1i}\,\hat{K}_1}{n_i\,\bar{K} + n_{2i}(\hat{K}_2-\bar{K})}$$

In order to show that $\hat{K}_1|\hat{K}_2 \to \hat{K}_1$, as $n \to \infty$, one needs to show that

$$\frac{n_{s_1}}{\hat{K}_1|\hat{K}_2} - \frac{n_{s_1}}{\hat{K}_1} \to 0 \text{ as } n \to \infty.$$

Alternately,

$$\frac{n_{s_1}}{\hat{K}_1|\hat{K}_2} - \frac{n_{s_1}}{\hat{K}_1} = \sum_{i=1}^{n} \frac{n_{1i}\,n_{2i}\,(\hat{K}_2-\bar{K})}{n_i^{\,2}\,\bar{K}^2 + n_i\,n_{2i}\,\bar{K}(\hat{K}_2-\bar{K})}$$

$$= \frac{1}{\bar{K}^2} \sum_{i=1}^{n} \frac{n_{1i}\,n_{2i}\,(\hat{K}_2-\bar{K})}{n_i^{\,2}\,\bar{K} + n_i\,n_{2i}(\hat{K}_2-\bar{K})}$$

$$= \frac{1}{\bar{K}^2} \sum_{i=1}^{n} \frac{\frac{n_{1i}n_{2i}}{n_i n_i}\left(\frac{\hat{K}_2}{\bar{K}} - 1\right)}{1 + \frac{n_{2i}}{n_i}\left(\frac{\hat{K}_2}{\bar{K}} - 1\right)}$$

$$\stackrel{(*)}{=} \frac{1}{\bar{K}^2} \sum_{i=1}^{n} \left\{ \frac{n_{1i}n_{2i}}{n_i n_i} \left(\frac{\hat{K}_2}{\bar{K}} - 1\right) - \frac{n_{1i}}{n_i}\left(\frac{n_{2i}}{n_i}\right)^2 \left(\frac{\hat{K}_2}{\bar{K}} - 1\right)^2 + O(n^{-4}) \right\}$$

$$= \frac{1}{\bar{K}^2} \sum_{i=1}^{n} \left\{ C_1 - C_2 + C_3 \right\}, \qquad (4.7.5)$$

where $(*)$: $\frac{aX}{1+X} \doteq aX(1 - X + X^2 - X^3 + \cdots)$.

Recall that, $n_{ji} \sim$ binomial $(n_{s_j}, \frac{n_i}{n})$ and $n_i = n - i + 1$, for $j = 1, 2$; $i = 1, 2, \cdots, n$, thus, $\frac{E(n_{ji})}{n_i} = \frac{n_{s_j}(n_i/n)}{n_i} = \frac{n_{s_j}}{n} = O(n^{-1})$. The independence of n_{1i} and n_{2i} by Lemma 4.3 implies that $E(n_{1i}n_{2i}) = E(n_{1i})E(n_{2i})$. By assumption (c), it can be shown that $(\frac{\hat{K}_2}{\bar{K}} - 1)$ is bounded, therefore, by Lemma 4.4 the first term (i.e. C_1) of equation (4.7.5) will converge to $O(n^{-2})$, $\forall$ i, the second term is of order $O(n^{-3})$, $\cdots$ etc. Therefore,

$$\frac{n_{s_1}}{\hat{K}_1|\hat{K}_2} - \frac{n_{s_1}}{\hat{K}_1} = \frac{1}{\bar{K}^2} O(n^{-1}) \to 0 \qquad \text{as } n \to \infty$$

and

$$\hat{K}_1|\hat{K}_2 \to \hat{K}_1 \quad \text{is independent of the value of } \hat{K}_2. \qquad \square$$

In summary, when the number of subjects within each stratum is fixed (i.e. the n_s are fixed), under assumptions (a)-(c), the $\hat{\beta}$'s are *i.i.d.* distributed as equation (4.7.4). When the stratum size is large the distribution of the stratum effect is approximately normal and is given in the following Corollary.

Corollary. If $n_{s_j} \to \infty$ as $n \to \infty$, then $\hat{\beta}_j$s are distributed $N(0, \frac{1}{n_{s_j}})$.

Proof: WLOG, let $\bar{K} = 1$. From equation (4.7.3), it can be shown that for stratum j,

$$\frac{n_{s_j}}{\hat{K}_j} \sim \text{Gamma}(n_{s_j}, 1).$$

Hence, the distribution of $n_{s_j}/\hat{K}_j$ may be thought of as the distribution of the sum of n_{s_j} *i.i.d.* exp(1) random variables, each with mean 1 and variance 1. Therefore,

$$\frac{1}{\hat{K}_j} = e^{-\hat{\beta}_j} \sim \text{Gamma}(n_{s_j}, n_{s_j}),$$

is the distribution of the mean of the n_{s_j} *i.i.d.* exp(1) random variables.

Therefore, by CLT, as $n_{s_j} \to \infty$,

$$n_{s_j}\left(\frac{1}{\hat{K}_j} - 1\right) \sim N(0,1).$$

Alternatively,

$$\frac{1}{\hat{K}_j} \sim N(\mu, \sigma^2), \text{ where } \mu=1 \text{ and } \sigma^2 = \frac{1}{n_{s_j}}.$$

Note that $\sigma^2 \to 0$ as $n_{s_j} \to \infty$. Let $g(x)=\log\left(\frac{1}{\hat{K}_j}\right) = -\hat{\beta}_j$, since $\log(\cdot)$ is a real-valued function differentiable at $x=\mu=1$, with $g'(\mu) \neq 0$ and $\sigma^2 \to 0$ as $n_{s_j} \to \infty$. By Theorem A from Serfling (1980, p.118), g(x) is Normal$(g(\mu),[g'(\mu)]^2/n_{s_j})$ as $n_{s_j} \to \infty$. Alternately, one may write

$$-\hat{\beta}_j \sim N(0, \frac{1}{n_{s_j}}), \text{ as } n_{s_j} \to \infty$$

or

$$\sqrt{n_{s_j}}\,\hat{\beta}_j \sim N(0, 1), \text{ as } n_{s_j} \to \infty . \qquad \square$$

For stratification with fixed number of strata (i.e. n_g fixed and $n_{s_j} \to \infty$ as $n \to \infty$), this corollary says that the $\hat{\beta}$'s are normally distributed with mean 0 and variance $\frac{1}{n_{s_j}}$. However, for stratification with fixed number within stratum (i.e. n_{s_j} fixed and

$n_g \to \infty$ as $n \to \infty$), this corollary does not hold. In general, if stratum sizes are small, assumption (b) will not hold and the above proof will not hold either.

The reason why assumption (b) fails to hold under these conditions is explained as follows. If n_{s_j} is fixed, at time 1,

$$\sum_{j=1}^{n_g} n_{j1}\hat{K}_j = \sum_{j=1}^{n_g} n_{s_j}\hat{K}_j \quad \text{will converge to } n\overline{K} \text{ by CLT.}$$

and at time 2

$$\sum_{j=1}^{n_g} n_{j2}\hat{K}_j = \sum_{j=1}^{n_g} n_{j1}\hat{K}_j - \hat{K}_l^*, \tag{4.7.6}$$

where $\hat{K}_l^*$ is the $\hat{K}$ in stratum l corresponding to the stratum of the person who died first. Again by the CLT equation (4.7.6) will converge to

$$\sum_{j=1}^{n_g} n_{j2}\,E\hat{K}_j = \sum_{j=1}^{n_g} n_{j1}\,\overline{K} - E\hat{K}_l^* .$$

In order for assumption (b) to hold, it is necessary that $E(\hat{K}_l^*) = \overline{K}$. However, for small n_{s_j} (e.g. n_s=2), it is obvious that $E(\hat{K}_l^*) > \overline{K}$. The $\hat{K}$ of stratum l is chosen because one of its members died first. Thus,

$$\sum_{j\in\Re_2} E\hat{K}_j = \sum_{j=1}^{n_g} n_{j2}\,E\hat{K}_j \neq \sum_{j=1}^{n_g} n_{j2}\,\overline{K}.$$

In general, if n_{s_j} is small, assumption (b) fails to hold. The principle goal of this research is to examine the case of highly stratified data where the stratum sizes are considered fixed and can be very small. Therefore, the distribution of strata effects, will be derived heuristically, without using assumption (b), in the following section.

§ 4.8 *Derivation of Expected Value of Strata Effects*

The distribution of stratum effect estimates were derived in a previous section for the case of large stratum sizes. For small stratum sizes, in order to derive the distribution of stratum effect estimates, $\hat{K}$, (will be shown in Section 4.9), the expected value of these estimates conditioned on the death order of one of the member in that stratum, the EK's, will be required. There are three methods discussed in this section for the derivation of the distribution of the EK's, direct analytical method, approximate method and numerical method. It was hoped that there is a closed form expression for the distribution of the EK's. However, neither the direct method (in Section 4.8.1) nor the approximate method (in Section 4.8.2) provides a closed form solution. Hence, a numerical method is discussed in Section 4.8.3.

4.8.1 *Analytical Method*

In order to derive the distribution of $\hat{K}_j$ without assumption (b), the conditional expected values of the K's are defined as follows.

Define

$$\begin{aligned} EK(1) &= E(\hat{K}_j | \text{ one of the members of stratum j died first}), \\ EK(2) &= E(\hat{K}_j | \text{ one of the members of stratum j died second}), \\ &\vdots \\ EK(n) &= E(\hat{K}_j | \text{ one of the members of stratum j died last}). \end{aligned} \tag{4.8.1}$$

Note that $\sum_i EK(i) = E\hat{K}_j = \bar{K}$. WLOG, define $\bar{K}=1$. Thus,

$$\sum_{i=1}^{n} EK(i) = n\bar{K} = n \text{ and } \sum_{\Re_2} e^{\underset{\sim}{\beta z}} = n\bar{K} - EK(1) = \sum_{i=2}^{n} EK(i) .$$

In general,

$$E\sum_{\Re_i} e^{\hat{\beta}\underset{\sim}{z}} = \sum_{j=i}^{n} EK(j) .$$

Again, let $m_{j(i)}$ be the ordered death time when the first i^{th} person in j^{th} stratum died and m_{ji} denotes the random death time. From the log likelihood equations (4.7.1), for stratum 1, the likelihood equation is:

$$\frac{n_{s_1}}{K_1} = \sum_{i=1}^{n} \frac{n_{1i}}{\sum_{j\in\Re_i} e^{\hat{\beta}\underset{\sim}{z}_j}}$$

$$\Rightarrow \sum_{i=1}^{n} \frac{n_{1i}}{\sum_{j=i}^{n} EK(j)}$$

$$= \sum_{i=1}^{m_{1(1)}} \frac{n_{s_1}}{\sum_{j=i}^{n} EK(j)} + \sum_{i=m_{1(1)}+1}^{m_{1(2)}} \frac{n_{s_1}-1}{\sum_{j=i}^{n} EK(j)} + \cdots + \sum_{i}^{m_{1(n_{s_1})}} \frac{1}{\sum_{j=i}^{n} EK(j)}$$

$$= \sum_{i=1}^{m_{1(1)}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{1(2)}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \cdots + \sum_{i=1}^{m_{1(n_{s_1})}} \frac{1}{\sum_{j=i}^{n} EK(j)}$$

$$= \sum_{i=1}^{m_{11}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \cdots + \sum_{i=1}^{m_{1n_{s_1}}} \frac{1}{\sum_{j=i}^{n} EK(j)}$$

$\Rightarrow$

$$\int_0^{r_{11}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_{s_1}}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy \qquad (4.8.2)$$

where $r_{ji} = \frac{m_{ji}}{n}$, $j = 1, 2, \cdots, n_g$.

Consider conditioning on one of the members of stratum 1 who died at a (random) time such as $r_1=a$. Equation (4.8.2) becomes,

$$\frac{n_{s_1}}{\hat{K}_1(a)} = \int_0^a \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_{s_1}}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy \quad (4.8.3)$$

$$\Rightarrow \quad \hat{K}_1(a) = \frac{n_{s_1}}{\int_0^a \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_{s_1}}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy} \quad (4.8.4)$$

Define,

$$F(z) = \int_0^z \frac{1}{\int_y^1 EK(x)\,dx}\,dy\,, \quad (4.8.5)$$

where,

$$F(1) = \int_0^1 \frac{1}{\int_y^1 EK(x)\,dx}\,dy \quad \text{and} \quad F(0) = 0.$$

Then,

$$F'(z) = \frac{1}{\int_z^1 EK(x)\,dx} \quad \text{with} \quad F'(1) = \infty\,, \quad F'(0) = 1,$$

and,

$$F''(z) = \frac{E\,K(z)}{\left(\int_z^1 EK(x)\,dx\right)^2} \quad \text{with} \quad F''(1) = \infty,\ F''(0) = EK(0).$$

Taking the expectation on both sides of equation (4.8.4) w.r.t. r_{1i}'s, then,

$$\text{LHS} = E(\hat{K}_1|a) = EK(a) = \frac{F''(a)}{\left(F'(a)\right)^2}$$

$$\text{RHS} = \int_0^1\int_0^1 \cdots \int_0^1 \frac{n_{s_1}}{F(a)+F(r_{12})+\cdots+F(r_{1n_{s_1}})}\,dr_{12}\,dr_{13}\cdots dr_{1n_{s_1}}\,.$$

If multiplied by $F'(a)$ and integrate w.r.t a on the above equations,

$$\Rightarrow \quad \text{LHS} = \log[\, F'(a)] + C \qquad \text{C: some constant}$$

$$\text{RHS} = n_{s_1} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{F'(a)}{F(a)+F(r_{12})+ \cdots +F(r_{1n_{s_1}})}\, dr_{12}\, dr_{13} \cdots dr_{1n_{s_1}}$$

$$= n_{s_1} \int_0^1 \int_0^1 \cdots \int_0^1 \log\{\, F(a)+F(r_{12})+ \cdots +F(r_{1n_{s_1}})\,\}\, dr_{12}\, dr_{13} \cdots dr_{1n_{s_1}} \quad (4.8.6)$$

Equation (4.8.6) is in the form of multidimensional integration equations. Expression of $EK(a)$ in closed form using equation (4.8.6) is analytically very difficult. Tremendous efforts were directed at solving this equation analytically. Many distinguished professors at both UNC and Duke University were consulted. However, the general consensus was that a direct analytical solution does not exist.

4.8.2. *Approximate analytical Method*

In this section, an alternative approach using an approximation suggested by Dr. P. K. Sen is examined. Expanding the right hand side of equation (4.8.3) in a Taylor series about $\hat{K}_1 = E\, K_1(a)$, one obtains

$$\frac{n_{s_1}}{\hat{K}_1(a)} = n_{s_1}\left\{\frac{1}{EK_1(a)} - \left(\hat{K}_1(a)-EK_1(a)\right)\frac{1}{E^2K_1(a)} + \left(\hat{K}_1(a)-EK_1(a)\right)^2\frac{1}{E^3K_1(a)}\right.$$

$$\left. - \left(\hat{K}_1(a)-EK_1(a)\right)^3\frac{1}{E^4K_1(a)} + \left(\hat{K}_1(a)-EK_1(a)\right)^4\frac{1}{E^5K_1(a)} - \cdots \right\}.$$

Taking the expectation on both sides,

$$\mathbb{E}\left(\frac{n_{s_1}}{\hat{K}_1(a)}\right)$$

$$\doteq n_{s_1}\left\{\frac{1}{\mathbb{E}K_1(a)}+\frac{\text{Var}\,\hat{K}_1(a)}{\mathbb{E}^3K_1(a)}-\frac{\mathbb{E}\left(\hat{K}_1(a)-\mathbb{E}K_1(a)\right)^3}{\mathbb{E}^4K_1(a)}+\frac{\mathbb{E}\left(\hat{K}_1(a)-\mathbb{E}K_1(a)\right)^4}{\mathbb{E}^5K_1(a)}\right\}. \quad (4.8.7)$$

The first step is to derive $\mathbb{E}K(a)$ using the approximation

$$\mathbb{E}\left(\frac{1}{\hat{K}_1(a)}\right) \doteq \frac{1}{\mathbb{E}K_1(a)}. \quad (4.8.8)$$

From equations (4.8.3) and (4.8.5), thus,

$$\mathbb{E}\left(\frac{n_{s_1}}{\hat{K}_1(a)}\right) = \int_0^a \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy + \mathbb{E}\left\{\int_0^{r_{12}} \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_{s_1}}} \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy\right\}$$

$$= F(a) + \int_0^1\int_0^1 \cdots \int_0^1 \{\, F(r_{12}) + \cdots + F(r_{1n_{s_1}}) \,\}\, dr_{12}\, dr_{13} \cdots dr_{1n_{s_1}}$$

$$= F(a) + (n_{s_1}-1)\int_0^1 F(z)dz. \quad (4.8.9)$$

WLOG, omit the subscript 1. Note that n_s is fixed here (i.e. as $n \to \infty$, $\frac{n_s}{n} \to 0$). Therefore, using the approximation of equation (4.8.8), equation (4.8.9) becomes

$$\frac{n_s}{\mathbb{E}K(a)} = F(a) + (n_s-1)\int_0^1 F(z)dz. \quad (4.8.10)$$

Differentiating with respect to a on both sides of equation (4.8.10) gives

$$\frac{-\left(\mathbb{E}K(a)\right)' n_s}{\left(\mathbb{E}K(a)\right)^2} = \frac{1}{\int_a^1 \mathbb{E}K(x)\,dx},$$

where $\left(\mathbb{E}K(a)\right)' = \frac{d\,\mathbb{E}K(a)}{d\,a}$,

$\Rightarrow$

$$\int_{a}^{1} \mathbb{E}K(x)\,dx = \frac{\left(\mathbb{E}K(a)\right)^2}{\left(\mathbb{E}K(a)\right)' n_s}. \qquad (4.8.11)$$

Define $g(a) = \int_{a}^{1} \mathbb{E}K(x)\,dx$, then $g'(a) = -\mathbb{E}K(a)$, so that equation (4.8.11) can be written as,

$$g(a) = \frac{\left(g'(a)\right)^2}{g''(a)\,n_s},$$

$\Rightarrow$

$$g(a)\,g''(a)\,n_s = \left(g'(a)\right)^2. \qquad (4.8.12)$$

Let $h = \frac{d\,g(a)}{d\,a} = g'(a)$, note that $g''(a) = \frac{d\,h}{d\,a}$, then

$$h\,h' = \frac{d\,g(a)}{d\,a}\cdot\frac{d\,h}{d\,g(a)} = \frac{d\,h}{d\,a} = g''(a).$$

Therefore, equation (4.8.12) can be written as

$$g(a)\,h\,h'n_s = h^2$$

$\Rightarrow$

$$g(a)\,h'n_s = h$$

$\Rightarrow$

$$n_s\frac{h'}{h} = \frac{1}{g(a)}$$

$\Rightarrow$

$$n_s \log(h) = \log\{g(a)\} + C_o, \qquad C_o\text{: some constant}$$

$\Rightarrow$

$$g'(a) = C_1\{g(a)\}^{1/n_s}$$

Integrating both sides,

$$g(a) = \left\{ \frac{n_s-1}{n_s}(C_1 a + C_2) \right\}^{\frac{n_s}{n_s-1}}.$$

Since $g(0)=EK = 1$ by definition, and using the equality of equation (4.8.10), it can be shown that,

$$g(a) = \int_a^1 EK(x)\,dx = (1-\kappa a)^{\frac{n_s}{n_s-1}}.$$

Therefore,

$$EK(a) = \frac{n_s}{n_s-1}(1-\kappa a)^{\frac{1}{n_s-1}}, \quad \text{with } \kappa\text{: some constant.} \tag{4.8.13}$$

From equation (4.8.5),

$$F(r_{1i}) = \int_0^{r_{1i}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy = \int_0^{r_{1i}} (1-\kappa y)^{\frac{-n_s}{n_s-1}}\,dy.$$

Recall that from equation (4.8.4),

$$\hat{K}(a) = \frac{n_s}{\int_0^a \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_s}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy}.$$

It was planned to derive the first 4 moments (i.e. $E\left\{\hat{K}(a)-EK(a)\right\}^r$, (r= 2, 3, 4), and then derive $EK(a)$ using the approximation,

$$E\left(\frac{n_s}{\hat{K}(a)}\right) \doteq n_s\left\{\frac{1}{EK(a)} + \frac{Var\,\hat{K}(a)}{E^3K(a)}\right\}.$$

However, the distribution of $\hat{K}(a)$ proved to be sufficiently complicated that the variance could not be obtained. Alternately, it was decided to use a numerical method to evaluate equation (4.8.4).

4.8.3. *Numerical Method*

In this section, the distribution of the expected value of the stratum effect will be examined numerically. Based on the log likelihood equations (4.7.1), the likelihood equation for stratum 1 is,

$$\frac{n_s}{\hat{K}_1} = \sum_{i=1}^{n} \frac{n_{1i}}{\sum_{j \in \Re_i} e^{\underset{\sim}{\beta} z_j}}.$$

This can be expressed in terms of the EK(j)'s as defined in equation (4.8.1), then,

$$\frac{n_s}{\hat{K}_1} = \sum_{i=1}^{n} \frac{n_{1i}}{\sum_{j=i}^{n} EK(j)}.$$

In order to numerically evaluate the EK(j)'s, the first step is to condition on one of the member of stratum j dying at time b, say, $m_{ij}=b$, specifically,

$$\frac{n_s}{\hat{K}_1(b)} = \sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \cdots + \sum_{i=1}^{m_{1n_s}} \frac{1}{\sum_{j=i}^{n} EK(j)},$$

$$\Rightarrow \quad \hat{K}_1(b) = \frac{n_s}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \cdots + \sum_{i=1}^{m_{1n_s}} \frac{1}{\sum_{j=i}^{n} EK(j)}}. \qquad (4.8.14)$$

The numerical problem here involves with the summations of functions of several variables over regions with dimensions of the order n_s. The number of function evaluations needed to sample an N-dimensional space increases as the N^{th} power of the number needed to do a one-dimensional summation. For example, when $n_s = 2$, equation (4.8.14) is,

$$E\hat{K}_1(b) = \sum_{m_{12}=1}^{n} \frac{2}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)}}.$$

When $n_s = 3$, equation (4.8.14) becomes,

$$E\hat{K}_1(b) = \sum_{m_{12}=m_{13}=1}^{n}\sum^{n} \frac{3}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{13}} \frac{1}{\sum_{j=i}^{n} EK(j)}},$$

and when $n_s = 4$, equation (4.8.14) becomes,

$$E\hat{K}_1(b) \tag{4.8.14}$$

$$= \sum_{m_{12}}^{n}\sum_{m_{13}}^{n}\sum_{m_{14}}^{n} \frac{4}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{13}} \frac{1}{\sum_{j=i}^{n} EK(j)} + \sum_{i=1}^{m_{14}} \frac{1}{\sum_{j=i}^{n} EK(j)}}.$$

It was numerically feasible to evaluate EK(a) for n_s=2, 3 and 4, which encompasses the strata sizes for most matched studies (e.g. 1 to 1, 2 to 1, 3 to 1, 2 to 2, matched designs). The algorithms for evaluating the EK(i)'s were programmed using

FORTRAN language. The starting value for the iterative computation of the $\mathbb{E}K(i)$'s was 1. The program also used the constraint that $\overline{K}=1$ at each iteration. All arithmetic operations were performed in double precision. The convergence criterion consisted of stopping the iteration, for each n given, when the absolute value of the difference between the $\mathbb{E}K$'s from two consecutive iterations is less than the TOLERANCE parameter ($=10^{-8}$). The n was chosen to be 19, 39, 79, $\cdots$, $2*(n+1)-1$, and compared to the $\mathbb{E}K$'s for the corresponding values from two successive runs until at least three digits of accuracy after the decimal point were obtained. The distribution of the $\mathbb{E}K$'s for different strata sizes are plotted and given in Figure 4.1, 4.2 and 4.3.

Figure 4.1 shows the distribution of $\mathbb{E}K(i)$'s when the stratum size is two. This corresponds to a matched pair design. Note that the $\mathbb{E}K(i)$'s go to infinity when i is very small. The reason for this is that if the stratum size is two, by knowing one of the members in that stratum died first, the expected value of the stratum effect for that stratum will be very large. While knowing that one of the members in the stratum died second leads to an expected value of the stratum effect large, but it is not as large as when one knows that the member died first. As a consequence, the distribution of $\mathbb{E}K$ is a monotone decreasing function in the case of no ties, and the distributions become more and more flat as the n_s increases. Figure 4.2 is the distribution of $\mathbb{E}K(i)$'s when the stratum size is three. This corresponds to a 1 to 2 matched design. The distribution of $\mathbb{E}K$'s shown here for i very small no longer goes to infinity as was true in Figure 4.1. For stratum size 4, it is shown in Figure 4.3 that the distribution of $\mathbb{E}K$'s is more flat than it was in Figure 4.2. For $n_s \to \infty$, (i.e. the stratum size is large), the distribution of $\mathbb{E}K$'s will become a flat line as shown in Figure 4.4. That is, when the stratum size is large, knowing the death order of one of the members in a given stratum will not affect the expected value of stratum effect.

Due to the complexity of the function which involve n_s-dimensional integration, evaluations were possible up to n_s=4 using the Cornell National Supercomputer Facility and the UNC VM system. The computer programs used to generate the EK's are included in the Appendix B.

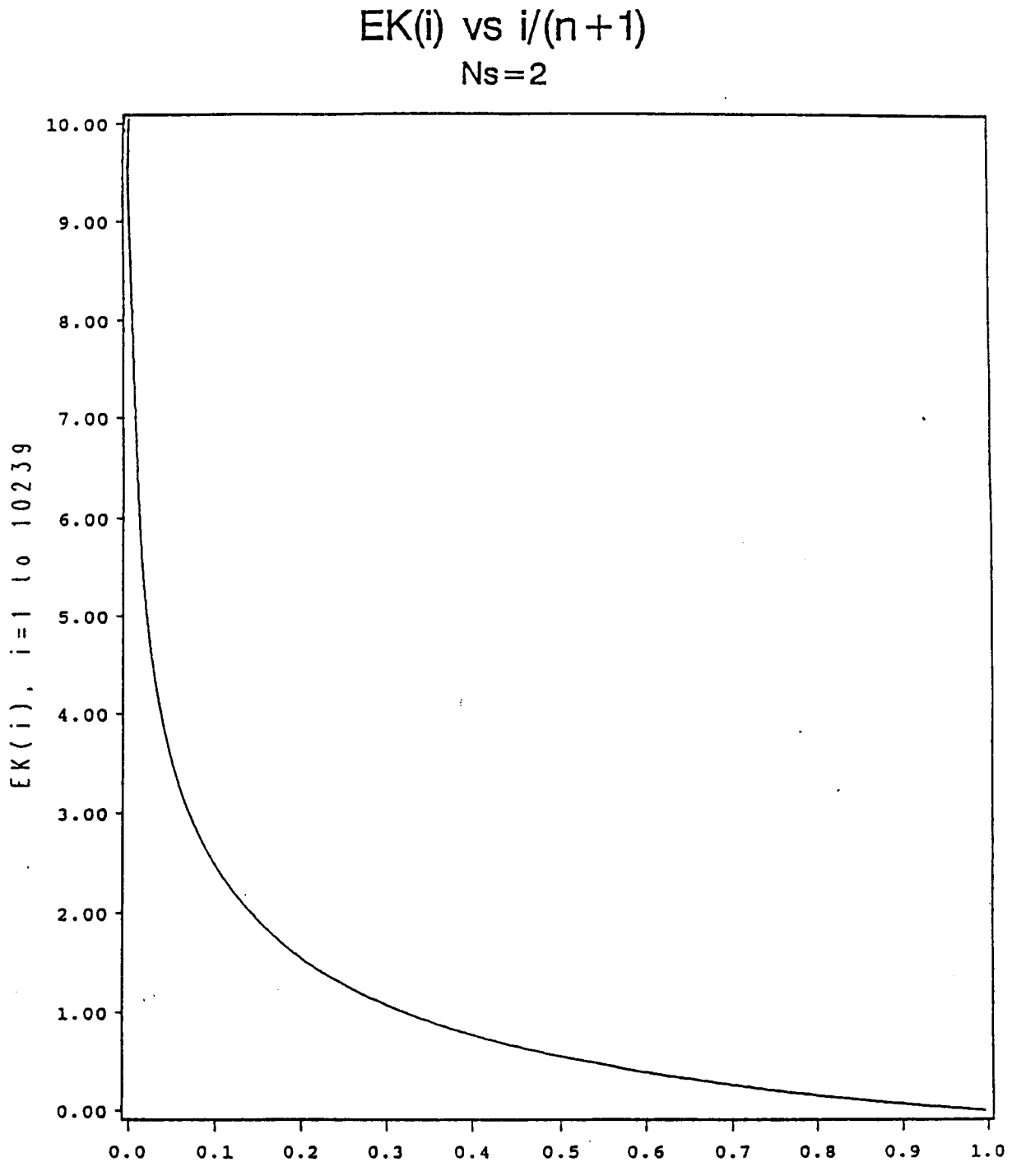

Figure 4.1: **The Distribution of EK when Stratum Size, ns, is 2**

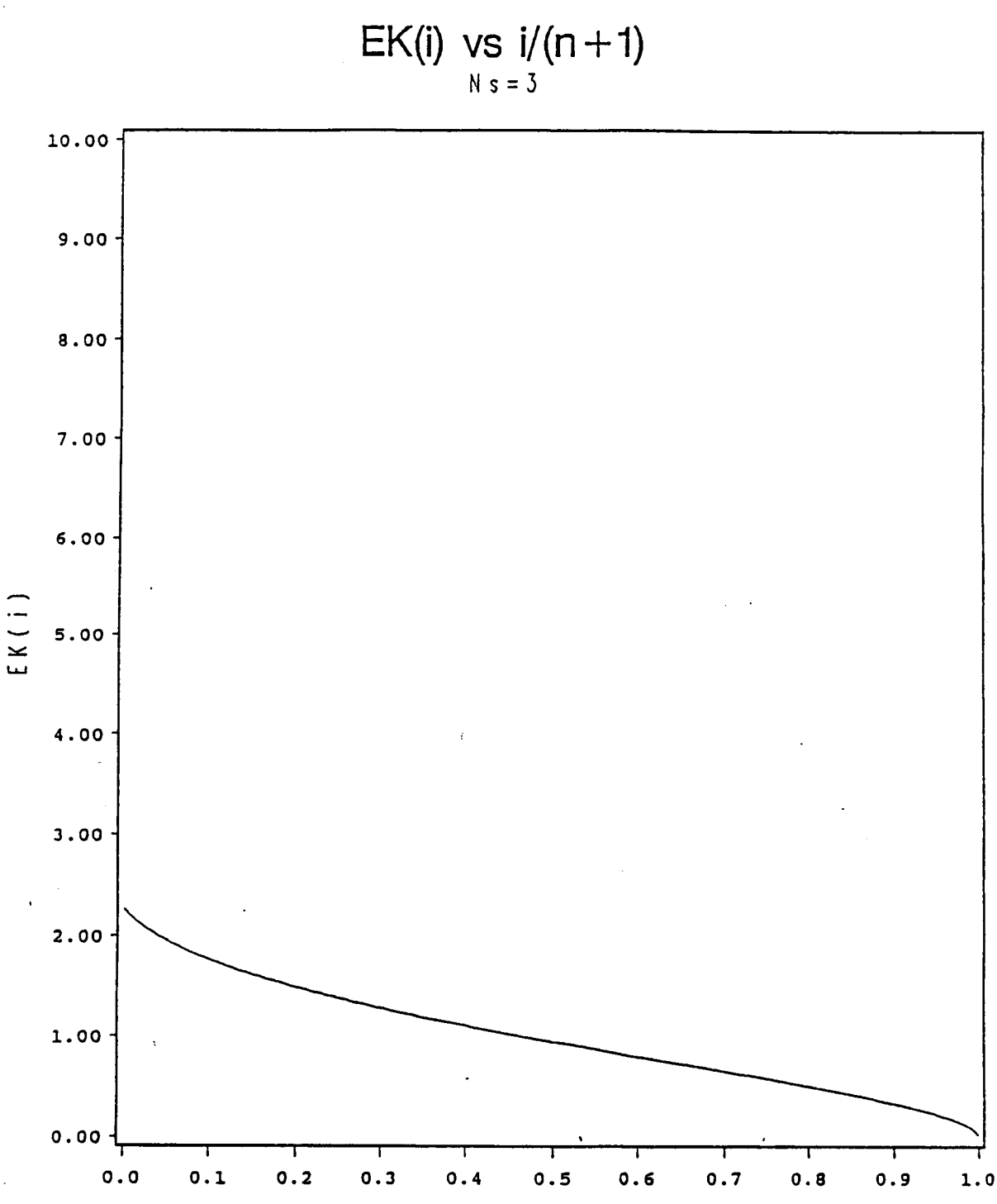

Figure 4.2: **The Distribution of EK when Stratum Size, ns, is 3**

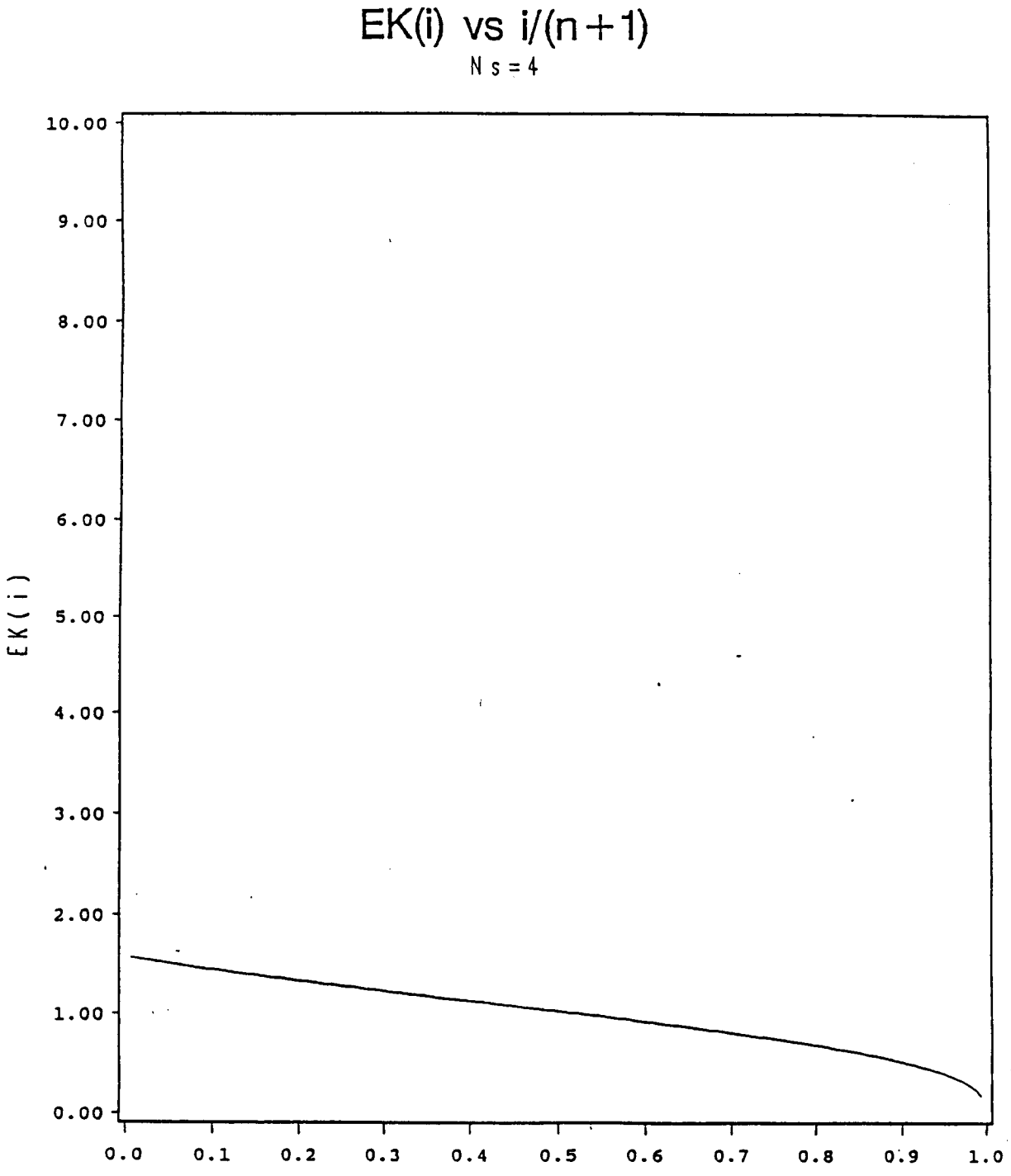

Figure 4.3: **The Distribution of EK when Stratum Size, ns, is 4**

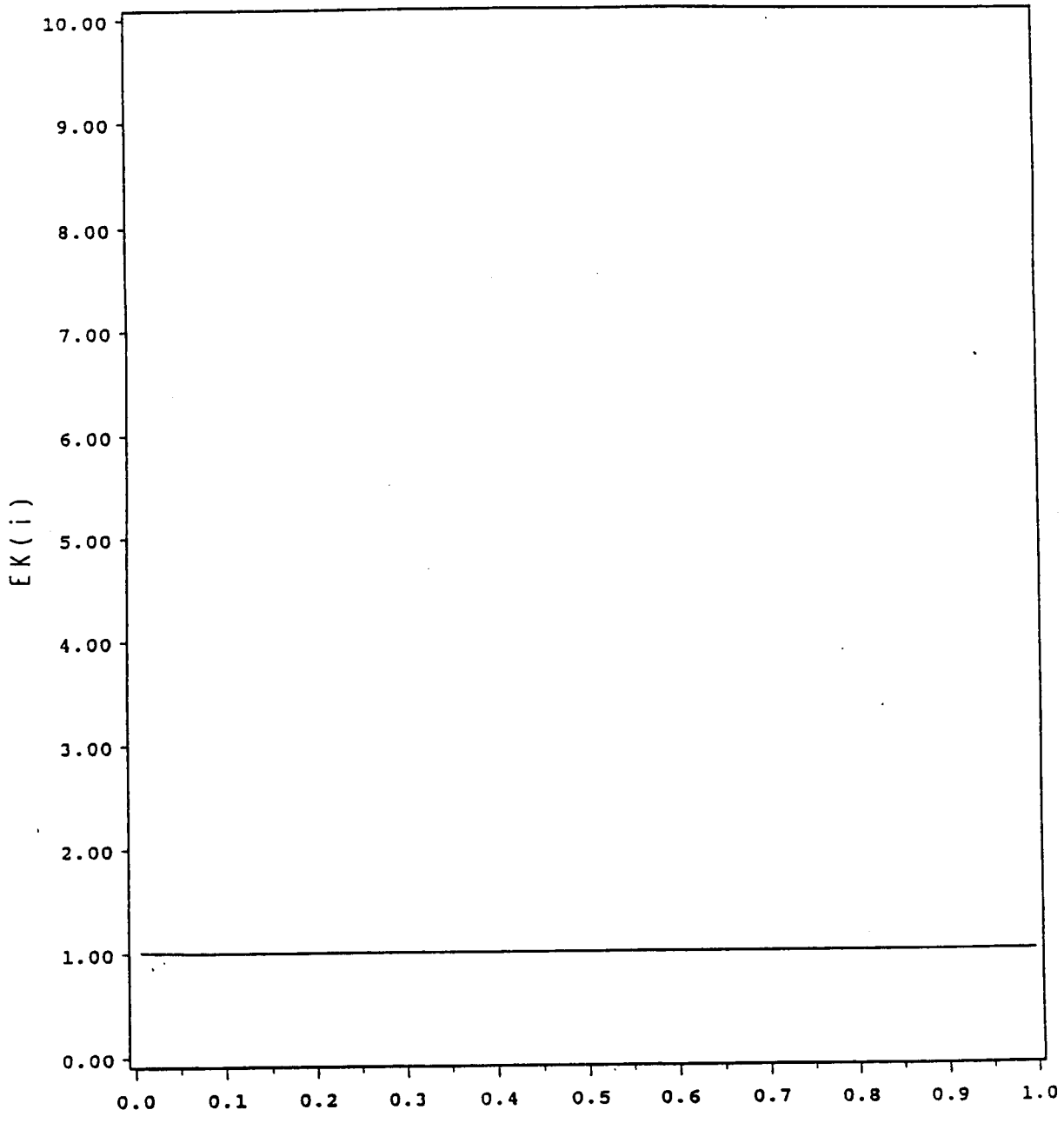

Figure 4.4: The Distribution of EK Corresponding to when Stratum Size, ns$\rightarrow\infty$

§ 4.9 *Distribution of Estimated Strata Effects*

Section 4.7 has shown that the distribution of the stratum effect is approximately normal if the stratum size is large. In section 4.8 the large stratum size assumption was relaxed and the distribution of the expected value of the stratum effect was derived. The function of the expected values of the stratum effects, the EK(i)'s, when stratum sizes are 2, 3 and 4 were numerically evaluated. This section uses the numerical results of the EK(i)'s computed from section 4.8 to generate the distribution of the stratum effect estimate. Recall from the log likelihood equations (4.7.1) and equation (4.8.2), that,

$$\frac{n_s}{\hat{K}_1} = \int_0^{r_{11}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{1n_s}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy$$

$$\frac{n_s}{\hat{K}_2} = \int_0^{r_{21}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{22}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{2n_s}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy$$

$$\vdots$$

$$\frac{n_s}{\hat{K}_{n_g}} = \int_0^{r_{n_g1}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{n_g2}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \cdots + \int_0^{r_{n_gn_s}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy \quad (4.9.1)$$

with the constraint that $E(\hat{K}_j) = \bar{K} = 1$. Notice that $\hat{K}_j$ can be written as a function of the EK(x)'s. Under the assumption of no stratum effect, the r_{ji}'s are *i.i.d.* Uniform (0,1), for j=1, 2, ···, n_g; i=1, 2, ···, n_s.

Define, $z = \int_0^r \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy$, then,

$$\frac{d\,r}{d\,z} = \int_r^1 \mathbb{E}K(x)\,dx,$$

and

$$f_z(z) = f_r(r)\cdot\frac{d\,r}{d\,z} = \int_r^1 \mathbb{E}K(x)\,dx. \tag{4.9.2}$$

Thus $\frac{n_s}{\hat{K}}$ can be written as the sum of the z's, for $n_s=2$, that is,

$$\frac{2}{\hat{K}_j} = \int_0^{r_{11}} \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 \mathbb{E}K(x)\,dx}\,dy = z_1 + z_2.$$

Similarly, for $n_s=3$,

$$\frac{3}{\hat{K}_j} = z_1 + z_2 + z_3,$$

and for $n_s=4$,

$$\frac{4}{\hat{K}_j} = z_1 + z_2 + z_3 + z_4.$$

Note that the distribution of $\frac{2}{\hat{K}_j}$ is the convolution of z_1 and z_2. Let $t_2=z_1+z_2$. The distribution of t_2 can be written as

$$f_{t_2}(t) = \int_0^t f_{z_1}(t-u) f_{z_2}(u)\,du. \tag{4.9.3}$$

Similarly, the distribution of $\frac{3}{\hat{K}_j}$ can be expressed as the convolution of t_2 and z_3, defined as t_3, and the distribution of $\frac{4}{\hat{K}_j}$ can be expressed as the convolution of t_3 and z_4. Therefore, the distribution of $\hat{K}_j$ for $n_s=2$ can be written in terms of t_2 as:

$\hat{K}_j = \frac{2}{t_2}$, for $n_s=3$, $\hat{K}_j = \frac{3}{t_3}$, and for $n_s=4$, it is $\hat{K}_j = \frac{4}{t_4}$. Since $\hat{K}_j = e^{\hat{\beta}_j}$, the distribution of stratum effect (β) can be easily obtained.

The distribution of z was computed by numerically integrating the $\mathbb{E}K(x)$ as shown in equation (4.9.2) using the results from Section 4.8. The distribution of f_{t_2} was obtained by numerically integrating the convolution of z_1 with z_2, as stated in equation (4.9.3). The distribution of t_3 was obtained by numerically integrating the convolution of t_2 with z_3, similarly, for the distribution of t_4.

The distributions of t_2, t_3 and t_4 corresponding to $n_s=2$, 3 and 4 are plotted in Figures 4.5, 4.6 and 4.7. For $n_s=2$, as was shown in Figure 4.5, the distribution is skewed to the right. For $n_s=3$, in Figure 4.6, two graphs are given. The dotted line is the distribution of $t_2(=z_1+z_2)$, and the solid line is the distribution of t_3. Note that t_3 is not as skewed as t_2. Figure 4.7 is for $n_s=4$, three graphs for t_2, t_3 and t_4 are given with the solid line indicating the distribution of t_4. The distribution of the stratum effect, $\hat{\beta}$, was obtained by the change of variable: $\hat{\beta} = \log(n_s) - \log(t)$, such that

$$f_{\hat{\beta}}(\beta) = n_s e^{-\beta} f_t(\beta).$$

The distributions of $\hat{\beta}$ are given for $n_s=2$, 3 and 4 in Figure 4.8, 4.9 and 4.10. These plots are symmetric about zero and as the n_s increased the variance became smaller. From the corollary in Section 4.7, when there are large stratum sizes, the distribution of the stratum effect estimate, $\hat{\beta}$, is Normal(0, $\frac{1}{n_s}$). The normal distributions, corresponding to $n_s=2$, 3 and 4, are considered as the limiting form of all three distributions as $n_s \to \infty$. Figure 4.11, 4.12 and 4.13 show the distribution of the strata effect each having the same expected value (zero) and variance $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ correspondingly, overlaid with the distribution of $\hat{\beta}$ computed numerically. The shape of the graphs is symmetric and fairly close to normal as predicted by asymptotic theory but the variance is bigger than would be predicted by asymptotic results.

§ 4.10 *Discussion*

A modified score test (MST) was presented in this chapter. The numerator of the MST is the score statistic of the proportional hazards model with indicator stratum covariables. The denominator of the MST is the proper variance estimate as proposed in Section 4.5. The properties of the MST was discussed in detail when there is no censoring. The distribution of the parameter estimates of stratum effect was derived, under the hypothesis of no stratum effect and the number of strata are proportional to sample sizes. The estimates are not consistent and traditional large sample theory does not apply. Simulation results are given in Section 4.6 and shown that the MST is always better than the SLRT and about 97% efficient for stratum size 10 when there is no stratum effect.

In summary, the MST is much more efficient than the SLRT and slightly less efficient than the LRT and the MLRT when there is no stratum effect and as efficient as the SLRT when there is "large" strata effects. For large stratum sizes, the MST is fully efficient for all sizes of strata effects and as efficient as the LRT and the MLRT when there is no stratum effect.

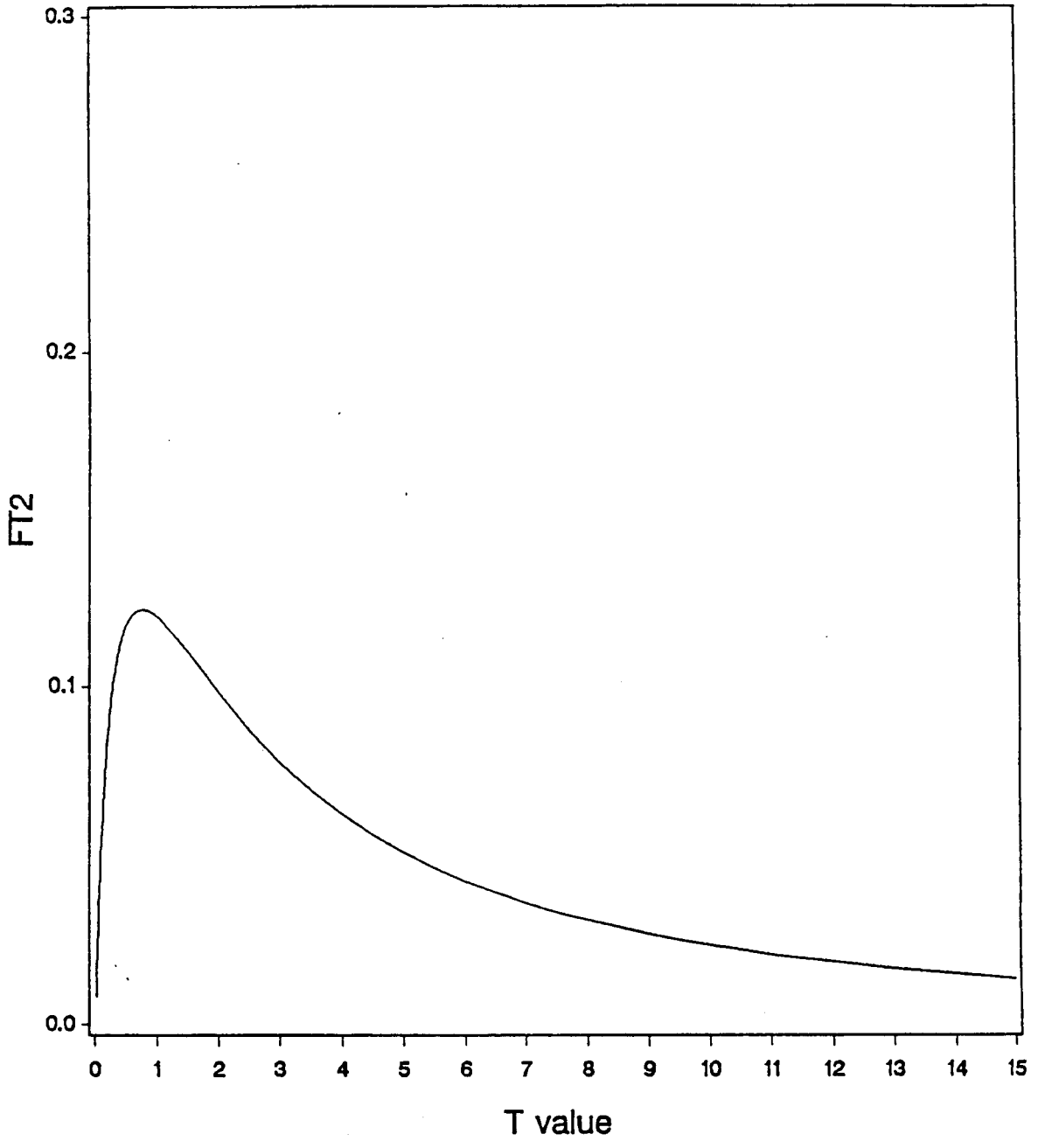

Figure 4.5: **The Distribution of t_2 when Stratum Size, ns, is 2**

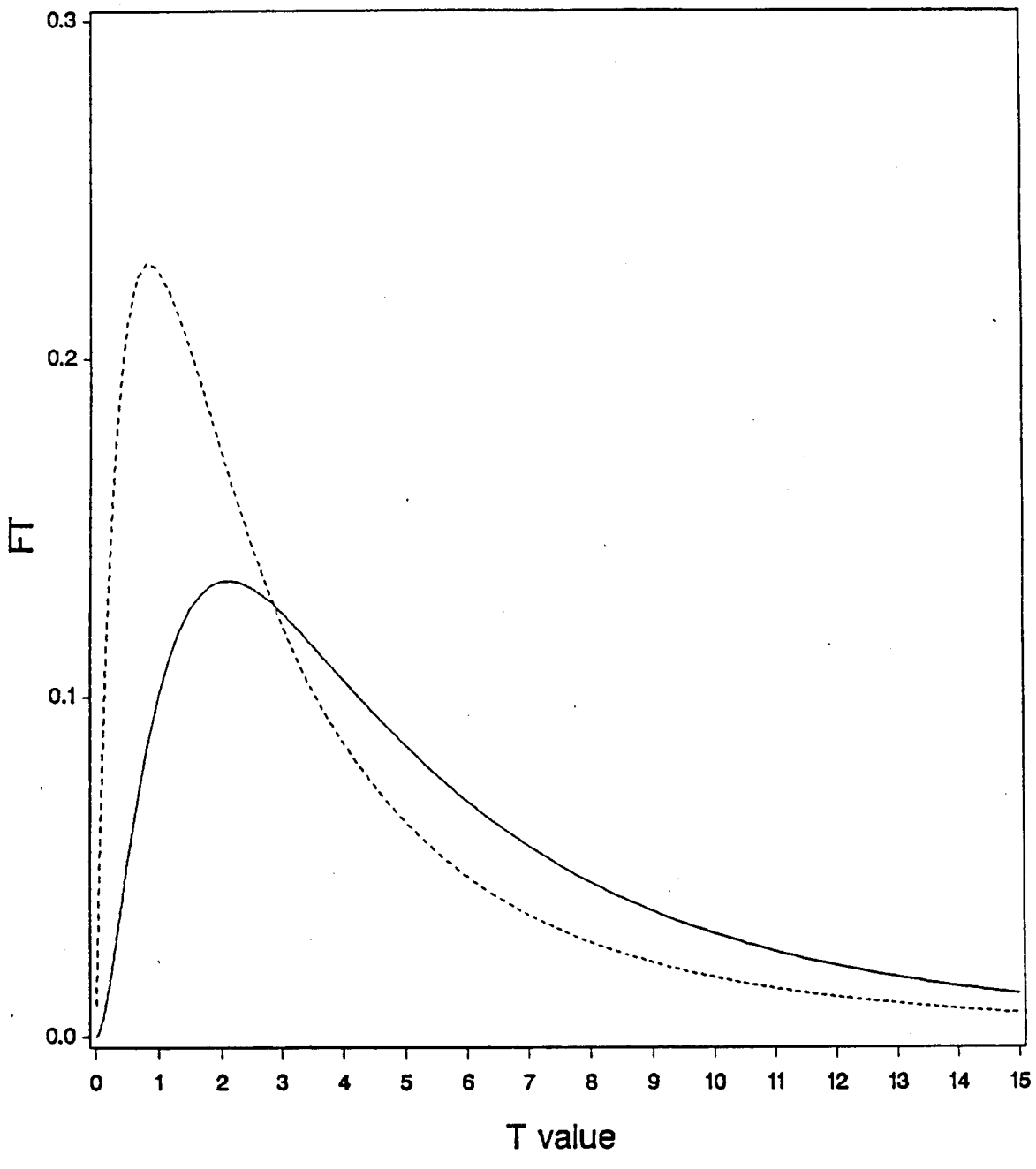

Figure 4.6: **The Distribution of t_2, t_3 when Stratum Size, ns, is 3**

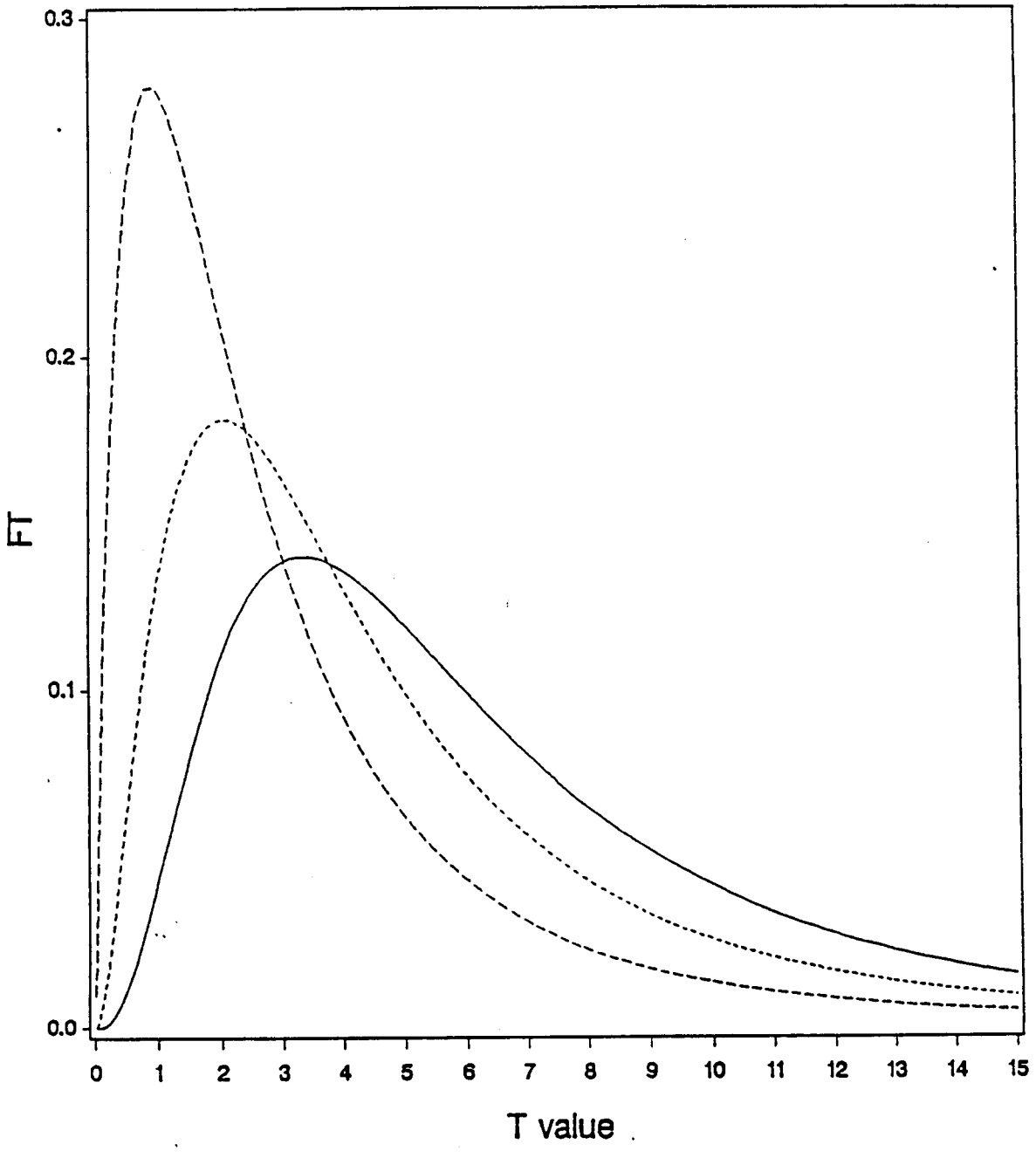

Figure 4.7: The Distribution of t_2, t_3 and t_4 when Stratum Size, ns, is 4

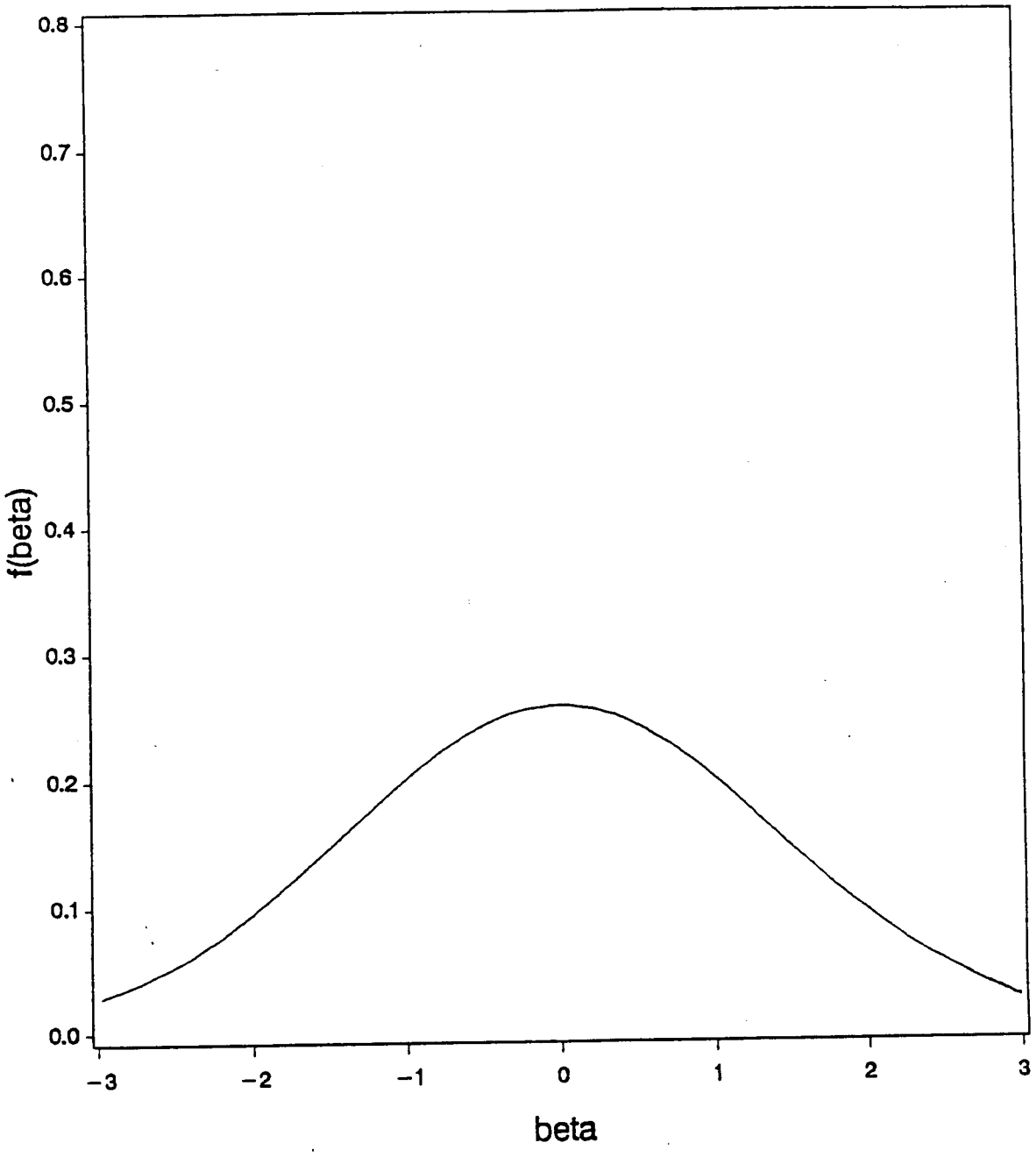

Figure 4.8: The Distribution of $\hat{\beta}$ when Stratum Size, ns, is 2

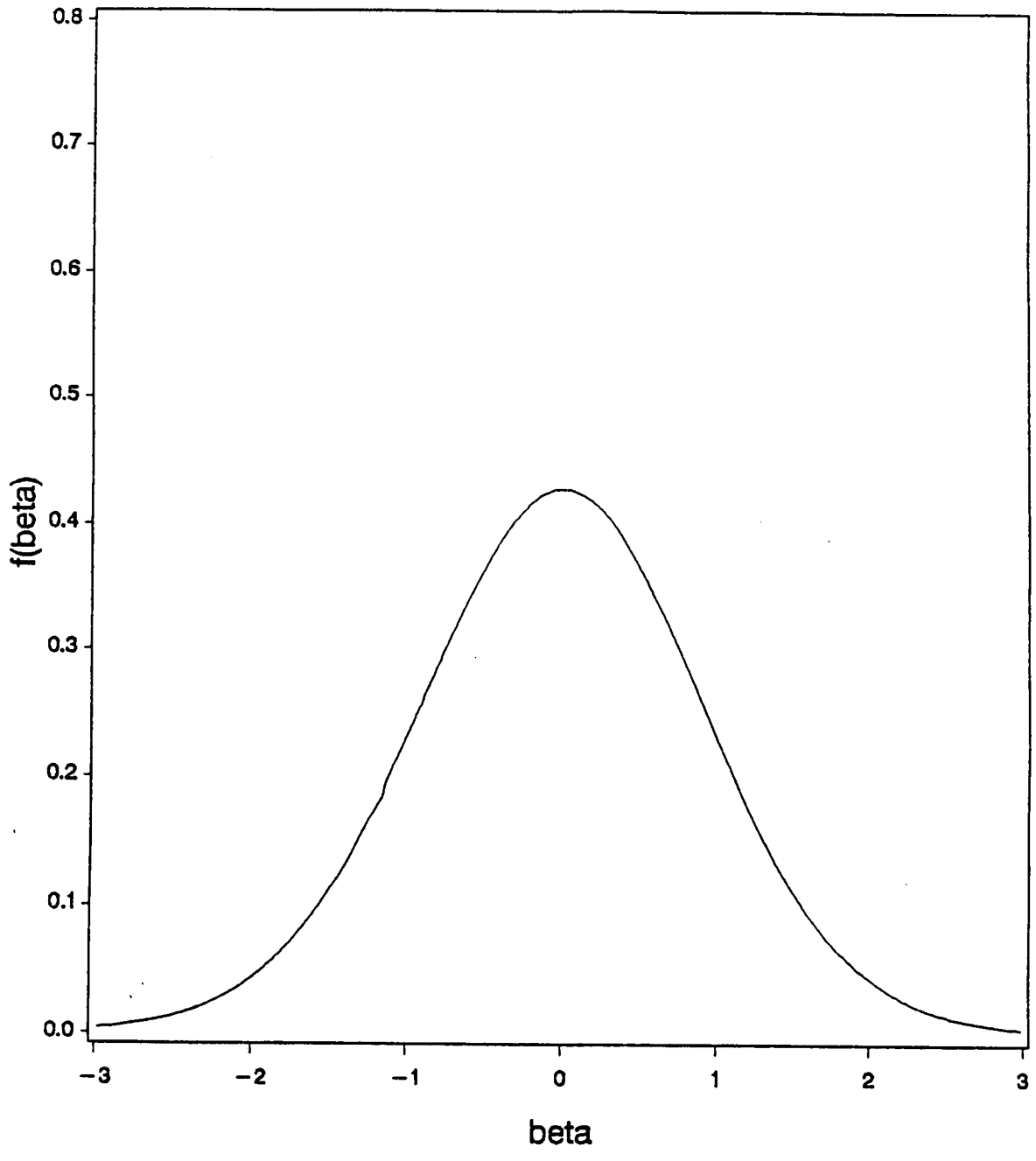

Figure 4.9: **The Distribution of $\hat{\beta}$ when Stratum Size, ns, is 3**

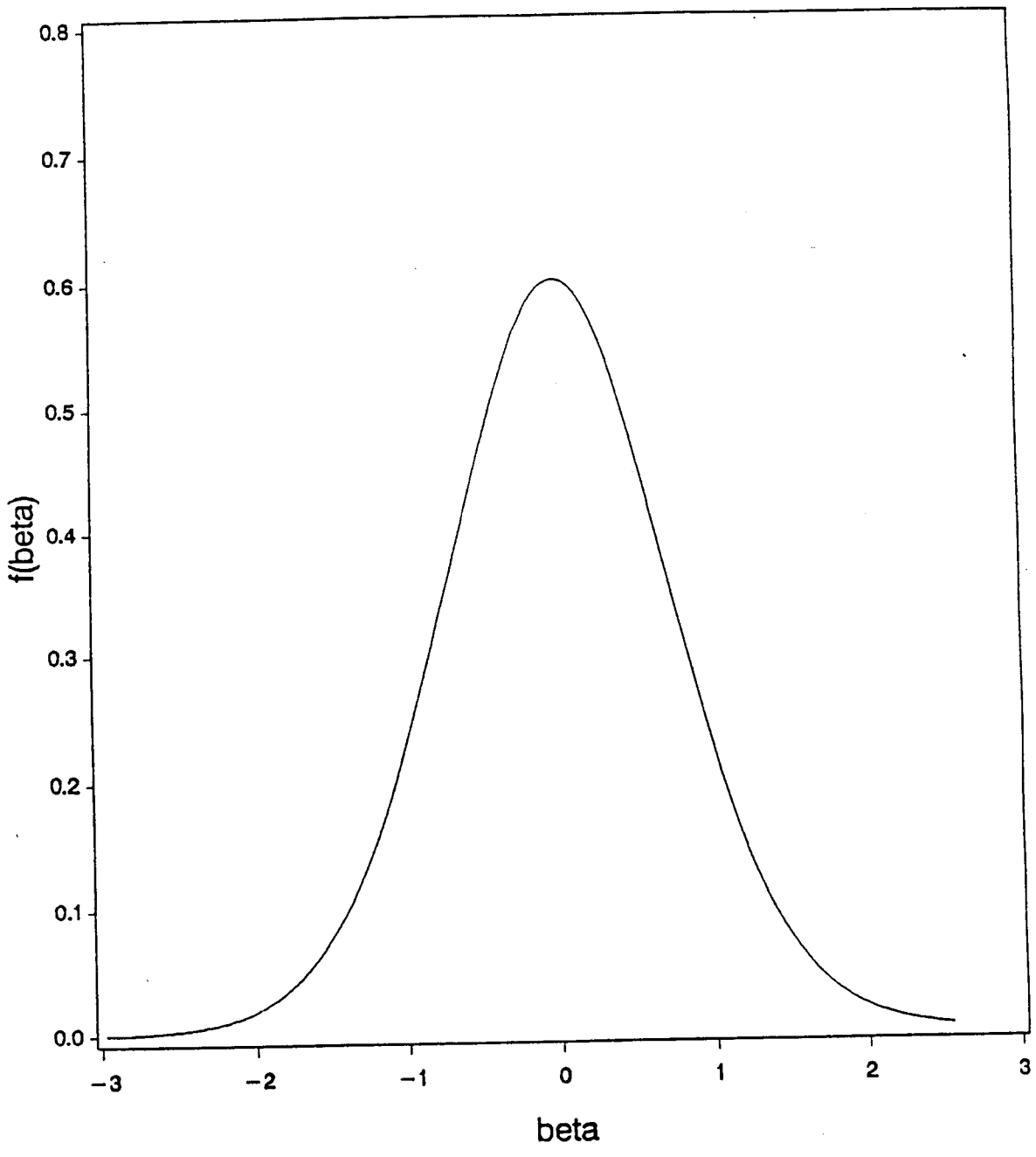

Figure 4.10: The Distribution of $\hat{\beta}$ when Stratum Size, ns, is 4

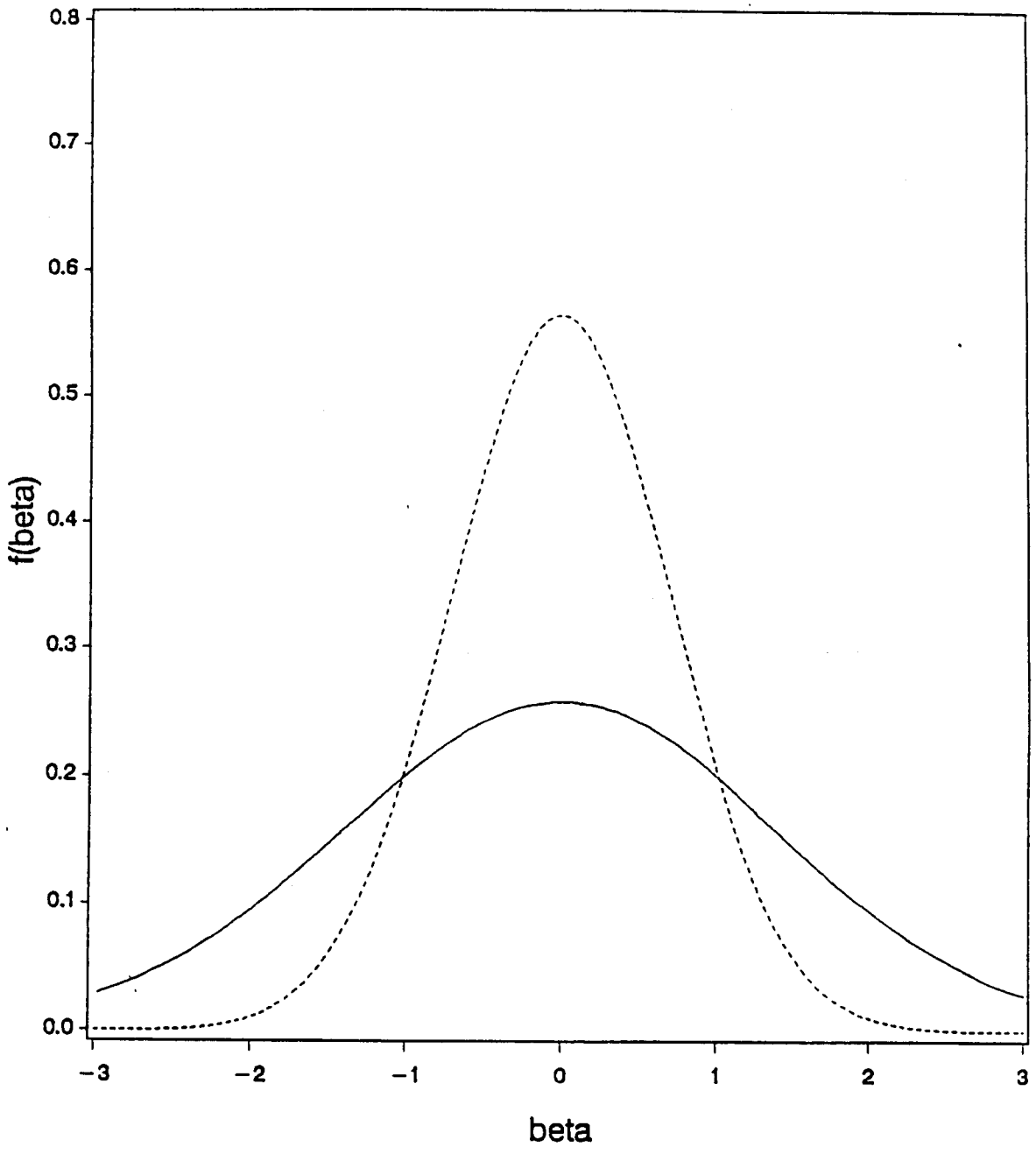

Figure 4.11: **The Distribution of $\hat{\beta}$ when Stratum Size, ns = 2, Overlay with the Distribution of $\hat{\beta} \sim$ Normal $(0, \frac{1}{2})$**

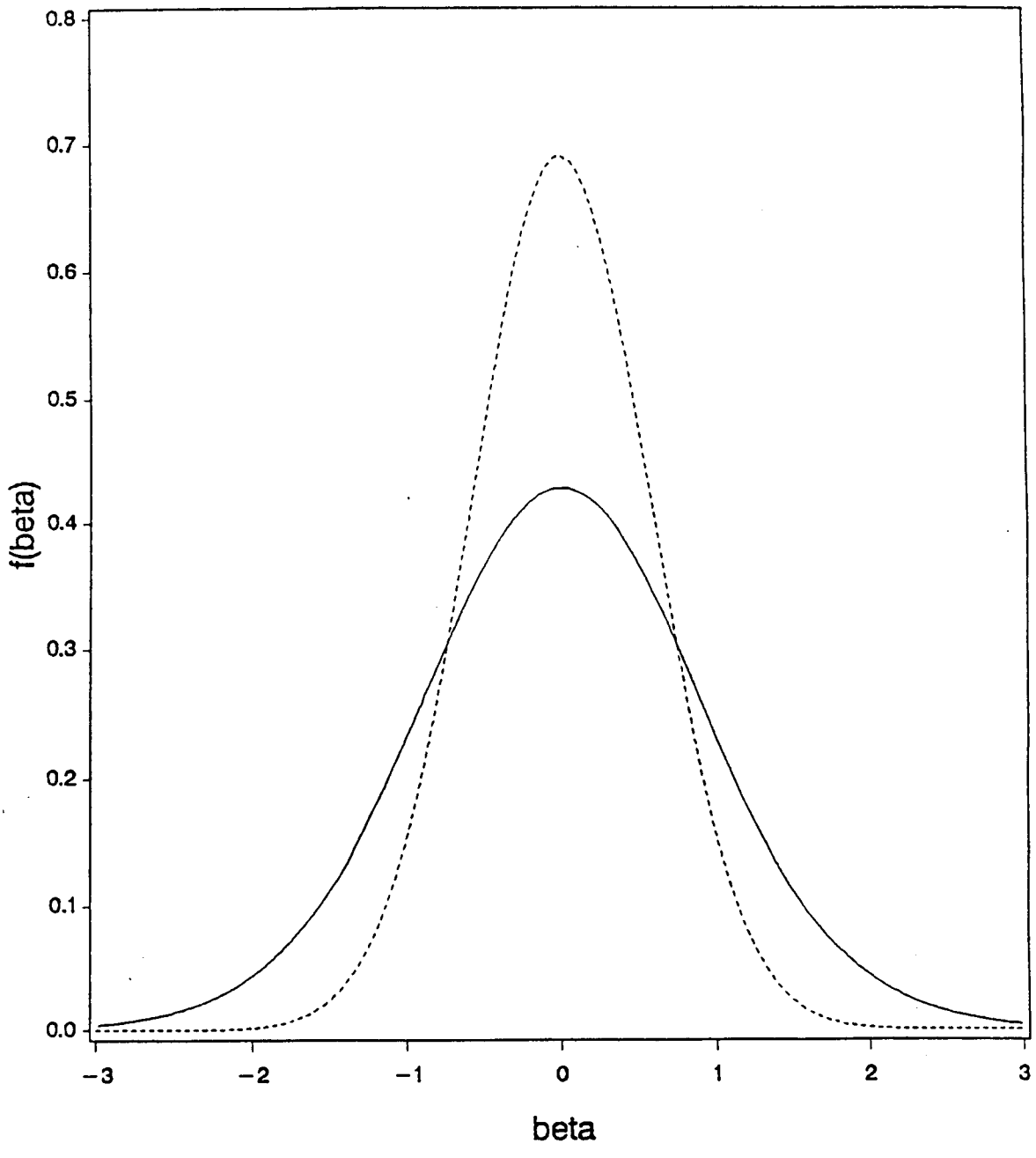

Figure 4.12: The Distribution of $\hat{\beta}$ when Stratum Size, ns = 3, Overlay with the Distribution of $\tilde{\beta} \sim$ Normal $(0, \frac{1}{3})$

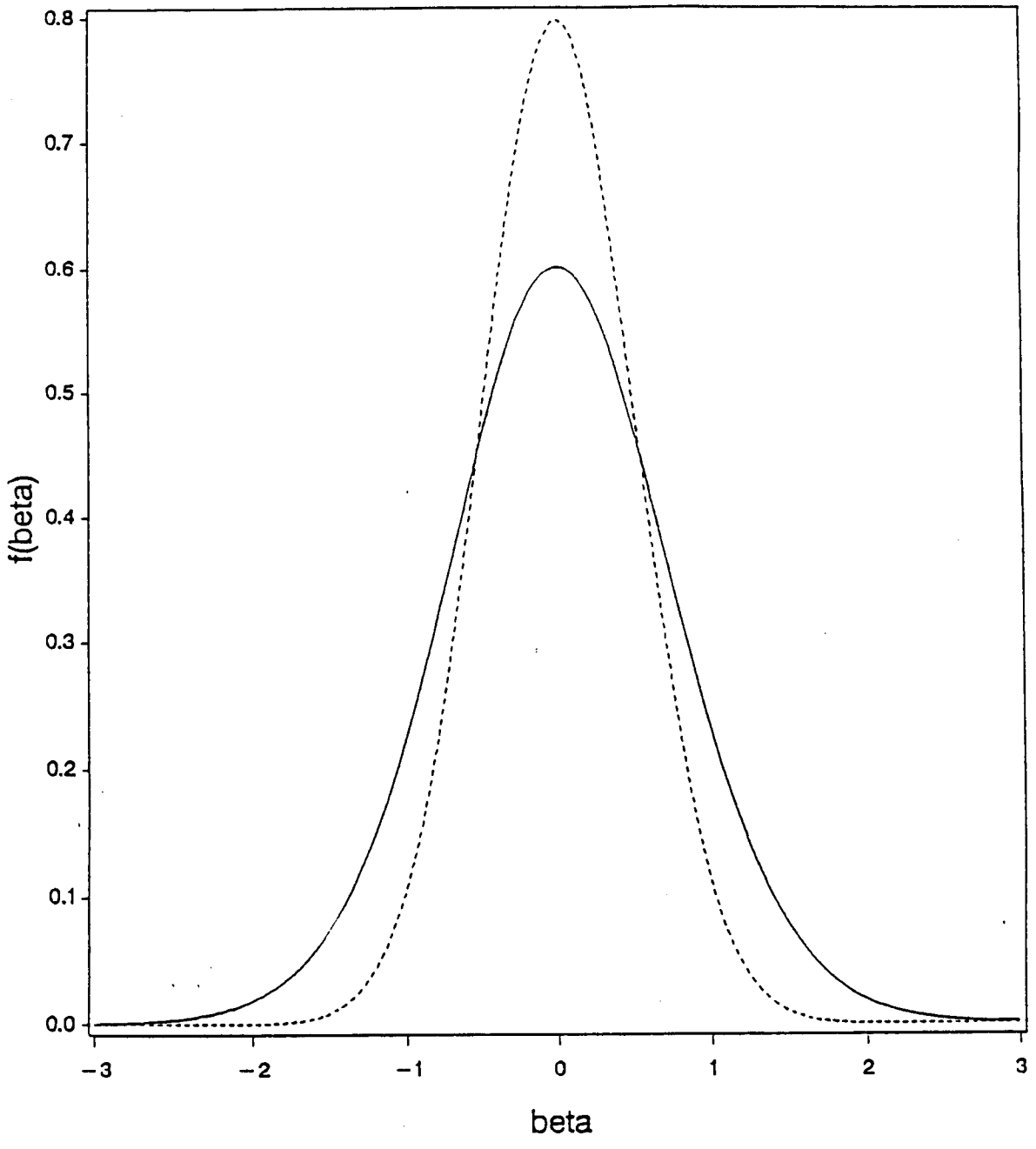

Figure 4.13: The Distribution of $\hat{\beta}$ when Stratum Size, ns = 4, Overlay with the Distribution of $\hat{\beta} \sim$ Normal $(0, \frac{1}{4})$

Chapter V

MODIFICATIONS FOR CENSORED DATA

§ 5. 1 *Introduction*

The modified score test (MST) was introduced in Chapter 4. The distribution and asymptotic properties when there is no censoring were examined. In this Chapter, the generalization of the MST for censored data will be discussed. Consideration is given to the case of Type I censoring at some fixed time c. The results for random Type I censoring can be easily extended.

As mentioned in previous chapters, censoring can be viewed as an example of incomplete or missing data. The failure time t_i of a censored subject is not observed, and is known only to exceed the censoring time c_i. In the case of fixed Type I censoring, the proposed test $\hat{S}$ is as given in equation (4.1.1) where δ_i=1 if the subject is uncensored, δ_i=0 otherwise. Some additional definitions are:

$\mathfrak{D}$: total number of deaths, $\mathfrak{D} = \sum_{i=1}^{n} E\delta_i$

$\mathfrak{D}_i$: number of deaths may occur at and after time i, $\mathfrak{D}_i = \sum_{l=i}^{n} E\delta_l$

$\mathfrak{D}_{ji}$: number of deaths may occur at and after time i in stratum j, $\mathfrak{D}_{ij} = \sum_{l=i}^{n} E\delta_{lj}$

d_{s_j}: number of death within stratum j

The $\underset{\sim}{\beta}$ is the parameter vector of strata effects and $\hat{\underset{\sim}{\beta}}$ is the solution to the following equations:

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} \delta_i \left[z_{1i} \frac{\sum_{j \in \Re_i} z_{1j} e^{\hat{\beta} z_j}}{\sum_{j \in \Re_i} e^{\hat{\beta} z_j}} \right]$$

$$= d_{s_1} - \sum_{i=1}^{n} \delta_i \frac{n_{1i} e^{\hat{\beta}_1}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0,$$

$$\frac{\partial \ell}{\partial \beta_m} = \sum_{i=1}^{n} \delta_i \left[z_{m_i} \frac{\sum_{j \in \Re_i} z_{mj} e^{\hat{\beta} z_j}}{\sum_{j \in \Re_i} e^{\hat{\beta} z_j}} \right]$$

$$= d_{s_m} - \sum_{i=1}^{n} \delta_i \frac{n_{m_i} e^{\hat{\beta}_m}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0,$$

$$\vdots$$

$$\frac{\partial \ell}{\partial \beta_{n_g}} = d s_{n_g} - \sum_{i=1}^{n} \delta_i \frac{n_{n_g i} e^{\hat{\beta}_{n_g}}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0.$$

§ 5.2 *The Effect of Censoring*

5.2.1 When There are "large strata effects"

As described in Section 4.2, the definition of "large" strata effects is: $e^{\beta_{j,i}} \rightarrow \infty$, $\forall\ j \leq i$, where $e^{\beta_{j,i}}$ is the hazard ratio of stratum j and stratum i. Using derivations similar to Section 4.2, it is easily shown that:

$$\hat{S} = n^{-1/2}\sum_i \delta_i \left[x_i - \frac{\sum_{\Re_i} x_j \, e^{\hat{\beta}_j}}{\sum_{\Re_i} e^{\hat{\beta}_j}} \right] \rightarrow n^{-1/2}\sum_i \delta_i \left[x_i - \frac{\sum_{\Re_i} x_j \, z_{ij}}{\sum_{\Re_i} z_{ij}} \right]$$

$$\rightarrow n^{-1/2} \sum_j \sum_i \delta_{ij} \left[a_{ij} - \frac{n_{Aij}}{n_{ij}} \right] \equiv SLRT_{NUM}$$

where,

$$z_{ij} = \begin{cases} 1 & \text{if } i^{th} \text{ and } j^{th} \text{ patients belong to the same stratum} \\ 0 & \text{otherwise} \end{cases}$$

n_{ij} = # of patients at risk at time i , stratum j

n_{Aij}= # of patients, group A, at risk at time i, stratum j

$$a_{ij} = \begin{cases} 1 & \text{death at time i, stratum j, group A} \\ 0 & \text{death at time i, stratum j, group B} \end{cases}$$

where δ_i=1 if the subject is uncensored, δ_i=0 otherwise.

5.2.2 Properties of $\hat{S}$ Conditioned on Estimated Strata Effects

Apply the definition given in equations (4.3.1a,b,c,d). Again, α is the parameter for treatment effect, $\hat{\alpha}$, is the parameter estimate. When $\hat{\beta}\underset{\sim}{z}$ is used in the model rather than $\beta\underset{\sim}{z}$, the parameter estimate for x is $\hat{\alpha}^*$ which is the solution to the following (incorrectly specified) log likelihood equation:

$$0 = \sum_i \delta_i \left[x_i - \frac{\sum_{\Re_i} x_j \, e^{\hat{\alpha}^* x_j + \hat{\beta}\underset{\sim}{z}_j}}{\sum_{\Re_i} e^{\hat{\alpha}^* x_j + \hat{\beta}\underset{\sim}{z}_j}} \right] \qquad (5.2.1)$$

Using methods similar to Morgan (1988, Appendix) on omitting a subset of covariate, $\hat{S}$ can be re-expressed as:

$$\hat{S} = n^{-1/2}\sum\delta_i\Big[\, x_i - U(t_i, \hat{\beta}\underset{\sim}{z})\Big] = S_1 + S_2 + S_3$$

where,

$$S_1 = n^{-1/2}\sum\delta_i\Big[\, x_i - U(t_i, \alpha x+\beta\underset{\sim}{z})\Big]$$

$$S_2 = n^{-1/2}\sum\delta_i\Big[U(t_i, \alpha x+\beta\underset{\sim}{z}) - U(t_i, \alpha^* x+\hat{\beta}\underset{\sim}{z})\Big]$$

$$S_3 = n^{-1/2}\sum\delta_i\Big[\, U(t_i, \alpha^* x+\hat{\beta}\underset{\sim}{z})-U(t_i, \hat{\beta}\underset{\sim}{z})\Big]$$

In the above, S_1 is the log likelihood equation for α under the true model, S_1 is asymptotically normally distributed with mean 0 and variance $p\,(1-p)(1-\pi(c))$, where $\pi(c) = \underset{\beta}{E}[\exp\{-\Lambda(c)\exp(\beta z)\}]$ is the proportion of censoring. If $p=\frac{1}{2}$, then,

$$S_1 \sim N(0, \tfrac{\pi(c)}{4}).$$

S_2 and S_3 can be similarly evaluated as in section (4.3). S_2 will converge to 0 and S_3 can be shown to converge to,

$$S_3 \rightarrow \alpha\, n^{1/2}\Big(\frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}\Big).\frac{[1-\pi(c)]}{4}$$

where the censored case of $\frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}$ is evaluated by the following Lemma.

Lemma 5.1. Let $\dot{\alpha}^{*c} = \frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}$ for the case of censoring, then

$$\dot{\alpha}^{*c} = \frac{\Lambda(c)V_0(c, \beta\underset{\sim}{z}) + \int_t \frac{\lambda(t)\Lambda(t)\, V_0(t, \beta\underset{\sim}{z}+\hat{\beta}\underset{\sim}{z})\, V_0(t, \beta\underset{\sim}{z})}{V_0(t, \hat{\beta}\underset{\sim}{z})}\, dt}{1-V_0(c, 0)}.$$

Proof By definition, α^*, the parameter value, is the solution to the asymptotic limit of the log likelihood equation (5.2.1) where the asymptotic limit is given by:

$$0 = \int_t \iint_{x\,z} \{ x \, d\, F_{TXZ}(t,x,z)\}\, I(t<c) - \int_t I(t<c) \iint_{x\,z} \{ d\, F_{T|XZ}(t,x,z)\, dF_{XZ}(x,z)\}$$
$$\cdot \frac{\iint_{x\,z} \{ x\, e^{\alpha^* x + \hat{\beta} z} d\, F_{TXZ}(-u,x,z)\}\, I(u>t)\}}{\iint_{x\,z} \{ e^{\alpha^* x + \hat{\beta} z} d\, F_{TXZ}(-u,x,z)\}\, I(u>t)\}} . \qquad (5.2.2)$$

with Type I fixed censoring at time c, when $x = \frac{1}{2}, -\frac{1}{2}$ this can be written as:

$$0 = \int_0^c \tfrac{1}{2}\, \lambda(t)\, \{ V_1(t, \tfrac{\alpha}{2}+\beta z) - V_2(t, -\tfrac{\alpha}{2}+\beta z)\}$$
$$- \lambda(t)\, \{V_1(t, \tfrac{\alpha}{2}+\beta z) + V_2(t, -\tfrac{\alpha}{2}+\beta z)\} \cdot \frac{\tfrac{1}{2}\{ V_1(t, \tfrac{\alpha^*}{2}+\hat{\beta} z) - V_2(t, -\tfrac{\alpha^*}{2}+\hat{\beta} z)\}}{\{V_1(t, \tfrac{\alpha^*}{2}+\hat{\beta} z) + V_2(t, -\tfrac{\alpha^*}{2}+\hat{\beta} z)\}} .$$

Subsequently, it can be shown that,

$$0 = \int_0^c \{ \tfrac{1}{4}\lambda(t)\Big(V_0(t, \beta z) - \Lambda(t)\, V_0(t, 2\beta z) + V_0(t, \beta z) - \Lambda(t)\, V_0(t, 2\beta z)\Big)$$
$$- \tfrac{1}{2}\lambda(t)\Big(V_0(t, \beta z) - \Lambda(t)\, V_0(t, 2\beta z) - V_0(t, \beta z) + \Lambda(t)\, V_0(t, 2\beta z)\Big) \cdot 0$$
$$- \tfrac{1}{2} \frac{\lambda(t)\, V_0(t, \beta z)}{V_0(t, \hat{\beta} z)} \cdot \Big(\dot{\alpha}^* V_0(t, \hat{\beta} z) - \Lambda(t)\, V_0(t, \beta z + \hat{\beta} z)\Big)$$
$$+ \frac{V_0(t, \beta z)}{4 V_0^2(t, \hat{\beta} z)} \cdot 0 \cdot 0 \}\, dt$$

$$= \int_0^c \lambda(t)\, \{ V_0(t, \beta z) - \Lambda(t)\, V_0(t, 2\beta z) + \frac{\Lambda(t)\, V_0(t, \beta z + \hat{\beta} z)\, V_0(t, \beta z)}{V_0(t, \hat{\beta} z)} \}\, dt$$
$$- \dot{\alpha}^* \int_0^c \lambda(t)\, V_0(t, \beta z)\, dt.$$

Note that,

$$\int_0^c \lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})\, dt = \int_0^c \lambda(t) \int e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, d\, F_Z(z)\, dt$$

$$= -\int_z \int_0^c \{-\lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, dt\, d\, F_Z(z)$$

$$= -\int_z [\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\,]\,|_0^c\, d\, F_Z(z)$$

$$= 1 - V_0(c, 0)$$

and

$$\int_0^c \lambda(t)\, \Lambda(t)\, V_0(t, 2\underset{\sim}{\beta}\underset{\sim}{z}) = \int_0^c \lambda(t)\, \Lambda(t) \int e^{2\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, d\, F_Z(z)\, dt$$

$$= \int_z \int_0^c \{\, \Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\}\, d\, \{-e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\}\, d\, F_Z(z)$$

$$= \int_z \{ \int_0^c \lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}} - \Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}\, e^{-\Lambda(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{z}}}\, \}\,|_0^\infty dF_Z(z)$$

$$= -\left[V_0(t, 0) + \Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})\,\right]|_0^c$$

$$= 1 - V_0(c, 0) - \Lambda(t)\, V_0(c, \underset{\sim}{\beta}\underset{\sim}{z})$$

Therefore,

$$\dot{\alpha}^{*c} = \frac{d\alpha^*}{d\alpha}\Big|_{\alpha=0}$$

$$= \frac{\Lambda(c)\, V_0(c, \underset{\sim}{\beta}\underset{\sim}{z}) + \int_t \dfrac{\lambda(t)\Lambda(t)\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z}+\hat{\underset{\sim}{\beta}}\underset{\sim}{z})\, V_0(t, \underset{\sim}{\beta}\underset{\sim}{z})}{V_0(t, \hat{\underset{\sim}{\beta}}\underset{\sim}{z})}\, dt}{1-V_0(c, 0)}. \qquad (5.2.3)$$

In the above, $\hat{S} = S_1+S_2+S_3$ in which S_1 is asymptotically normal with mean 0 and variance $\frac{1-\pi(c)}{4}$, S_2 converges to 0 and S_3 converges to $\alpha \cdot n^{1/2} \dot{\alpha}^{*c} \cdot \frac{[1-\pi(c)]}{4}$. Therefore, with Type I censoring fixed at time c, the score statistic for the proportional hazards model with indicator stratum variables conditioned upon the estimated stratum effects is asymptotically normally distributed with mean $\alpha \cdot n^{1/2} \dot{\alpha}^{*c} \cdot \frac{[1-\pi(c)]}{4}$ and

variance $\frac{1-\pi(c)}{4}$ where $\dot{\alpha}^{*c}$ is given in equation (5.2.3). Noted that the above results hold only when none of the strata effects are dominating, i.e. equation (4.3.5) holds. The variance of the numerator of the proposed test when conditioned upon the estimated stratum effect is similarly derived as in equation (4.3.6) and is given by:

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = \frac{p(1-p)}{n}\sum_{i=1}^{n}(w_i^c)^2 - \frac{p(1-p)}{n(n_s-1)}\sum_{i^*\neq j^*}\sum w_{i^*}^c w_{j^*}^c, \qquad (5.2.4a)$$

for balanced treatments within stratum, where i^* and j^* are in the same stratum and

$$\mathrm{Var}(\hat{S}|\hat{\beta}) = \frac{p(1-p)}{n}\sum_{i=1}^{n}(w_i^c)^2, \qquad (5.2.4b)$$

for the unbalanced cases where

$$w_i^c = \left(\delta_i - \hat{K}_{(i)}\sum_{j=1}^{i}\delta_j \frac{1}{\sum_{l\in\Re_j}\hat{K}_{(l)}}\right).$$

Thus, with Type I censoring, the numerator of the proposed test when conditioned upon estimated strata effects is normally distributed with mean $\alpha\cdot n^{1/2}\dot{\alpha}^{*c}.\frac{[1-\pi(c)]}{4}$ and variance given in equation (5.2.4a,b).

5.2.3 Asymptotic Distribution of $\hat{S}$

In the case of censoring, when there are "large" stratum effects (defined as all $\hat{K}$'s were diverging) the numerator of the proposed test, $\hat{S}$, converges to the numerator of the SLRT as was shown in Section 5.2.1. If none of the $\hat{K}$'s are diverging, with Type I censoring fixed at time c, the conditional distribution of $\hat{S}$ given $\hat{\beta}$ is normal with mean $\alpha\cdot n^{1/2}\dot{\alpha}^{*c}.\frac{[1-\pi(c)]}{4}$ and variance $\frac{1-\pi(c)}{4}$, where $\dot{\alpha}^{*c}$ is given in equation (5.2.3). The variance is independent of the stratum effect, therefore, the unconditional distribution of

$\hat{S}$ will have the same variance as the conditional one. If some of the strata effects are large, the variance of the proposed test when conditioned upon $\hat{\beta}$'s is given in equation (5.2.4).

5.2.4 Variance Estimate of $\hat{S}$

The variance estimate of the proposed test can be easily generalized to incorporate censoring. For balanced treatments within stratum, that is,

$$\hat{V}_c = \frac{p(1-p)}{n} \sum_{i=1}^{n} (W_i^c)^2 - \frac{p(1-p)}{n(n_s-1)} \sum_{i^* \neq j^*} \sum_{i^*} W_{i^*}^c W_{j^*}^c, \qquad (5.2.5a)$$

for unbalanced cases, it is,

$$\hat{V}_c = \frac{p(1-p)}{n} \sum_{i=1}^{n} (W_i^c)^2, \qquad (5.2.5b)$$

where $W_i^c = \left(\delta_i - \hat{K}_{(i)} \sum_{j=1}^{i} \delta_j \frac{1}{\sum_{l \in \Re_j} \hat{K}_{(l)}} \right)$.

Therefore, with Type I censoring, the proposed modified score test is given by:

$$MST = \frac{\left[n^{-1/2} \sum x_i W_i^c \right]^2}{\hat{V}_c}. \qquad (5.2.6)$$

5.2.5 Distribution of Stratum Effect Estimates

The distribution of $\hat{K}_j$ for the case of large stratum sizes with type I censoring is studied below. The assumptions made here are similar to those given in Section 4.7. The assumptions are:

(a) no strata effects, i.e. $K_1 = K_2 = \cdots = K_{n_g}$

(b) $\sum_{j=1}^{n_g} \mathfrak{D}_{ji} \hat{K}_j$ converges to $\sum_{j=1}^{n_g} \mathfrak{D}_{ji} \bar{K} = \bar{K} \cdot \mathfrak{D}_i$, $\forall$ i, $\bar{K}$, some constant

(c) There exist an $M < \infty$, such that $Pr(|\hat{K}_j| > M) \rightarrow 0$, for all j, as $n \rightarrow \infty$.

The distribution of strata effects for large stratum sizes in the case of Type I censoring can be similarly derived as the case of no censoring in Section 4.7. The proof of the following lemma is virtually identical as of Lemma 4.4. The only changes to make are as follows. The j^{th} stratum size denoted by n_{s_j} replaced by the number of death in that stratum denoted as d_{s_j}. The total sample size, n, replaced by the total number of deaths, d. The number at risk at time i, n_i, replaced by the number of deaths may occur at and after time i, $\mathfrak{D}_i$, and the number at risk at time i in stratum j, n_{ji}, replaced by $\mathfrak{D}_{ji}$. The results are as follows.

Lemma 5.2 If no strata effects, as $\mathfrak{D} \rightarrow \infty$, $\mathfrak{D}_{ji}$ and $\mathfrak{D}_{li}$ ($j \neq l$, j, l, stratum indicator) are independently distributed as binomial(d_{s_j}, $\frac{\mathfrak{D}_i}{\mathfrak{D}}$), where d_{s_j} is fixed.

Using the results of Lemma 5.2 and Lemma 4.5, the following theorem concerning the distribution of the $\hat{\beta}$'s can be similarly established for censoring cases.

Theorem 5.1: Under assumptions (a) - (c) in the above, the $\hat{\beta}_j$'s are independent and identically distributed with density function

$$f_Y(y) = \frac{d_{s_j}^{d_{s_j}}}{\Gamma(d_{s_j})} e^{- d_{s_j} y} e^{- d_{s_j} e^{-y}},$$

where $j = 1, 2, \cdots, n_g$.

Recall from Section 4.8 and Section 4.9, even in the case of no censoring, the distribution of strata effects can not be explicitly written in closed form. The distribution of strata effects must be numerically evaluated. In the case of fixed Type I censoring, the distribution of strata effects can also be numerically determined and the distribution is dependent on the number within stratum as well as the censoring distribution. Due to the vast amount of computer time involved in the numerical evaluation, the actual evaluation will not be included in this research, but the steps in performing this will be discussed below.

The distribution of the expected value of the stratum effect was examined numerically in Section 4.8.3. The results can be easily generalized in the case of Type I censoring. Using the same notation given in Section 4.8.3, and define $\mathfrak{D} = \sum_{i=1}^{n} \mathbb{E}\delta_i$. Conditioned upon the total sample size, n, the total number of deaths is a random variable following a binomial distribution. Let $p_d = \Pr(\mathfrak{D} \mid n)$ and $p_{d_i} = \Pr(d_{s_i} \mid n_{s_i})$. For $\mathbb{E}(d_{s_1}) = 2$,

$$\mathbb{E}\hat{K}_1(b) = \sum_{\mathfrak{D}=1}^{n} p_d \sum_{m_{12}=1}^{\mathfrak{D}} \frac{2}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{d} \mathbb{E}K(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{d} \mathbb{E}K(j)}}.$$

When $\mathbb{E}(d_{s_1}) = 3$,

$$\mathbb{E}\hat{K}_1(b) = \sum_{\mathfrak{D}=1}^{n} p_d \sum_{m_{12}=m_{13}=1}^{\mathfrak{D}} \sum^{\mathfrak{D}} \frac{3}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{\mathfrak{D}} \mathbb{E}K(j)} + \sum_{i=1}^{m_{12}} \frac{1}{\sum_{j=i}^{\mathfrak{D}} \mathbb{E}K(j)} + \sum_{i=1}^{m_{13}} \frac{1}{\sum_{j=i}^{\mathfrak{D}} \mathbb{E}K(j)}},$$

and when $\mathbb{E}(d_{s_1}) = 4$, equation (4.8.14) becomes,

$$\mathbb{E}\hat{K}_1(b) = \sum_{\mathfrak{D}=1}^{n} p_d \sum_{m_{1*}}^{\mathfrak{D}} \frac{4}{\sum_{i=1}^{b} \frac{1}{\sum_{j=i}^{\mathfrak{D}} \mathbb{E}K(j)} + \sum_{l=2}^{4} \sum_{i=1}^{m_{1l}} \frac{1}{\sum_{j=i}^{\mathfrak{D}} \mathbb{E}K(j)}},$$

where $\sum_{1*}^{\mathfrak{D}} = \sum_{m_{12}}^{\mathfrak{D}} \sum_{m_{13}}^{\mathfrak{D}} \sum_{m_{14}}^{\mathfrak{D}}$.

The numerical results of the $\mathbf{E}K(i)$'s computed from above can be use to generate the distribution of the $\hat{K}$'s and of the stratum effect estimates. Notice that the $\hat{K}$'s can be written as a function of the $\mathbf{E}\,K(i)$'s as was shown in equations (4.9.1). With Type I censoring equations (4.9.1) can be modified as:

$$\frac{d_{s_1}}{\hat{K}_1} = \int_0^{r_{11}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \cdots + \int_0^{r_{1d_{s_1}}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy$$

$$\frac{d_{s_2}}{\hat{K}_2} = \int_0^{r_{21}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \int_0^{r_{22}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \cdots + \int_0^{r_{2d_{s_2}}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy$$

$$\vdots$$

$$\frac{d_{s_{n_g}}}{\hat{K}_{n_g}} = \int_0^{r_{n_g1}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \int_0^{r_{n_g2}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy + \cdots + \int_0^{r_{n_gd_{s_{n_g}}}} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy \quad (5.2.7)$$

The constraint used here is $\mathbf{E}\,\hat{K}_j = \bar{K} = 1$. Under the assumption of no stratum effect, the r_{ji}'s are *i.i.d.* Uniform (0,1), for j=1, 2, ···, n_g; i=1, 2, ···, d_{s_j}. As defined in Section 4.9,

$$z = \int_0^{r} \frac{1}{\int_y^1 \mathbf{E}K(x)\,dx}\,dy$$

and

$$f_z(z) = f_r(r)\cdot\frac{d\,r}{d\,z} = \int_r^1 \mathbf{E}K(x)\,dx.$$

Then, the $\frac{d_{s_j}}{\hat{K}_j}$ in equations (5.2.7) can be written as the sum of the z's. For $d_{s_j} = 2$,

that is,

$$\frac{2}{\hat{K}_j} = \int_0^{r_{11}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy + \int_0^{r_{12}} \frac{1}{\int_y^1 EK(x)\,dx}\,dy = z_1 + z_2,$$

and so forth. Define $t_2 = z_1 + z_2$. The distribution of t_2 is the convolution of z_1 and z_2 and can be written as

$$f_{t_2}(t) = \int_0^t f_{z_1}(t\text{-}u) f_{z_2}(u)\,du.$$

Similarly, the distribution of $\frac{3}{\hat{K}_j}$ can be expressed as the convolution of t_2 and z_3, defined as t_3. The distribution of $\hat{K}_j$ for $E(d_{s_j})=2$ can be written in terms of t_2 as: $\hat{K}_j = \frac{2}{t_2}$, and for $E(d_s)=3$, $\hat{K}_j = \frac{3}{t_3}$, etc. The distribution of the stratum effect estimates, $\hat{\beta}_j$, can be obtained by change of variables, that is, $\hat{\beta}_j = \log(d_{s_j}) - \log(t)$, such that

$$f_{\beta_j}(\beta) = d_{s_j} e^{-\beta} f_t(\beta).$$

With the procedures discussed above, the distribution of the stratum effect estimates can be numerically evaluated in the case of Type I censoring.

§ 5.3 *Simulation Results*

A simulation study is conducted to examine the effect of censoring has on the efficiencies of the tests, the MST, the LRT, the SLRT and the MLRT, for stratum sizes of 2 and 10. The comparisons were made based on the AREs of each test and the AREs are computed from the means and variances of the test statistics determined over the simulations. The simulations were conducted as discussed in Section 4.6. In addition:

(1) Assume Type I fixed censoring where everyone in the study has potential censoring time t_0, which was chosen to be ∞, 0.3721, 0.1413, and 0.05 to correspond to 0%, 25%, 50% and 75% censoring.

(2) The stratum effects used in this simulation were generated to be $\beta \sim$ Uniform (0, A), where A=3 corresponds to hazard ratio of the interquartile range of 4.5.

The results of these simulations for different proportion of censoring are summarized in Table 5.1b. Note that the efficiency of the MST decrease slightly as the proportion of censoring increase while the efficiency of the LRT is lower for small amount of censoring and then increase as the proportion of censoring gets large. The efficiency of the SLRT remains constant with the changes of percent censoring and the efficiency of the MLRT increase with the increase of percent censoring. The MST seems performing very well over a wide range of different censoring patterns.

Table 5.1

Simulation results, balanced design
Assume $\beta \sim$ Uniform (0, 3)

Stratum size	% Censoring	ARE of LRT	ARE of SLRT	ARE of MLRT	ARE of MST
2	0 %	0.4865	0.4923	0.7102	0.6960
2	25%	0.5075	0.5755	0.8308	0.7830
2	50%	0.6260	0.6412	0.8671	0.7880
2	75%	0.7747	0.7139	0.8658	0.8119

Table 5.2

Simulation results, balanced design
Assume $\beta \sim$ Uniform (0, 3)

Stratum size	% Censoring	ARE of LRT	ARE of SLRT	ARE of MLRT	ARE of MST
10	0%	.7437	.7670	.7424	.9554
10	25%	.5226	.8734	.7979	.9939
10	50%	.6363	.8612	.8698	.9356
10	75%	.7667	.8576	.8868	.8703

Chapter VI

COMPUTER ALGORITHM AND EXAMPLES

§ 6.1 *Introduction*

In this Chapter a computer algorithm is developed for an iterative procedure to estimate the large number of parameters of strata effects. These parameter estimates are subsequently used to compute the modified score test. An example of comparing the time of the infection effect indwelling vascular catheters had in cancer patients, from a randomized clinical trial is provided. The clinical trial is a two to one matched design.

§ 6.2 *Computer Algorithm*

Once the parameter estimates for strata effects, $\hat{\beta}$, are computed, the proposed test statistic, $\hat{S}$ as given in equation (4.1.1), is easily calculated. The SAS PHGLM procedure may be used to compute the estimated strata effects for the proposed model by assigning n_g-1 indicator stratum variables. However, the method (modified Gauss-Newton method) used by PHGLM and BMDP2L to compute the maximum-likelihood estimates (MLEs) requires calculation and inversion of the matrix of second derivatives. This can be a very time consuming procedure even for a moderate number of strata, 10 to 20, and practically impossible for $n_g > 20$. When the number of strata increase with the sample size, the PHGLM failed to compute the estimates and test statistic. Others (Holt & Prentice, 1974; Clayton & Cuzick, 1985; Huster, Brookmeyer and Self, 1989)

are given the considerations of the proposed test statistic, but the computational difficulties in finding $\hat{\beta}$ appeared insurmountable.

An iterative procedure that does not involve the inversion of the covariance matrix is considered here. The MSTEST program listed in the Appendix C is written in FORTRAN and can be easily modified to different computer environments and available resources as described in the PROGRAM GUIDE (Appendix C). The rationale behind the program will be discussed below:

To find the maximum likelihood estimator, note that the objective function is $\ell = \log L$, the estimated stratum effect, $\hat{\underset{\sim}{\beta}}$, is the solution to the following equations,

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} \left(z_{1i} - \frac{\sum_{j \in \Re_i} z_{1j} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}} \right) = n_s - \sum_{i=1}^{n} \frac{n_{1i} e^{\hat{\beta}_1}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0$$

$$\frac{\partial \ell}{\partial \beta_2} = \sum_{i=1}^{n} \left(z_{2i} - \frac{\sum_{j \in \Re_i} z_{2j} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}} \right) = n_s - \sum_{i=1}^{n} \frac{n_{2i} e^{\hat{\beta}_2}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0$$

$$\vdots$$

$$\frac{\partial \ell}{\partial \beta_{n_g}} = \sum_{i=1}^{n} \left(z_{n_g} \frac{\sum_{j \in \Re_i} z_{n_g j} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}}{\sum_{j \in \Re_i} e^{\hat{\underset{\sim}{\beta}} \underset{\sim}{z}_j}} \right) = n_s - \sum_{i=1}^{n} \frac{n_{n_g i} e^{\hat{\beta}_{n_g}}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}} = 0$$

Let $\dot{\underset{\sim}{\beta}}$ be the arbitrary initial values for the estimates of $\hat{\underset{\sim}{\beta}}$, the estimated stratum effect, $\hat{\underset{\sim}{\beta}}$, can be approximated by the following equations. Type I fixed censoring can be easily incorporated, that is,

$$e^{\hat{\beta}_1} = \frac{d_{s_1}}{\sum_{i=1}^{n} \delta_i \frac{n_{1i}}{\sum_{j=1}^{n_g} n_{ji} e^{\dot{\beta}_j}}}$$

$$e^{\hat{\beta}_2} = \frac{d_{s_2}}{\sum_{i=1}^{n} \delta_i \frac{n_{2i}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}}}$$

$$\vdots$$

$$e^{\hat{\beta}_{n_g}} = \frac{d_{s_{n_g}}}{\sum_{i=1}^{n} \delta_i \frac{n_{n_g i}}{\sum_{j=1}^{n_g} n_{ji} e^{\hat{\beta}_j}}},$$

where d_{s_j} denotes the number of deaths in j^{th} stratum. The values, $\hat{\beta}$, computed from the first iteration are then used as the initial values, $\hat{\beta}$, in the second iteration, and so forth. Convergence will be assumed when the differences in any two parameter estimates between two successive iterations is less than 10^{-6}. To avoid the underflow/overflow problems, when an estimated stratum effect diverges to greater than 10^6, it is set to be 10^6, similarly when an estimated stratum effect diverges to less than 10^{-6}, it is set to be 10^{-6}. The program also computed the modified score statistic given in equation (4.1.1) using the previously obtained parameter estimates under the hypothesis of no treatment effect.

§ 6.3 *Example*

This section contains an example of applying the new proposed test statistic to a highly stratified randomization clinical trial. The example is kindly given by Dr. Peter Axelrod at the Department of Infectious Diseases, School of Medicine, Temple University. This retrospective clinical trial was designed to compare the rate of infection in two types of indwelling vascular catheters in cancer patients. One type of catheter, the Hichman, is tunneled under the skin but ends externally like a "heparin-lock". The

other type, the "infusaport" is entered by a needle through the skin. It was hoping that the "infusaport" would reduce the incidence of infection, but it was possible that it could increase infections. The study population consists of 47 patients with Hichman catheters and 94 with infusaports matched on tumor type, presence of metastases, age, gender, and time of catheter insertion.

This example is analyzed using the log rank test (LRT), the stratified log rank test (SLRT), with 47 strata, each stratum has 3 patients, and the proposed modified score test (MST). The results of testing for treatment effect using these three tests are summarized in Table 6.1. All three tests show a significant treatment effect at Type I error, $\alpha = 0.05$. The heavy censoring pattern shown in this particular data set may be the reason why the MST is less significant than the other two (the LRT and the SLRT). The score test computed from SAS/PHGLM, the numerator is the same as the numerator of the MST, only the variance estimates (denominator of the test) are different. Note that the computation time used to compute the MST using MSTEST program was reduced by a factor of 1:21 over the SAS/PHGLM procedure (7.21 cpu seconds vs. 157.91 cpu seconds). The time used to compute the LRT and the SLRT using SAS/PHGLM procedure took 8.76 and 8.12 cpu seconds respectively, which is longer than the MST computed from MSTEST program, even though the latter involved more parameter estimations than the former one did.

Table 6.1.

Catheter Trial Data Analyzed by the LRT, the SLRT and the MST
for testing the treatment effect
using SAS/PHGLM procedure and MSTEST program (Appendix C)

Test	Statistic	p-value	CPU† time (in seconds)
LRT/PHGLM	7.92	0.005	8.79
SLRT/PHGLM	7.90	0.005	8.12
SCORE/PHGLM	13.54	0.0002	157.91
MST/MSTEST	4.52	0.03	7.28

† the comparisons were made using a Micro VAX II.

Chapter VII

DISCUSSION AND FUTURE RESEARCH

§ 7.1 *Summary and Discussion*

A modified score test (MST) has been proposed as an alternative to the log-rank test (LRT), the stratified log-rank test (SLRT), and the modified log-rank test (MLRT) in the testing of treatment effect for highly stratified survival data. The proposed test is based on the score statistic of Cox's proportional hazards model with indicator strata covariates. A major focus of this research was to examine the asymptotic properties of the proposed test as compared with the existing tests. The MST has been shown to be much more efficient than the SLRT and slightly less efficient than the LRT and the MLRT when there is no stratum effect and as efficient as the SLRT when there is "large" stratum effect. Simulations show the MST is more efficient than the LRT, the SLRT and the MLRT for "moderate" strata effects. A computationally feasible FORTRAN program is developed for the computation of the proposed test. The computation involved an estimation of large number of parameters which is computationally infeasible using existing software packages such as SAS PHGLM or BMDP2L.

The distribution of parameter estimates, where the number of estimates increase proportional to the total sample sizes has special difficulties. It is common in clinical trials to have quite large numbers of small groups of observations where the group sizes are 2 (corresponding to matched pair design), 3 or 4. Maximum likelihood theory does not apply in estimating these parameters. Small sample theory based on asymptotic

conditional means of the strata estimates that used to derive explicit equations which then solved by numerical methods to derive the distribution of the estimates.

§ 7.2 *Future Research*

The proposed test is based on the assumption of proportional treatment effect and proportional stratum effect. Generalization can be made for the case of nonproportional treatment effects and nonproportional stratum effect.

The results based on proportional hazards model in this dissertation are easily generalizable to nonproportional hazards treatment effects. This can be done by generalizing to time dependent treatment effects in a time dependent proportional hazards model. This approach is similar to the extension of the LRT to the weighted version of the LRT.

The assumption that the multiplicative relation between hazard and stratum covariates includes most models for survival in the literature. For example, Gompertz (1825) observed an approximate exponential relation between the hazard function and age of the patient. Different doses of radiation on experimental animals was also found to operate multiplicatively on the hazard (Storer, 1965). If there is stratum effects that are not proportional, the performance of the MST is unclear and may become inefficient. Thus, attempts at verifying the proportionality assumption and research into the robustness of the MST should be considered.

The proposed test was based on the score for the proportional hazards model and therefore weighted all observations equally. Extension of this work could be performed allowing for unequal weights to the strata, for example, unequal weights which are proportional to some function of the stratum sizes. Extensions of this test to

unequal weights would have important applications in survey sampling, nonproportional sampling of risk groups, and pooling of data from strata with different quality control efforts.

The test proposed in this dissertation is for survival data when there is a single, possibly censored, survival time on each study subjects. Extensions to the case of competing risks (i.e. the failure on an individual may be one of several distinct types or causes) and the case of multivariate failure times (i.e. repeated occurrences of some similar events) could be developed.

In Section 4.6 to Section 4.8, the distribution of a larger number of estimated stratum effects were evaluated by numerical methods. Applications of the distributions of a large number of inconsistent estimates to other area's of statistical theory are of great interest and are highly recommended. This research can form the basis for further work on the developing of tests for heterogeneity in highly stratified data and estimates of the distribution of the heterogenious effects.

BIBLIOGRAPHY

Anderson, P. K. 1982. Counting processes and survival analysis. *Survival Analysis. Inst. Math. Stat. Lecture Notes - Monogr. Ser.* 2. Hayward, Calif.: Inst. Math. Stat. pp. 1 - 13

Andersen, P. K., Borgan, O., Gill, R. & Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Int. Statist. Rev.* 50, 219-258

Anderson, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: regression model for counting processes: A large sample study. *Ann. Stat.* 10: 1100-20

Armitage, P. (1959). The comparison of survival curves. *Proc. R Stat. Soc. Ser. A* 122:279-300

Armitage, P. & Gehan, E. (1974). Statistical methods for the identification and use of prognostic factors. *Int. J. Cancer*, 13, 16-26

Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension. *JASA*. 69, 1031-1037

Bernstein, D. and Lagakos, S. W. (1978). Sample size and power determination for stratified clinical trials. *Journal of Statistical Computing and Simulation* 8, 65-73.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* 30.

Breslow, N. & Crowley, J. (1974). A large sample study of the life-table and product limit estimates under random censorship. *Ann. Statist.* 2, 437-453

Bulpitt, C. J. (1983). *Randomized controlled clinical trials.* Martinus Nijhoff.

Byar, D. P., Simon, R. M., Friedewald, W. T., Schlesselman, J. J., DeMets, D. L., Ellenberg J. H., Gail, M. H. & Ware, J. H. (1976). Randomized clinical trials. *New Eng. J. Med.* 295. 74-80

Campbell, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* 68: 417-22

Campbell, G. (1982). Asymptotic properties of several nonparametric multivariate distribution function estimators under random censorship. *Survival Analysis. Inst. Math. Stat. Lecture Notes - Monogr. Ser.* 2. Hayward, Calif.: Inst. Math. Stat. pp. 243-56

Cederlof, R., Epstein, F. H., Friberg, L., Hrubec, Z. & Radford, E. (1971)/ Twin registries in the study of chronic disease. *Acta. Med. Scand. Suppl.* 523, 9-40.

Cheng, K. F. (1984). Asymptotically nonparametric tests with censored paired data. *Commun. Statist. Theor. Meth.*, 13(12), 1453-1470.

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations fo the proportional hazards model. *J. R. Statist. Soc. A* 148, pp. 82-117

Cox, D. R. (1958). *Planning of experiments.* New York: Wiley

Cox, D. R. (1964). Some applications of exponential ordered scores. *J. R. Statist. Soc.* B 34, 187-220.

Cox, D. R. (1970). *The analysis of binary data.* London: Methuen

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B* 34, 187-220

Cox, D. R. (1975). Partial likelihood. *Biometrika* 62, 269-276

Crowley, J. 1979. Some extensions of the logrank test. In *medical Informatics* ed. D. A. B. Lindberg, P. L. Reichetz, Lecture Notes 4, pp. 213-23. Berlin: Springer-Verlag

Crowley, J., Breslow, N. (1975). Remarks on the conservatism of $\sum(O-E)^2/E$ in survival data. *Biometrics* 31: 957-61

Crowley, J., Hu, M. (1977). Covariance analysis of heart transplant survival data. *J. Am. Stat. Assoc.* 72:27-36

Crowley, J., Johnson, R. A. eds (1982) *Survival Analysis. Inst. Math. Stat. Lecture Notes - Monogr. Ser.* 2. Hayward, Calif.: Inst. Math. Stat. 301 pp.

Efron, B. (1967). The two-sample problems with censored data. Proc. 5^{th} Berkeley Symp. 4, 831-853

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* 58, 403-417

Elandt-Johnson, R. C., Johnson, N. D. (1980). *Survival Models and Data Analysis.* New York: Wiley. 457 pp.

Feinstein, A. R. (1972). The purposes of prognostic stratification. *Clinical Pharmacology and Therapeutics, pp.* 285-289

Feinstein, A. R. and Landis, J. R. (1976). The role of prognostic stratification in preventing the bias permitted by random allocation of treatment. *J. Chron. Dis. Vol.* 29, *pp.* 277-284.

Fisher R. A., (1966). The design of experiments, 8th edition. *Edinburgh: Oliver and Boyd*

Friedman L. M., Furberg C. D., and DeMets D. L., *Foundamentals of clinical trials.* John Wright, Inc.

Gail, M. H. (1988). The effect of pooling across strata in perfectly balanced studies. *Biometrics* 44, 151-162.

Gail, M. H., Tan, W. Y. and Piantadosi, S. (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika, 75, pp.* 57-64.

Gail, M. H., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71, 3, 432-44

Gehan, E. A. (1965). A Generalized Wilcoxon Test for comparing arbitrarily singly-censored samples. *Biometrika*, 52, 203-223

Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124. Amsterdam: Mathematisch Centrum

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on the new mode of determining the value of life contingencies. *Phil. Trans. R. Soc.* A 115, 513-80

Green, S. B. and Byar, D. P. (1979). The effect of stratified randomization on size and power of statistical tests in clinical trials. *J. Chron. Dis. Vol.* 31, *pp.* 445-454.

Grizzle, J. E. (1987). Completely random assignment versus stratification in clinical trials. *New Perspectives in Theoretical and Applied Statistics.* 13-20.

Harrington, D. P., Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* 69:553-66

Hammond, E. C. (1964). Smoking in relation to mortality and morbidity. Findings in first thirty-four months of follow-up in a prospective study started in 1959. *J. Nat. Cancer Inst.* (*U.S.*) 32, 1161-1188

Hill, B. A. (1951) The clinical trial. *Br. Med. Bull.* 7: 278-282

Hijek, J. & Sidak, Z. (1967). *Theory of rank tests.* New York: Academic Press

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73, 3, pp. 671-678

Hsieh, F. Y., Crowley, J., Tormey, D. C. (1983). Some test statistics for use in multistate survival analysis. *Biometrika* 70; 111-19

Huster, W. J., Brookmeyer R. and Self S. G. (1989). Modelling paired survival data with covariates. *Biometrics* 45, 145-156

Johnson,, M. E., Tolley, H. D., Bryson, M. C. and Goldman, A. S. (1982). Covariate Analysis of survival data: a small-sample study of Cox's model. *Biometrics* 38, *pp.* 685-698.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The statistical analysis of failure time data.* Wiley

Kalbfleisch, J. D. & Sprott, D. (1970). Application of likelihood methods to models involving large numbers of parameters (with discussion). *J. R. Statist. Soc.* B 32, 175-208.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *JASA* 53, 457-481

Kempthorne O. (1977). Why randomize? *J. Statist. Plan. Infer.* 1, 1-25

Koch, G. G., Sen, P. K. & Amara, I. (1985). Log-rank Scores, Statistics, and Tests. *Encyclopedia of Statistical Sciences, Vol.* 5. 136-142

Latta, R. (1977). Generalized Wilcoxon statistics for the two-sample problem with censored data. *Biometrika* 64, 633-635

Lehmann E. L., (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Lagakos, S. W. (1981). The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika* 68: 93-98

Lagakos, S. C. & Schoenfeld, D. (1984). Properties of proportional-hazards score tests under miss-specified regression models. *Biometrics* 40, *pp.* 1037-1048.

Latta, R. B. (1981). A Monte Carlo study of some two-sample rank tests with censored data. *J. Am. Stat. Assoc.* 76:713-19

Liang, K-Y and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models *Biometrika*, 73, 1, pp. 13-22.

Lindley, D. V. and Singpurvalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *J. Appl. Prob.* 23, pp. 418-431.

Lininger, L., Gail, M. H., Grddn, S. B. and Byar, D. P. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika*, 66, 3, *pp.* 419-428.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports, vol.*50, *No.* 3.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* 22, 719-748.

Mantel, N. and Ciminera, J. L. (1979). Use of logrank scores in the analysis of litter-matched data on time to tumor appearance. *Cancer research* 39, *pp.*4308-4315.

Mantel, N. (1984). Pre-stratification or post-stratification. *letter to the editor of Biometrics*

Miller, R. G. Jr. (1981). *Survival analysis.* Wiley Series

Morgan, T. M. (1986). Omitting covariates from the proportional hazards model. *Biometrics* 42, 993-995

Morgan, T. M. (1988). Effect of censoring and correlation on omitting covariates from the Proportional Hazards Model. (submitted)

Peto, R. & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *J.R. Statist. Soc. A*, 135, II., 185-207

Peto, R. (1972b). Rank tests of maximal power against Lehmann-type alternatives. *Biometrika*, 59, 472-

Peto, R. and Pike, M.C. (1973). Conservatism of the approximation $\sum(O-E)^2/E$ in

the logrank test for survival data or tumor incidence data. *Biometrics*, 29, 579-

Peto, R. et al (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and Design. *Br. J. Cancer* 34, 585-612

Peto, R. et al (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and Examples. *Br. J. Cancer pp.* 1-39

Pike, M. C. (1966). A method of analysis of certain class of experiments in carcinogenesis, *Biometrics*, 22, 142-161

Popock, S. J. & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 31, 103-115

Prentice. R. L. (1978). Linear rank tests with right censored data. *Biometrika* 65: 167-79

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 69:331-42

Prentice, R. L. and Kalbfleisch, J. D. (1979). Hazard rate models with covariates. *Biometrics* 35: 25-39

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V. Jr., Flournoy, N., Farewell, V. T. et al (1978). The analysis of failure times in the presence of competing risks. *Biometrics* 34: 541-54

Prentice, R. L., Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Stat.* 11:804-13

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley

Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68, 316-319.

Schoenfeld, D. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics* 39, 499-503.

Schoenfeld, D. & Tsiatis, A. A. (1987). A modified log rank test for highly stratified data. *Biometrika*, 74, 1, *pp.* 168-175.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley

Storger, J. B. (1965). Radiation resistance with age in normal and irradiated populations of mice. *Radiation Res.* 25, 435-59

Tarone, T E. & Ware J. (1977). On distribution-free tests for equality of survival distribution. *Biometrika* 64, 156-160

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* 9, 93-108

W.H.O. (1966). The use of twins in epidemiological studies. *Acta Gen. et Med. Gem.*

15, 109-128.

Wei, L. J. (1980). A generalized Gehan and Gilbert test for paired observations that are subject to arbitrary right censorship. *JASA* 75, *pp.* 634-637

Woolson, R.F. and Lachenbruch, P.A. (1980). Rank tests of censored matched pairs. *Biometrika*, 67, 597-606.

APPENDICES

The APPENDICES consist of three parts, APPENDIX A, APPENDIX B and APPENDIX C. APPENDIX A contains three programs for the computation of the ARE's of the LRT, the SLRT and the MLRT for different assumptions about the distribution of stratum effect discussed in detail in Chapter 3. The program shown in APPENDIX B determines the distribution of the EK's described in Chapter 4. APPENDIX C includes the PROGRAM GUIDE and the program itself for computing the proposed modified score test. The integration subroutine used in MLRTV programs (in APPENDIX A) incorporates the DQDAGS subroutine which can be found in most computer algorithm books or in IMSL packages.

APPENDIX A

Computation of the ARE of the LRT:

```
C This is the FORTRAN program to compute the asymptotic mean of the LRT,
C where the stratum effect, beta is uniform(0,A) among strata.
C The parameters used are:
C    k:  no. of stratum
C    nr: no. of replications used in the integration subroutine
C    p:  proportion of subjects in treatment group
C    d:  stratum effect
C
C NOTE:  This program can be easily modified for beta is normal(0,sigma)
C        ( see MLRTV program ) and beta is dichotomous by redefining
C        the wt function.

      program LRTM2
      implicit double precision (A-H,O-Z)
      external f
      common /param/ a,p,xlamda,xlj,nr,k
      data one /1d0/
      data two /2d0/
C
      OPEN (UNIT=12)
C
      xlamda=one
      k=100
      nr=3600
      P=0.5
      do 111 na=0,70
        a=0.1d0*na
        sumbeta=0.0d0
        sum=0.0d0
      DO 50 JJ=1,k
        beta2=((dfloat(jj)-one)/(dfloat(k)-one))*a
        xlj=xlamda*dexp(beta2)
        call integ(F,V,nr)
       sumbeta=sumbeta+beta2
       sum=sum+V
   50  CONTINUE
      beta=sumbeta/dfloat(k)
      value=sum/dfloat(k)
      write (2,80) k,beta,a,value
      write(*,80)k,beta,a,value
   80  format (i5,2f5.2,f12.8)
C
  111  CONTINUE
      STOP
      END
```

```
C-----------------------------------------------------------------
      function f(x)
      implicit double precision (A-H,O-Z)
      common /param/ a,p,xlamda,xlj,nr,k
C
      data one /1d0/
C
      cn1=0.0
      cn2=0.0
      d1=0.0
      d2=0.0
C
      DO 50 I=1,K
        beta3=((dfloat(i)-one)/(dfloat(k)-one))*a
        ri=dexp(beta3)
        xli=xlamda*ri
        cn1=cn1+xli*x*dexp(-xli*x)
        d1=d1+dexp(-xli*x)
        cn2=cn2+xli*x*ri*dexp(-xli*x)
        d2=d2+ri*dexp(-xli*x)
 50   CONTINUE
C
      s1=cn1/d1
      s2=cn2/d2
      f=p*(1-p)*(one+s1-s2)*xlj*dexp(-xlj*x)
C
      RETURN
      END


      SUBROUTINE integ(F,V,nr)
      implicit double precision (a-h,o-z)
c     integer nr
      h=1.d0/dfloat(nr)
      x=-h/2.d0
      sum=0.d0
      do 10 i=1,nr
      x=x+h
      t=-dlog(x)
 10   sum=sum+F(t)/x
      v=sum*h
      RETURN
      END
```

Computation of the ARE of the SLRT:

```
C This is the program to compute the asymptotic mean, variance and
C ARE of the SLRT.  Note that the mean and variance only depend on
C stratum size, ns, and are independent of strata effects, beta.

      program SLRT
      implicit double precision (a-h,o-z)
      integer n,k,m
c
      open (UNIT=12)
c
      do 20 n=1,200
      sum=0
c
      m=2*n
      do 10 k=1,m
        sum=sum+real(m-k)/real(((m-1)*k*m))
 10   continue
      eff=1-sum
      x=0.25*eff
      write(12,70)
      write(12,80)m,eff,x
 70   format('  straum size, ARE, SLRT(variance):')
 80   format(i5,2F10.5)
 20   continue
      stop
      end
```

Computation of the ARE of the MLRT:

```
C This is the program to compute the variance of the MLRT,
C where the stratum effect, beta is normal (0, sigma) among strata.
C The parameters used are:
C    k:  no. of stratum
C    nr: no. of replications used in the integration subroutine
C    p:  proportion of subjects in treatment group
C    d:  stratum effect
C
C NOTE:  This program can be easily modified for beta is uniform(0, A),
C        and beta is dichotomous by redefining the wt function.

      program MLRTV
      implicit double precision (A-H,O-Z)
      external f1,f2
      common /param/ sig,p,xlamda,xlj,nr,k
      data one /1d0/
      data two /2d0/
C
      OPEN (UNIT=12)
C
      xlamda=one
      k=250
      nr=1600
      p=0.5
      do 111 na=0,20
        sig=0.1d0*na
        sum1=0.0
        sum2=0.0
      do 50 JJ=1,K
        beta1=((dfloat(jj)-one)/(dfloat(k)-one))*sig*7.0-3.5*sig
        x1=((dfloat(jj)-one)/(dfloat(k)-one))*7.0-3.5
        wt=dexp(-0.5*x1**2)/(dsqrt(2.0d0*3.14159265))
        xlj=xlamda*dexp(beta1)
      call integ(F1,V1,nr)
      call integ(F2,V2,nr)
      sum1=sum1+V1*wt
      sum2=sum2+V2*V2*wt
 50   continue
      value=P*(ONE-P)*(TWO*sum1-sum2)*7.0/dfloat(k)
      write(*,80)k,sig,value
      WRITE (2,80) K,sig,value
 80   FORMAT (I5,f5.2,F12.8)
C
 111  continue
      STOP
      END
```

```
C-----------------------------------------------------------------
      FUNCTION F1(X)
      implicit double precision (A-H,O-Z)
      common /param/ sig,p,xlamda,xlj,nr,k
      data one /1d0/
C
      s1=0.0
      s2=0.0
C
      do 50 I=1,K
        beta2=((dfloat(i)-one)/(dfloat(k)-one))*sig*7.0-3.5*sig
        x2=((dfloat(i)-one)/(dfloat(k)-one))*7.0-3.5
        wt=dexp(-0.5*x2**2)/(dsqrt(2.0d0*3.14159265))
        xli=xlamda*dexp(beta2)
        s1=s1+xli*x*wt
        s2=s2+xli*wt*x**xli
 50   continue
C
      s1=s1*7.0/float(k)
      s2=s2*7.0/float(k)
      f1=-dlog(s1)*(s2/s1)*x**(xlj-1)
C
      RETURN
      END
C-----------------------------------------------------------------
      FUNCTION F2(X)
      implicit double precision (A-H,O-Z)
      common /param/ sig,p,xlamda,xlj,nr,k
      data one /1d0/
C
      S1=0.0
      S2=0.0
C
      do 50 I=1,K
        beta3=((dfloat(i)-one)/(dfloat(k)-one))*sig*7.0-3.5*sig
        x3=((dfloat(i)-one)/(dfloat(k)-one))*7.0-3.5
        wt=dexp(-0.5*x3**2)/(dsqrt(2.0d0*3.14159265))
        xli=xlamda*dexp(beta3)
        S1=S1+xli*x*wt
        S2=S2+xli*wt*x**xli
 50   continue
C
      f2=(s2/s1)*x**(xli-1)
C
      RETURN
      END
```

```
C-----------------------------------------------------------------
      SUBROUTINE integ(F,V,nr)
      PARAMETER (TOL=1.0D-10,RELTOL=0.0)
      implicit double precision (a-h,o-z)
      real*16 sum
      EXTERNAL F
      h=1.d0/nr
      y=h/2.d0
      sum=0.q0
      do 10 i=1,256
      x=i*h-y
      t=F(x)
      sum=sum+t
 10   CONTINUE
      do 20 i=nr-255,nr
      x=i*h-y
      t=F(x)
      sum=sum+t
      CONTINUE
 20   v1=sum*h
      a=256.0d0/nr
      b=(nr-256.0d0)/nr
      call dqdags(f,a,b,tol,reltol,v2,errest)
      v=v1+v2
      RETURN
      END
```

APPENDIX B

```
C          ***This is the CORNELLF version ***
C
C     This is a program to calculate sets of EKs for an NS of 3.
C
C    > The array W is defined for holding V(I)+V(I) in the first N
C      elements, and V(I)+V(J) for I<>J in the remaining elements.
C      This allows for cutting out 1/2 of the additions and reduces
C      summation from a double DO loop to a single loop with a very
C      long vector.
C    > The read in starting values are renormalized before starting
C      the iterations.
C    > The direction of the summations is changed to try and add from
C      the smallest to the largest, in order to decrease roundoff due
C      to shifting for alignment.
C    > Code to calculate the sum of squares of (fnew(i)-fold(i)) is
C      put back in. This is to hopefully give better indication on
C      procedure is converging or diverging or oscillating.
C    > Code is added to dump current iteration results if program is
C      in danger of running out of time.

      program EK2NS3
      parameter (N=639,TOL=1.0D-8)
      parameter (nn=(n*(n+1))/2)
      parameter (TLIM=8.4D2)
      implicit double precision (a-h,o-z)
      dimension fold(n),fnew(n),u(n),v(n),w(nn)
      integer ssdct
      dimension time1(4),time2(4),time3(4)
      OPEN (UNIT=10)
      read(10,1000)(fold(i),i=1,n)
1000  format(d20.14)
      CLOSE(UNIT=10)
      sum=0.0
      do 2 i=n,1,-1
        sum=sum+fold(i)
    2 continue
      do 3 i=1,n
      fold(i)=fold(i)*(dfloat(n)/sum)
    3 continue
      OPEN (UNIT=11)
      rn2=1.0d0/dfloat(n)**2
      ssdct=0
      ssold=dfloat(n**3)
C  Initialize the job timer
      call vmtime(time1)
C  and iteration timer from job timer
        TIME3(1)=TIME1(1)
        TIME3(2)=TIME1(2)
```

```
      TIME3(3)=TIME1(3)
      TIME3(4)=TIME1(4)
      l=0
    5 continue
      l=l+1

      u(n)=fold(n)
      do 10 i=n-1,1,-1
      u(i)=u(i+1)+fold(i)
   10 continue
      do 20 i=1,n
      u(i)=1.0/u(i)
   20 continue
      v(1)=u(1)
      do 30 i=2,n
      v(i)=v(i-1)+u(i)
   30 continue
      do 31 i=1,n
      w(i)=v(i)+v(i)
   31 continue
      k=n
      do 33 i=1,n-1
      do 32 j=i+1,n
      w(k-i+j)=v(i)+v(j)
   32 continue
      k=k-i+n
   33 continue
C  This is the central time intensive section of the program
      do 40 k=1,n
      fnew(k)=0
      do 51 i=n,1,-1
      fnew(k)=fnew(k)+3.0d0/(v(k)+w(i))
   51 continue
      do 52 i=nn,n+1,-1
      fnew(k)=fnew(k)+6.0d0/(v(k)+w(i))
   52 continue
      fnew(k)=fnew(k)*rn2
C  Now renormalize the fnew to sum to n
   40 CONTINUE
      sum=0.0d0
      do  43 i=n,1,-1
        sum=sum+fnew(i)
 43   continue
      do  44 i=1,n
        fnew(i)=fnew(i)*(dfloat(n)/sum)
 44   continue
C  Calculate the sum of squares of changes from fold to fnew.
C  If this has not decreased then issue a warning. If this happens five
C  times in a row then kill the job on suspicion of numeric problems.
C  if the sum decreases then reset the counter that keeps track of incre
      ssnew=0.0
      do 45 i=1,n
   45 ssnew=ssnew+(fnew(i)-fold(i))**2
      if(ssnew.ge.ssold) then
       write(11,9001)l,ssold,ssnew
```

```
      ssdct=ssdct+1
 9001 FORMAT('  For iteration ',I4,' SS Changes grew from ',D20.14,'to'
     *,D20.14)
      if(ssdct.gt.5) then
       write(11,9002)
 9002 FORMAT(' SS changes has not decreased in the last five iterations'
     * ,'* PROGRAM TERMINATED ON SUSPICION OF NUMERICAL PROBLEMS*')
      WRITE(11,9003)L,SSOLD,SSNEW

 9003  FORMAT(' Iteration',I4,' SSOLD=',D20.14,' SSNEW=',D20.14)
       goto 900
       endif
      else
       ssdct=0
      endif
C  Now calculate time used and check against timelimit.
C  If over timelimit then write message, dump results
C  and quit.
      call vmtime(time2)
      if(time2(3)-time1(3).gt.tlim) then
       write(11,2002)
 2002 format('  program terminating at end of specified timelimit ')
      WRITE(11,9004)L,SSOLD,SSNEW
 9004 FORMAT( ' Last iteration was ',I4,' SSOLD=',D20.14,' SSNEW=',
     * D20.14)
      goto 900
      endif
C  Now for testing output iteration times if # of iterations<=5
      if(I.le.5)then
       WRITE(11,9005)L,(TIME2(1)-TIME3(1)),(TIME2(2)-TIME3(2)),
     *  (TIME2(3)-TIME3(3)),(TIME2(4)-TIME3(4))
 9005 FORMAT('  iteration ',I4,' wallclock time =',d20.14,' sec.'/
     *  17x,' virtual cpu time=',d20.14,' sec.'/
     *  17x,' total  cpu time=',d20.14,' sec.'/
     *  17x,' vector  cpu time=',d20.14,' sec.')
       time3(1)=time2(1)
       time3(2)=time2(2)
       time3(3)=time2(3)
       time3(4)=time2(4)
      endif
C Now test the absolute changes between fnew and fold to see if they all
C satisfy the tolerance criterion for stopping.
      do 65 i=1,n
      if (abs(fold(i)-fnew(i)).gt.tol) goto 67
   65 continue
      write(11,2001)
 2001 format(' Successful convergence achieved ')
      OPEN(UNIT=12)
      WRITE(12,1000)(FNEW(1)-.5D0*(FNEW(1)-FNEW(2))),
     *            (FNEW(I),.5D0*(FNEW(I)+FNEW(I+1)),I=1,N-1),
     *             FNEW(N),FNEW(N)+.5D0*(FNEW(N)-FNEW(N-1))
      CLOSE(UNIT=12)
       WRITE(11,9005)L,(TIME2(1)-TIME1(1)),(TIME2(2)-TIME1(2)),
     *  (TIME2(3)-TIME1(3)),(TIME2(4)-TIME1(4))
      go to 900
```

```
   67 do 70 i=1,n
      fold(i)=fnew(i)
   70 continue
      goto 5
900    CONTINUE
      WRITE(11,1001) I
 1001 FORMAT(I5)
      WRITE(11,1000)(FNEW(I),I=1,N)
      CLOSE(UNIT=11)
      stop
      end
```

APPENDIX C

Program for computing the MST:

```
C MSTEST is a FORTRAN program used to compute the proposed modified
C score test.  The program assume that the data are sorted on survival
C time in ascending order with failures (deaths) preceding censorings
C in the event of ties.  The data file is assume to be an ASCII filel,
C each record contains TIME, CENSOR, STRATA, TRT (treatment) variables.
C The survival time, TIME, can be real time or the rank of the time.
C The censorship variable, CENSOR,is 0 if censoring occurs and is 1 if
C death occurs.  STRATA takes value 1 to NG (the number of strata).
C The treatment indicator, TRT, takes value 0 and 1.
C
C NOTE THAT:
C > If sample size is greater than 500, the dimension need to be modified,
C   however, the maximum array dimension depends on machine configuration.
C > The read and write unit varied from system to system.
      program MSTEST
      implicit double precision (A-H,O-Z)
      integer is(501),iss(101),nds(501)
      dimension d(501),d2(501),e(101),ee(101),sumd(501),
     *trt(501),censor(501),time(501),wi(501)
      character*50 infnam
      character*50 outnam
C
      write(*,*)'enter the input file name(specify the directory):'
      read(*,1001) infnam
C
      write(*,*)'enter the output file name(specify the directory):'
      read(*,1001) outnam
 1001 format(a50)
C
      OPEN (10,FILE=infnam)
      OPEN (12,FILE=outnam)
C
C  SET UP # OF MAXIMUN ITERATION & STOPCRIT
C
      maxi=10000
      stopcrit=1.0d-6
      sum2=0.0d0
      sum=0.0d0
C
C  INPUT TOTAL SAMPLE SIZE - N
C      & # OF STRATA - NG
C
      print *,'enter total sample size:'
      read *,n
      print *,'enter number of strata:'
      read *,ng
      read(10,*)(time(i),censor(i),is(i),trt(i),i=1,n)
C
```

```
      ns=float(n)/float(ng)
      do 5 i=1,ng
C
C  TO INITIALIZE THE EXP(BETA)'S, e(i)=exp(beta)
C
  5   e(i)=1.d0
C
C  ki IS THE NUMBER OF ITERATIONS REQUIRED FOR CONVERGENCE
C
      ki=0
C
  1   ki=ki+1
C
C  nds IS THE UNCENSORED SUBJECTS (DEATHS) AT EACH STRATUM
C
      do 6 i=1,ng
  6   nds(i)=0.0d0
      d(1)=0
C
C  Compute the proportion of being trt=1, p = Pr(trt=1)
C
      sum=0.0d0
      do 8 i=1,n
  8   sum=sum+trt(i)
      p=sum/n
C
C  Time 1, d(1) is the risk set at time 1
C
      do 10 i=1,n
      j=is(i)
      nds(j)=nds(j)+censor(i)
 10   d(1)=d(1)+e(j)
C
C  Time 2 to Time n, calculate the risk set at each time
C
      do 20 i=2,n
      j=is(i-1)
      d(i)=d(i-1)-e(j)
 20   continue
C
C  Compute the number of deaths
C
      do 30 i=1,ng
      sumd(i)=0.d0
 30   iss(i)=0

      do 40 i=1,n
      k=n+1-i
      j=is(k)
C
      iss(j)=iss(j)+1
      do 50 l=1,ng
      sumd(l)=sumd(l)+censor(k)*dfloat(iss(l))/d(k)
 11   format(2f5.2,i5,f5.2)
C
```

```
C   IN CASE OF THE DENOMINATOR BEING 0
C
      if (sumd(l).gt.0) goto 50
      sumd(l)=0.1d-6
 50   continue
      if (nds(l).ne.0) goto 35
      e(l)=0.d0
 35   continue
 40   continue
      do 60 i=1,ng
 60   ee(i)=dfloat(nds(i))/sumd(i)
      do 61 i=ng,1,-1
      ee(i)=ee(i)/ee(1)
      if (ee(i) .gt. 1.0d6) ee(i)=1.0d6
 61   if (ee(i) .lt. 1.0d-6) ee(i)=1.0d-6

      amaxreld=0
      do 70 i=2,ng
      rd=abs(ee(i)/e(i)-1.d0)
      if (rd .gt. amaxreld) amaxreld=rd
 70   e(i)=ee(i)
C
      write(*,1000)ki
C
      if (amaxreld .gt. stopcrit .and. ki .lt. maxi) goto 1
C
C  nd is the total number of deaths
C
      nd=0.0d0
      do 99 i=1,ng
 99   nd=nd+nds(i)
C
C   TO WRITE OUT THE EXP(BETA)'S
C
      write(*,1002)ki
C
      write(12,1002)ki
      write(12,999)
      write(12,*)'  n: sample size,  ng: # of strata,  ns: stratum size'
      write(12,1003)n,ng,ns
      write(12,1004)nd
      write(12,999)
      write(12,1005)
      write(12,1006)e(1)
      write(12,1007)(i,e(i),i=2,ng)
 999 format('                                        ')
1000 format('  iteration:',i5)
1002 format('  *** number of iterations went through:',i5,'   ***')
1003 format(' n=',i5,8x,'ng=',i5,8x,'ns=',i5)
1004 format(' total number of deaths :',i5)
1005 format(' The estimated stratum effect, exp(beta):')
1006 format('  stratum     1:',2x,f11.6,2x,'  (by definition)')
1007 format('  stratum  ',i4,':',2x,f11.6)
C
C   TO COMPUTE SCORE TEST STATISTIC
```

```
      d(1)=0.0
      d2(1)=0.0
      do 100 i=1,n
      j=is(i)
      d(1)=d(1)+e(j)
      d2(1)=d2(1)+trt(i)*e(j)
  100 continue
C
      do 200 i=2,n
      j=is(i-1)
      d(i)=d(i-1)-e(j)
      d2(i)=d2(i-1)-trt(i-1)*e(j)

  200  continue
C
      sumn1=trt(n)*censor(n)
      sumn2=censor(n)*d2(n)/d(n)
C
      do 510 i=n-1,1,-1
      sumn1=sumn1+trt(i)*censor(i)
      sumn2=sumn2+censor(i)*d2(i)/d(i)
  510  continue
C
      score=sumn1-sumn2
C
      sumi=0.0d0
      do 540 i=1,n
      j=is(i)
      sumi=sumi+censor(i)/d(i)
      wi(i)=censor(i)-sumi*e(j)
  540  continue
C
      var=0.0d0
      do 550 i=1,n
  550  var=var+wi(i)*wi(i)
C
      cov=0.0d0
      do 570 i=1,n-1
      do 580 j=i+1,n
      if (is(i).eq.is(j)) cov=cov+wi(i)*wi(j)
  580  continue
  570  continue
C
      varest=p*(1.0d0-p)*(var-2.0d0*cov/(dfloat(ns)-1.0d0))
      test=score*score/varest
C
      write(*,1008)score
      write(*,1009)test
      write(12,1008)score
      write(12,1009)test
 1008 format('  Score Statistic :',2x,f12.6)
 1009 format('  Score Test      :',2x,f12.6)
C
      stop
      end
```

PROGRAM GUIDE

The purpose of this guide is to present instruction in the use of MSTEST which is a computer program written in FORTRAN to compute the proposed modified score test based on the proportional hazards model. It is to used in testing of no treatment effect for highly stratified data with complete or censored data.

An Overview of MSTEST

MSTEST is a special-purpose computer program that uses iterative methods described by Cheng (1989) to compute the proposed score test. The program produces an array of statistics, including estimates of the stratum parameters of the proportional hazards model with indicator stratum covariate and the modified score test (Cheng, 1989) in testing of no treatment effect. The output contains estimated regression coefficient is in exponential form (i.e. $e^{\hat{\beta}}$'s) and a MSTEST statistic whose p value can be found by comparing to any standard Chi-square table with one degree of freedom. All calculations are performed in double precision. The initial parameter estimates for strata effects are assumed to be zero. Convergence is assumed when the differences in any two parameter estimates between two successive iterations is less than 10^{-6}.

Entering The Data to MSTEST

The data are assumed sorted on survival time from smallest to largest with failures preceding censorings in the event of ties. The PC SORT routine, SAS PROC SORT procedure, or any other sort routines that accomplishes the same task can be

applied prior to this program to get the data properly sorted. Each data record is assumed to contain 4 variables, survival time (TIME), censorship (CENSOR), strata indices (STRATA) and treatment indicator (TRT), this ordering must be followed. The survival time can be real time or the rank of the time, the censorship variable is assumed to take the values 0 if censoring occurs and 1 if death occurs. The strata variables take the value 1 to ng, where ng is the number of strata. The treatment variable takes values 0 and 1.

Modifications May be Done

Note that in the program, unit 10 is used for input unit and unit 12 used as output unit. The choice of the unit for input/output device depends on the specific computer machine configurations. For example, if you would like to change the input unit to 5 and output unit to 6, change the program as the following:

go to the program line,

Change:
```
OPEN (10, FILE=infnam)
OPEN (12, FILE=outnam)
```
To:
```
OPEN (5, FILE=infnam)
OPEN (6, FILE=infnam)
```

The program here is written to read from a file with the restrictions stated in previous section and to write out to both the screen and pre-specified file. The maximum number of sample size allowance of this program depends on the computer memory limitation which varies from machine to machine. The program presented here is allowing for total sample size, $n \leq 500$, and number of stratum, $n_g \leq 100$, which is due to the limitation of maximum array dimensions PC FORTRAN had. For example, when running this program under VAX machine, the dimension can be changed as the following:

go to the program line,

Change:
```
integer is(501),iss(101),nds(501)
dimension d(501),d2(501),e(101),ee(101),sumd(501),
*trt(501),censor(501),time(501),p(501)
```

To:

```
integer is(1001),iss(501),nds(1001)
dimension d(1001),d2(1001),e(501),ee(501),sumd(1001),
*trt(1001),censor(1001),time(1001),p(1001)
```

Limitation of the Program

1. The program is written by setting the parameter estimates of first stratum to be one (as reference), problems may result if the members in first stratum are all censored. If there is no deaths (failures) in the first stratum, the program can be modified to allow the other stratum as reference stratum or the stratum indices can be relabelled in the data file to allow the other stratum as stratum one.
2. To avoid underflow/overflow problems, when an estimated stratum effect diverges to greater than 10^6, it is set to be 10^6, similarly when an estimated stratum effect diverges to less than 10^{-6}, it is set to be 10^{-6}.

Example

A small dataset has been arbitrarily created as follows:

Example: n = 40, matched pair design (ns=2) and 20 strata (ng=20)

TIME	CENSOR	STRATA	TRT
1	1	1	0
3	0	2	0
4	1	3	1
6	0	4	1
7	1	5	0
8	0	6	0
10	0	7	1
13	1	8	1
15	1	9	0
17	1	10	0
18	1	11	0
24	0	12	1
31	1	13	0
32	0	14	0
43	1	15	1
51	0	16	1
54	1	17	0
56	1	18	0
62	1	19	0
68	1	20	0
69	0	20	1
71	0	19	1
74	1	18	1
77	1	17	1
83	1	16	0
85	1	15	0
86	1	14	1
94	1	13	1
99	0	12	0
107	1	11	1
119	1	10	1
127	1	9	1
135	1	8	0
138	1	7	0
146	1	6	1
149	1	5	1
151	1	4	0
157	1	3	0
158	1	2	1
171	1	1	1

The above data are assumed being sorted in ascending order on survival time as was required and stored in the file, sample.dat, in a: drive. The steps to run MSTEST are

showing below, with > sign indicating where to type in information:

After compiling the MSTEST program, say, using PC FORTRAN 4.1,

```
> MSTEST    (or  RUN MSTEST if using VAX)
enter the input file name (specify the directory):
> A:\SAMPLE.DAT
enter the output file name (specify the directory):
> A:\SAMPLE.OUT
enter total sample size:
> 40
enter number of strata:
> 20
```

The program should run immediately after you type in the above information, the results will be showing in both the screen and the output file specified. The output file contains the parameter estimates of the e^{β}, the score statistic of the proposed test and the score test as listed below:

```
This is the parameter estimates of the exp(beta):

 stratum    1:     1.000000    (by definition)
 stratum    2:     0.999998
 stratum    3:     3.990708
 stratum    4:     2.991869
 stratum    5:     8.925233
 stratum    6:     5.943808
 stratum    7:     7.916622
 stratum    8:    20.824293
 stratum    9:    25.555313
 stratum   10:    29.967229
 stratum   11:    34.004952
 stratum   12:     0 .000001
 stratum   13:    37.618064
 stratum   14:    21.007387
 stratum   15:    45.319642
 stratum   16:    24.958185
 stratum   17:    52.539377
 stratum   18:    54.303498
 stratum   19:    27.589948
 stratum   20:    26.020738

==============================================
Score Statistic :  -4.180024
Score Test      :   2.501307
```

Reference:

Cheng, Y. L. (1989), Modified score test for highly stratified survival data in randomized clinical trials. Ph. D. dissertation, UNC- Chapel Hill.