ABSTRACT

Title of dissertation: Comparing Survival Distributions in the Presence of Dependent Censoring: Asymptotic Validity and Bias Corrections of the Logrank Test

Yabing Mai
Doctor of Philosophy, 2008

Dissertation directed by: Professor Eric V. Slud
Department of Mathematics

We study the asymptotic properties of the logrank and stratified logrank tests under different types of assumptions regarding the dependence of the censoring and the survival times.

When the treatment group and the covariates are conditionally independent given that the subject is still at risk, the logrank statistic is asymptotically standard normally distributed under the null hypothesis of no treatment effect. Under this assumption, the stratified logrank statistic has asymptotic properties similar to logrank statistic.

However, if the assumption of conditional independence of the treatment and covariates given the at risk indicator fails, then the logrank test statistic is generally biased and the bias generally increases in proportional to the square root of the sample size. We provide general formulas for the asymptotic bias and variance. We also establish a contiguous alternative theory regarding small violations of the assumption as well as of the usually considered small differences between treatment

and control group survival hazards.

We discuss and extend an available bias-correction method of DiRienzo and Lagakos (2001a), especially with respect to the practical use of this method with unknown and estimated distribution function for censoring given treatment group and covariates. We obtain the correct asymptotic distribution of the bias-corrected test statistic when stratumwise Kaplan-Meier estimators of the conditional censoring distribution are substituted into it. Within this framework, we prove the asymptotic unbiasedness of the corrected test and find a consistent variance estimator.

Major theoretical results and motivations of future studies are confirmed by a series of simulation studies.

Comparing Survival Distributions in the Presence of Dependent Censoring: Asymptotic Validity and Bias Corrections of the Logrank Test

by

Yabing Mai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Dr. Eric V. Slud, Chair/Advisor
Dr. Abram M. Kagan
Dr. Benjamin Kedem
Dr. Paul J. Smith
Dr. Deborah Rohm Young

ACKNOWLEDGEMENTS

I would like to thank many people who have helped me in completing this thesis.

I am deeply indebted to my advisor Dr. Eric Slud. Without his selfless help, this thesis would have never been completed. Without his continuous inspiration, I would have given up a long time ago. I do not know how to express my thanks in words, because to what I should thank to, all words sounds pale. I am privileged to be his student. I am proud of having such a great person being my advisor. He is my role model in my graduate study, and he will continue to be the one in my entire life.

I also want to thank my committee members: Drs. Abram Kagan, Benjamin Kedem, Paul Smith and Deborah Young for their comments and suggestions to this thesis, more important, for their continuous help to me during my graduate study at University of Maryland.

Many thanks to my supervisor Dr Guoxing Soon at the Food and Drug Administration for his support to me during my two-year internship, from which I gained valuable experience.

My sincere thanks to all my friends at the Math department, especially Nadarajasundaram Ganesh, Shihua Wen, Tinghui Yu and Linbao Zhang for their sincerely help and all the good time shared with them. Thanks also go to the warmhearted Math Department staff: Haydee Hidalgo, Fletcher Kinne, Celeste Regalado, William Schidknecht, Sharon Welton, Tony Zhang...

Finally, I would like to give special thanks to my parents and my girlfriend. Their endless care and love enable me complete this work and keep me moving forward.

Table of Contents

List of Tables vi

List of Notation and Abbreviations vii

1 Background and Preliminaries 1
- 1.1 Background . . . 1
- 1.2 Overview of thesis . . . 4
- 1.3 Definitions and Assumptions . . . 5
  - 1.3.1 General Setting . . . 5
  - 1.3.2 Assumptions . . . 7
  - 1.3.3 Tests for Treatment Effectiveness . . . 9
  - 1.3.4 Logrank Test Statistic . . . 11
  - 1.3.5 Stratified Logrank Test Statistic . . . 12
  - 1.3.6 W-Stratified Logrank Test Statistic . . . 13

2 Logrank Rank Test with Covariate-mediated Dependent Censoring 14
- 2.1 Large Sample Null Distribution and Consistency . . . 14
- 2.2 Biased Logrank Test . . . 16
  - 2.2.1 The Bias of the Logrank Statistic . . . 17
  - 2.2.2 The Variance . . . 19
- 2.3 Logrank Test under the Kong-Slud Assumption . . . 22
  - 2.3.1 Kong-Slud Assumption . . . 23
  - 2.3.2 Large Sample Null Distribution . . . 23

3 Stratified Logrank Tests under Two Types of Stratifications 32
- 3.1 Stratified Logrank Test with Completely Discrete Covariate . . . 32
- 3.2 The W-Stratified Logrank with the Kong-Slud Assumption . . . 35
- 3.3 Comparisons . . . 44
  - 3.3.1 Homogeneous Model . . . 46
  - 3.3.2 Cox Proportional Hazard Model . . . 47
  - 3.3.3 Accelerated Failure Model . . . 48
- 3.4 Remark . . . 49

4 A New Class of Contiguous Alternatives 51

5 A Bias-corrected Logrank Test 58
- 5.1 The $\varphi(\cdot)$ Function and the Weighted at Risk Indicator . . . 58
- 5.2 Bias Correction with Kaplan-Meier Estimation . . . 61
- 5.3 Aalen's Additive Model . . . 78

6 Simulations and Examples 80
6.1 The Bias 80
6.2 Comparisons between the Logrank and Stratified Logrank Test 82
6.3 Simulations on the Bias-corrected Logrank Test 84
6.3.1 Bias-corrected Logrank Test Under H_0 86
6.3.2 Bias-corrected Logrank Test under H_A 88

7 Summary of Results and Future Research Problems 90

A Appendix: Lemmas and Proofs 93

Bibliography 123

List of Tables

6.1 Biased Logrank, Numerical Calculation and Simulation, $n = 100, 400$, $R = 5000$ 81

6.2 Logrank and Stratified Logrank Under KS and H_0, $N = 10000$ 84

6.3 Logrank and Stratified Logrank Under KS and H_n, $N = 10000$ 85

6.4 Logrank and Stratified Logrank Without KS and under H_0, $N = 10000$ 85

6.5 Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_0, Example 1 87

6.6 Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_0, Example 2 88

6.7 Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_A 89

List of Notation and Abbreviations

Ω	a space of outcomes of an experiment
ω	an arbitrary element of Ω
$\bar{A}$	for any set A in Ω, $\bar{A}$ denotes its complement
$\mathcal{F}$	a σ-algebra of subsets of Ω
P	a probability measure on Ω
R	the real line, $(-\infty, \infty)$
R^+	the non-negative real line, $[0, \infty)$
$\mathcal{B}$	the Borel σ-algebra of the real line
$\sigma\{\cdot\}$	the σ-algebra generated by random variables in $\{\cdot\}$
$E(Y \mid \mathcal{G})$	the conditional expectation of Y w.r.t. a σ-algebra $\mathcal{G} \subset \sigma\{X\}$
$E(Y \mid X)$	$E\{Y \mid \sigma(X)\}$, for random variable X and Y.
$I_{[A]}$	indicator of an event or a set A
$\mathcal{F}_{t+}$	the σ-algebra $\cap_{h>0}\mathcal{F}_{t+h}$
$\mathcal{F}_{t-}$	the smallest σ-algebra containing all the sets in $\cup_{h>0}\mathcal{F}_{t-h}$
$X(t-)$	$\lim_{h\downarrow 0} X(t-h)$
$\Delta X(t)$	$X(t) - X(t-)$, if X is right continuous with left hand limit
$\vee$	$x \vee y = \max(x, y)$
$\wedge$	$x \wedge y = \min(x, y)$
$\perp\!\!\!\perp$	Probability independence
$\perp\!\!\!\perp \mid Z$	$X \perp\!\!\!\perp Y \mid Z$ means X and Y are conditionally independent w.r.t. $\sigma(Z)$
$\langle M_1, M_2 \rangle$	the predictable covariation process for M_1, M_2
$N(\mu, \sigma^2)$	the normal distribution with mean μ and variance σ^2
Φ	the cumulative distribution function for the $N(0, 1)$ distribution
$\equiv$	equal, by definition
$\overset{d}{\sim}$	$X \overset{d}{\sim} F$ means X has distribution function F
$\overset{as}{\to}$	converge almost surely
$\overset{p}{\to}$	convergence in probability
$\overset{d}{\to}$	convergence in distribution
$\overset{L^p}{\to}$	L^p convergence
$o_p(1)$	a r.v. sequence $X_n = o_p(1)$ means $X_n \overset{p}{\to} 0$
$O_p(1)$	a r.v. sequence $Y_n = O_p(1)$ means Y_n is bounded in probability
$f^{(r)}(\cdot)$	the r^{th} derivative of function f
$\mathrm{Var}(X)$	variance of the random variable X
r.v.	"random variable"
iid	"independent and identically distributed"

Throughout the thesis, we use the convention $0/0 = 0$.

Chapter 1

Background and Preliminaries

1.1 Background

Randomized clinical trials generally use hypothesis testing to compare the survival experience of two groups of individuals. The logrank statistic is the most popular statistic used in these tests. A great deal of work has been done on the properties of the log rank test of no treatment effect for two sample right-censored survival data (Mantel, 1966; Cox, 1972; Peto and Peto, 1972; Green and Byar, 1978; Schoenfeld, 1981; Morgan, 1986; Schoenfeld and Tsiatis, 1987).

The logrank test statistic, as defined in Section 1.3, compares the Nelson-Aalen estimator (Nelson, 1972; Aalen, 1978) of hazard functions from two groups at each observed event time. The statistic combines the observed minus expected numbers of events in the treatment group at each observed event time, across all event time points for an overall comparison. Generally there is no need for parametric model assumptions because the log rank test is a nonparametric test procedure. But under some special model assumptions, the log rank test may have particularly good properties. For example, the logrank test, with independent death and censoring, is locally most powerful among the family of rank tests under the proportional hazard model (Peto and Peto, 1972).

The logrank test has very different properties under various types of censoring assumptions regarding the dependence of the censoring, covariates and the survival times. First and foremost, to use the classsical logrank test one must assume non-informative censoring because informative censoring may result in a non-identifiable distribution of the time to event (Tsiatis 1975, Slud and Rubinstein 1983). Though the non-informative censoring assumption (See 3.1 of Chapter 1 in Fleming and Harrington, 1991) is slightly weaker, that is, less restrictive, than the assumption of statistical independence of the survival and censoring, the independent censoring assumption is often imposed. Censoring may depend on survival through some covariates or the treatment group. It is not rare that one group in a clinical trial may have higher dropout rate than the other, or that patients with some specified covariate patterns are more likely to drop out than others. In this thesis, we make the assumption of conditional independence given the treatment group, or the covariates, or both. With this assumption the logrank test may still have good properties. For example, if the censoring time is conditionally independent of the survival time given the treatment group, the logrank test statistic is asymptotically normal with mean zero under the null hypothesis and is consistent against the stochastic ordering alternative (Gill, 1980).

If the distribution of the censoring time depends on both the treatment group and the covariates, then according to Dirienzo and Lagakos (2001a), the null asymptotic distribution of the score test is generally not centered at zero when the model is misspecified. As they concluded, the logrank test is generally biased under this kind of dependent censoring. However, the null asymptotic distribution of the logrank test will still be a standard normal as long as the following dependent censoring assumption holds: "Under the null hypothesis, the treatment group is conditionally independent of the covariates given that the subject is still at risk." This assumption was first introduced in Kong and Slud (1997) for the purpose of finding a robust variance estimator. We refer to it as the Kong-Slud Assumption in this thesis.

Though the Kong-Slud Assumption is not fully general, it is reasonable in many situations. For example, it holds when the censoring depends on both the treatment group and the covariates as long as there is no treatment-by-covariates interaction in the conditional distribution function of the censoring time. Kong and Slud (1997) proposed a general scenario in which their Assumption holds, in which a clinical study has purely administrative censoring that occurs at a fixed calendar time, but patients enter the study at earlier staggered random times that may depend upon their covariates. Suppose further that patients may decide to withdraw from the study for reasons not depending on medical covariates, which may be related to side effects of the therapy. Withdrawals due to side effects are dependent on treatment-group, but as long as the side-effects are not materially dependent upon the covariates, the Kong-Slud Assumption holds approximately.

When the Kong-Slud Assumption does not hold, the log rank test is generally biased and the size of the test will be inflated regardless of whether the bias is positive or negative. Moreover, simply increasing the sample size would not reduce the bias. On the contrary, it would generally increase in the proportional to the square root of the sample size. Hence ignoring the potential bias may have serious consequences for the validity of clinical trials.

The main purpose of this thesis is to study the asymptotic validity of the logrank test when Kong-Slud Assumption holds and the bias corrections for the logrank test when this Assumption does not hold.

1.2 Overview of thesis

In Chapter 2, we study the properties of the logrank test under various assumptions about dependent censoring. The chapter starts with a short review of the large sample null distribution and consistency under the assumption that the censoring is conditionally independent of the survival given the treatment group. Our contributions in this chapter are: (1) identifying the potential bias and providing general formulas for the bias and variance of the logrank statistic under the assumption that the censoring depends on both the treatment group and the covariates; (2) proving that under the Kong-Slud assumption, the large sample null distribution of the logrank statistic is standard normal, asymptotically.

In Chapter 3, we study the properties of the stratified logrank test under various assumptions of dependent censoring. The contributions are, primarily, showing that the class of W-stratified logrank statistics (defined in Section 1.3.6) have asymptotic standard normal distributions under the Kong-Slud Assumption and are generally biased when the assumption does not hold. We also show that under the Kong-Slud Assumption, the variance estimators for the logrank statistic and the W-stratified logrank statistic are asymptotically equivalent.

In Chapter 4, we establish a contiguous alternative theory regarding small violations of the Kong-Slud Assumption as well as of the usually considered small differences between treatment and control group survival hazards. This theory enables us to calculate the asymptotic distribution, with small violations of the Kong-Slud Assumption, of the logrank statistic under contiguous alternatives to models satisfying to the Kong-Slud Assumption.

In Chapter 5, we discuss and extend a bias correction method proposed by DiRienzo and Lagakos (2001a), especially with respect to the practical use of this method with unknown and estimated distribution function for censoring given treatment group and covariates. We contribute by obtaining the correct asymptotic distribution of the corrected test statistic when stratumwise Kaplan-Meier estimators of conditional censoring distribution are substituted. Within this framework, we prove the asymptotic unbiasedness of the corrected test and find a consistent variance estimator.

In Chapter 6, we provide simulation studies confirming and illustrating the theoretical results of the previous chapters.

In Chapter 7, we summarize our results with a comprehensive discussion on bias corrections and future research problems.

Appendix A contains many lemmas and proofs cited earlier in this thesis.

1.3 Definitions and Assumptions

1.3.1 General Setting

Assume n patients are randomly assigned to two different treatment groups. The ith patient has latent survival time T_i and censoring time C_i.

Define

$$X_i \equiv T_i \wedge C_i \ ; \ \delta_i \equiv I_{[T_i \le C_i]},$$

the counting process

$$N_i(t) \equiv \delta_i I_{[X_i \le t]}$$

and the at risk indicator

$$Y_i(t) \equiv I_{[X_i \ge t]}.$$

Assume $(X_i, \delta_i, Z_i, V_i)$, for $i = 1, 2, \ldots, n$, are iid realizations of (X, δ, Z, V), where Z is the treatment group indicator that only takes values of 0 and 1 and V is a q-dimensional vector of covariates.

The conditional hazard function of survival time T of a patient given treatment group Z and covariate V is generally denoted as

$$\lambda(t \mid z, v) = \lambda(t, \theta z, v) \tag{1.1}$$

where θ is an unknown scalar parameter and the null hypothesis of no treatment effect will be $H_0: \ \theta = 0$. At many places in this thesis we use the notation $\lambda(t, 0, v)$ or $\lambda(t, v)$ for the conditional hazard function of T given $V = v$ and use $\Lambda(t, 0, v)$ or $\Lambda(t, v)$ for the conditional cumulative hazard function for the survival time under H_0. Further regularity assumptions on $\lambda(\cdot)$ will be imposed later.

Denote the conditional survival function for a patient, given treatment Z and covariate V, as

$$S(t, Z, V) = Pr\{T \ge t \mid Z, V\}$$

and the conditional survival function based on censoring for this patient as

$$S_C(t, Z, V) = Pr\{C \geq t \mid Z, V\}.$$

The survival function of T under the null hypothesis $H_0 : \theta = 0$ is $S(t, 0, V)$ and in some places of this thesis, $S(t, V)$.

Define the history $\mathcal{F}_t$ generated by the observable data as

$$\mathcal{F}_t = \sigma(N_i(s), Y_i(s), Z_i, V_i;\ 0 \leq s \leq t, i = 1, 2, \dots) \tag{1.2}$$

By the Doob-Meyer decomposition theorem (Section II.3 of Anderson et. al. 1992), for each i,

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda(s, 0, V_i)ds$$

is an $\mathcal{F}_t$ martingale under H_0. From Section II.3.2 and II.4.1 of Andersen et al (1992), the predictable variation process $\langle M_i \rangle$ of M_i satisfies:

1. $M_i^2 - \langle M_i \rangle$ is an $\mathcal{F}_t$ martingale equals zero at time zero;

2. $\langle M_i \rangle(t) \;=\; \int_0^t Y_i(s)\lambda(s, 0, V_i)ds.$

1.3.2 Assumptions

In this section we list all assumptions that will be used in later chapters. Note that these assumptions are used in different combinations in different places.

Assumption 1.1 *The treatment group indicator Z and the prognostic covariates V are independent: $Z \perp\!\!\!\perp V$.*

The unconditional independence of treatment and covariates in Assumption 1.1 is assumed throughout this thesis.

Assumption 1.2 *(Noninformative Censoring I) The survival time T is conditionally independent of the censoring time C given the treatment group indicator Z only, that is: $T \perp\!\!\!\perp C \bigm| Z$.*

Assumption 1.3 *(Noninformative Censoring II) The survival time T is conditionally independent of the censoring time C given the treatment group indicator Z and the covariates V, that is: $T \perp\!\!\!\perp C \bigm| (Z, V)$.*

Assumption 1.3 is the same dependence assumption used in the Cox model.

At various places in this thesis, we may assume either Assumption 1.2 or Assumption 1.3 but not both of them at the same time.

Assumption 1.4 *(Kong-Slud I) Under the null hypothesis $H_0 : \lambda(t \mid 1, v) = \lambda(t \mid 0, v)$, the treatment group indicator Z is conditionally independent of the covariates V given that the subject is still at risk:*

$$E_0[Z \mid Y(t) = 1, V] \;=\; E_0\{Z \mid Y(t) = 1\} \equiv \mu(t).$$

Assumption 1.4 was first introduced by Kong and Slud (1997) and is an important assumption in this thesis.

Assumption 1.5 *(Kong-Slud II) The survival function of the censoring time C satisfies*

$$-\log S_C(t, Z, V) \; = \; a(t, Z) + b(t, V)$$

for some positive functions $a(\cdot)$ and $b(\cdot)$.

Assumption 1.5 says that there is no interaction between treatment and covariates in the conditional cumulative hazard function for censoring given Z and V. It is easy to show that Assumption 1.5 implies Assumption 1.4 (See Lemma A.2.). Though a little more restrictive than Assumption 1.4, Assumption 1.5 is employed as the "Kong-Slud Assumption" in many examples and simulations throughout this paper due to its easy form in calculation.

Assumption 1.6 *(DiRienzo-Lagakos) The censoring time C is either conditionally independent of the treatment group Z given the covariates V or is conditionally independent of the covariates V given the treatment group Z:*

$$C \perp\!\!\!\perp V \mid Z \;\; \text{or} \;\; C \perp\!\!\!\perp Z \mid V.$$

Assumption 1.6 was first introduced by DiRienzo and Lagakos (2001b) and is more restrictive than the Kong-Slud Assumption. As shown in Lemma A.2, Assumptions 1.4 and 1.5 hold whenever Assumption 1.6 holds.

1.3.3 Tests for Treatment Effectiveness

The null hypothesis of the test for treatment effectiveness is that there is no effect of the treatment, that is, that there is no difference between the conditional survival or hazard functions of the two groups. We define the null hypothesis as

$$H_0 : S(t, 1, V) = S(t, 0, V)$$

which is equivalent to the hypothesis that $\lambda(t, 1, V) = \lambda(t \mid 0, V)$ for all t. If the hazard function is parameterized as in (1.1), the null hypothesis can also be written as

$$H_0 : \theta = 0$$

In this thesis we study properties of the tests for treatment effectiveness under various choices of alternative hypothesis.

Definition 1.1 *The alternative $H_1 : \lambda(t, 0, V) \geq \lambda(t, 1, V)$ for all t and V is called the ordered hazards alternative.*

Definition 1.2 *The alternative $H_2 : S(t, 1, V) \geq S(t, 0, V)$ for all t and V is called the alternative of stochastic ordering.*

It is clear that H_1 implies H_2, since $S(t, z, v) = \exp\{-\Lambda(t, z, v)\}$ with $\Lambda(t, z, v) \equiv \int_0^t \lambda(s, z, v)ds$.

Another type of alternative that is of interest in this thesis is related to contiguous sequences of probabilities:

Definition 1.3 *Let sequences P_n and Q_n be the probability measures under the null hypothesis H_0 and the alternative H_n, respectively. If $P_n(A_n) \to 0$ implies $Q_n(A_n) \to 0$ for every sequence of measurable sets A_n, we say that Q_n is contiguous with respect to P_n (Section 6.1, van der Vaart, 1998) and H_n is a contiguous alternative to H_0.*

As in Section VIII.1.2 of Anderson et. al. (1992), it is well known that under suitable regularity conditions, the alternatives

$$H_n : \theta = b/\sqrt{n}$$

are contiguous to $H_0 : \theta = 0$. A more general result is proved in Theorem 4.1 of this thesis.

Finally we define consistency of a test:

Definition 1.4 *Let* $\mathbf{X}$ *be a random population,* $T_n, n = 1, 2, \ldots,$ *be a sequence of test statistics used to test a hypothesis* H*, and* $R_n = \{\mathbf{X} : T_n(\mathbf{X}) \geq c_n\}, n = 1, 2, \ldots,$ *be an associated set of level* α *rejection regions. The sequence* T_n *is said to be consistent against a family of alternative hypotheses* H_A *if*

$$\lim_{n\to\infty} P(T_n(\mathbf{X}) \in R_n) = 1$$

whenever the probability P *governing* $\mathbf{X}$ *lies in* H_A*.*

1.3.4 Logrank Test Statistic

Define

$$\bar{N}_1(t) = \sum_{i=1}^{n} Z_i N_i(t); \;\; \bar{N}(t) = \sum_{i=1}^{n} N_i(t); \; \bar{N}_0(t) = \bar{N}(t) - \bar{N}_1(t);$$

$$\bar{Y}_1(t) = \sum_{i=1}^{n} Z_i Y_i(t); \;\; \bar{Y}(t) = \sum_{i=1}^{n} Y_i(t); \; \bar{Y}_0(t) = \bar{Y}(t) - \bar{Y}_1(t).$$

Then the numerator of the logrank test statistic is defined as

$$n^{-\frac{1}{2}}\hat{U}_L \equiv n^{-\frac{1}{2}} \int \left\{ d\bar{N}_1(t) - \frac{\bar{Y}_1(t)}{\bar{Y}(t)} d\bar{N}(t) \right\} \tag{1.3}$$

Note that (1.3) can also be written as

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\hat{U}_L \equiv n^{-\frac{1}{2}}\int\Big\{Z_i - \frac{\sum_{i=1}^{n} Z_i Y_i(t)}{\sum_{i=1}^{n} Y_i(t)}\Big\}dN_i(t). \tag{1.4}$$

The square of the denominator of the logrank statistic is defined as

$$\hat{V}_L \equiv \int \frac{1}{n}\cdot\frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)^2}d\bar{N}(t). \tag{1.5}$$

1.3.5 Stratified Logrank Test Statistic

The term "stratified logrank test" will be used to refer to a stratified logrank test based on only on the complete covariates V, and even then, only when V is discrete. Here we assume V is discrete with finite values and let $\mathcal{V}$ be the set of all discrete values of V. For any $v \in \mathcal{V}$, define

$$\xi_i^v \equiv I_{[V_i=v]};$$

$$\bar{N}_{1v}(t) = \sum_{i=1}^{n}\xi_i^v Z_i N_i(t);\;\; \bar{N}_v(t) = \sum_{i=1}^{n}\xi_i^v N_i(t);\; \bar{N}_{0v} = \bar{N}_v - \bar{N}_{1v};$$
$$\bar{Y}_{1v}(t) = \sum_{i=1}^{n}\xi_i^v Z_i Y_i(t);\;\; \bar{Y}_v(t) = \sum_{i=1}^{n}\xi_i^v Y_i(t);\; \bar{Y}_{0v} = \bar{Y}_v - \bar{Y}_{1v}.$$

Then the numerator of the stratified logrank test statistic is defined as

$$n^{-\frac{1}{2}}\hat{U}_S \equiv n^{-\frac{1}{2}}\sum_v\int\Big\{d\bar{N}_{1v}(t) - \frac{\bar{Y}_{1v}(t)}{\bar{Y}_v(t)}d\bar{N}_v(t)\Big\}. \tag{1.6}$$

Note that (1.6) can also be written as

$$n^{-\frac{1}{2}}\hat{U}_S \equiv n^{-\frac{1}{2}}\sum_v\int\Big\{Z_i - \frac{\sum_{i=1}^{n}\xi_i^v Z_i Y_i(t)}{\sum_{i=1}^{n}\xi_i^v Y_i(t)}\Big\}\cdot\xi_i^v dN_i(t). \tag{1.7}$$

The square of the denominator of the stratified logrank statistic is defined as

$$\hat{V}_S \equiv \sum_v\int\frac{1}{n}\cdot\frac{\bar{Y}_{1v}(t)\bar{Y}_{0v}(t)}{\bar{Y}_v(t)^2}d\bar{N}(t). \tag{1.8}$$

1.3.6 W-Stratified Logrank Test Statistic

In this section we define a so-called W-stratified logrank test statistic within which stratification is on a smaller set of covariates than the full set of covariates V appearing in Assumptions 1.3 and 1.4. Now let V remain the same as previously defined: a q-dimensional vector of covariates that may be discrete or continuous. Let a p-dimensional vector $W = h(V)$ be discrete with $p \leq q$ and $n(W) \leq n(V)$, where $n(V)$ is defined as the maximum number of levels of V and $n(V) = \infty$ if the support of V is infinite. Let (V_i, W_i), for $i = 1, 2, \ldots,$ be iid realizations of (V, W). Similar to the definition of the stratified log rank statistics, here we define

$$\xi_i^w \equiv I_{[W_i = w]};$$

$$\bar{N}_{1w}(t) = \sum_{i=1}^{n} \xi_i^w Z_i N_i(t); \;\; \bar{N}_w(t) = \sum_{i=1}^{n} \xi_i^w N_i(t); \; \bar{N}_{0w}(t) = \bar{N}_w(t) - \bar{N}_{1w}(t);$$

$$\bar{Y}_{1w}(t) = \sum_{i=1}^{n} \xi_i^w Z_i Y_i(t); \;\; \bar{Y}_w(t) = \sum_{i=1}^{n} \xi_i^w Y_i(t); \; \bar{Y}_{0w}(t) = \bar{Y}_w(t) - \bar{Y}_{1w}(t).$$

Then the numerator of the W-stratified logrank test statistic is defined as

$$n^{-\frac{1}{2}} \hat{U}_W \equiv n^{-\frac{1}{2}} \sum_w \int \Big\{ d\bar{N}_{1w}(t) - \frac{\bar{Y}_{1w}(t)}{\bar{Y}_w(t)} d\bar{N}_w(t) \Big\}. \tag{1.9}$$

where (1.9) can also be written as

$$n^{-\frac{1}{2}} \hat{U}_W \equiv n^{-\frac{1}{2}} \sum_w \sum_{i=1}^{n} \int \Big\{ Z_i - \frac{\sum_{i=1}^{n} \xi_i^w Z_i Y_i(t)}{\sum_{i=1}^{n} \xi_i^w Y_i(t)} \Big\} \cdot \xi_i^w dN_i(t). \tag{1.10}$$

The square of the denominator of the W-stratified logrank statistic is defined as

$$\hat{V}_W \equiv \sum_w \int \frac{1}{n} \cdot \frac{\bar{Y}_{1w}(t) \bar{Y}_{0w}(t)}{\bar{Y}_w(t)^2} d\bar{N}_w(t). \tag{1.11}$$

Chapter 2

Logrank Rank Test with Covariate-mediated Dependent Censoring

In this chapter we study properties of the logrank test under various assumptions about dependent censoring. In Section 2.1 we review the classical result that under Assumptions 1.1 and 1.2, the logrank test with the test statistic defined at Section 1.3.4 asymptotically achieves the nominal significance level and is consistent against stochastically ordered alternatives. In Section 2.2 we identify the potential bias and provide general formulas for the bias and variance of the logrank statistic under the assumption that the censoring depends on both the treatment group and the covariates. In Section 2.3, we prove that under the Kong-Slud Assumption, the large sample null distribution of the logrank statistic is asymptotically standard normal.

2.1 Large Sample Null Distribution and Consistency

In this section we study the performance of the logrank statistic $\hat{U}_L/\hat{V}_L^{\frac{1}{2}}$ defined in Section 1.3.4 under Assumption 1.1 and 1.2.

Note that when Assumptions 1.2 holds, that is, when the survival time T is conditionally independent of the censoring C given the treatment group Z, the conditional hazard function of T becomes

$$\lambda(t, z, v) = \lambda(t \mid z).$$

Then from Theorems 1.3.1 and 1.3.2 in Fleming and Harrington (1991),

$$\bar{M}_1(t) \equiv \bar{N}_1(t) - \int_0^t \bar{Y}_1(s) d\Lambda(s \mid 1)$$

and

$$\bar{M}_0(t) \equiv N_0(t) - \int_0^t \bar{Y}_0(s) d\Lambda(s \mid 0)$$

are both martingales with respect to $\mathcal{F}_t$, where $\Lambda(t \mid z) \equiv \int_0^t \lambda(s \mid z) ds$.

Hence the numerator $n^{-\frac{1}{2}} \hat{U}_L$ of the logrank statistic can be written as

$$\int n^{-\frac{1}{2}} \frac{\bar{Y}_0(t)}{\bar{Y}(t)} d\bar{M}_1(t) - \int n^{-\frac{1}{2}} \frac{\bar{Y}_1(t)}{\bar{Y}(t)} d\bar{M}_0(t) + \int n^{-\frac{1}{2}} \frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)} \{d\Lambda(t|1) - d\Lambda(t|0)\}.$$

According to Gill (1980), A statistic W_K of the "class $\mathcal{K}$" is defined as

$$W_K = \int_0^\infty K(s) \{d\hat{\Lambda}_1(s) - d\hat{\Lambda}_0(s)\}$$

with $\hat{\Lambda}_1(t) = \int_0^t d\bar{N}_1(s)/\bar{Y}_1(s)$, $\hat{\Lambda}_0(t) = \int_0^t d\bar{N}_0(s)/\bar{Y}_0(s)$ and a $\mathcal{F}_s$ predictable $K(s)$. Then by Section 3.3 of Fleming and Harrington (1991), the statistic $n^{-\frac{1}{2}} \hat{U}_L$ is a statistic of the "class $\mathcal{K}$". Hence from Section 7.2 of Fleming and Harrington (1991) and as a result of the Martingale Central Limit Theorem, the logrank statistic $\hat{U}_L / \hat{V}_L^{\frac{1}{2}}$, as a statistic of "class $\mathcal{K}$", is asymptotically normal with mean 0 and variance 1 as $n \to \infty$.

Furthermore, from Theorems 7.3.1 and 7.3.2 of Fleming and Harrington (1991), the logrank test based on $\hat{U}_L / \hat{V}_L^{\frac{1}{2}}$ is consistent against the alternative of stochastic ordering $H_A : S_T(t \mid 1) \geq S_T(t \mid 0)$ for all t and with strict inequality for some t, where $S_T(t \mid z) = e^{-\Lambda(t \mid z)}$.

The main result of this section is summarized in the following proposition:

Proposition 2.1 *Under Assumptions 1.1 and 1.2, that is, when* $Z \perp\!\!\!\perp V$*,* $T \perp\!\!\!\perp V$ *and* $T \perp\!\!\!\perp C \mid Z$*, the logrank test statistic* $\hat{U}_L/\hat{V}_L^{\frac{1}{2}}$ *defined in Section 1.3.4 is asymptotically normal with mean 0 and variance 1 under* $H_0 : \lambda(t \mid 1) = \lambda(t \mid 2)$ *for all* t *and is consistent under the stochastic ordering alternative* $H_A : S_T(t \mid 1) \geq S_T(t \mid 0)$ *for all* t *and with strict inequality for some* t*.*

2.2 Biased Logrank Test

Regarding the dependence of the censoring and the survival times, the logrank test has very different properties under different types of censoring assumptions. If Assumption 1.2 in Section 2.1 is replaced by Assumption 1.3, that is, $T \perp\!\!\!\perp C \mid (Z, V)$, the survival time T is generally dependent on censoring C given Z, and the hazard function of T depends on the covariate V. It is easy to show that the logrank statistic is no longer a "class $\mathcal{K}$" test statistic. Hence the logrank statistic may no longer have the good properties introduced in Section 2.1 such as the asymptotic standard normal null distribution and consistency with respect to stochastic ordering alternatives. Actually we find that the logrank statistic is generally biased under Assumptions 1.1 and 1.3.

2.2.1 The Bias of the Logrank Statistic

The following lemma gives us the general formula for the bias of $n^{-\frac{1}{2}}\hat{U}_L$, the numerator of the logrank statistic.

Lemma 2.1 *Under Assumptions 1.1 and 1.3, that is, $Z \perp\!\!\!\perp V$ and $T \perp\!\!\!\perp C \mid (Z, V)$, the asymptotic distribution of the numerator of the logrank statistic is no longer centered at 0 under $H_0 : \theta = 0$ and the bias of the statistic $n^{-\frac{1}{2}}\hat{U}_L$ is*

$$E_0\{n^{-\frac{1}{2}}\hat{U}_L\} = \sqrt{n}\int E_0\{[Z - \mu(t)]Y(t)\lambda(t, 0, V)\}dt + o(\sqrt{n}),$$

where $\mu(t) = E_0\{Z \mid Y(t) = 1\}$.

Proof:

The logrank statistic $n^{-\frac{1}{2}}\hat{U}_L$ defined in Section 1.3.4 is

$$\frac{1}{\sqrt{n}}\hat{U}_L = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int\left[Z_i - \frac{\sum_j Y_j(t)Z_j}{\sum_j Y_j(t)}\right]dN_i(t).$$

Writing

$$dN_i(t) = \{dN_i(t) - Y_i(t)\lambda(t, 0, V_i)dt\} + Y_i(t)\lambda(t, 0, V_i)dt$$

in $n^{-\frac{1}{2}}\hat{U}_L$ we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int\left[Z_i - \frac{\sum_j Y_j(t)Z_j}{\sum_j Y_j(t)}\right]dM_i(t) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int\left[Z_i - \frac{\sum_j Y_j(t)Z_j}{\sum_j Y_j(t)}\right]Y_i(t)\lambda(t, 0, V_i)dt$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda(s, 0, V_i)ds$ is an $\mathcal{F}_t$ martingale. Since each $Z_i - \sum_j\{Y_j(t)Z_j\}/\sum_j Y_j(t)$ is $\mathcal{F}_t$ predictable,

$$\int\left[Z_i - \frac{\sum_j Y_j(t)Z_j}{\sum_j Y_j(t)}\right]dM_i(t)$$

is also an $\mathcal{F}_t$ martingale with mean 0 (Section II.3.3, Andersen et al 1992). Therefore,

$$E(n^{-1/2}\hat{U}_L) = \frac{1}{\sqrt{n}} E \sum_{i=1}^{n} \int \left\{ [Z_i - \mu(t)] - \frac{\sum_j [Z_j - \mu(t)] Y_j(t)}{\sum_j Y_j(t)} \right\} \lambda(t, 0, V_i) Y_i(t) dt \tag{2.1}$$

By the definition of $\mu(t)$ and independence of the data vectors with different indices, for all $j \neq i$,

$$E\{Z_j - \mu(t) \mid V_i, \{Y_k(t)\}_{k=1}^{n}\} = 0.$$

Hence

$$E\left\{ \frac{\sum_j [Z_j - \mu(t)] Y_j(t)}{\sum_j Y_j(t)} \lambda(t, 0, V_i) Y_i(t) \right\} = E\left\{ [Z_i - \mu(t)] \frac{Y_i(t)}{\sum_j Y_j(t)} \lambda(t, 0, V_i) \right\}.$$

Therefore

$$E(n^{-1/2}\hat{U}_L) \;=\; \frac{1}{\sqrt{n}} E \sum_{i=1}^{n} \int [Z_i - \mu(t)] \left[1 - \frac{Y_i(t)}{\sum_j Y_j(t)} \right] \lambda(t, 0, V_i) Y_i(t) dt. \tag{2.2}$$

Furthermore,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} E \sum_{i=1}^{n} \int [Z_i - \mu(t)] \frac{Y_i(t)}{\sum_j Y_j(t)} \lambda(t, 0, V_i) Y_i(t) dt \\
= \;& \frac{1}{\sqrt{n}} E \sum_{i=1}^{n} \int [Z_i - \mu(t)] \frac{Y_i(t)}{1 + \sum_{j:j\neq i} Y_j(t)} \lambda(t, 0, V_i) Y_i(t) dt \\
= \;& \sqrt{n} E \int [Z_1 - \mu(t)] Y_1(t) \lambda(t, 0, V_1) E\{1/(1 + \sum_{j=2}^{n} Y_j(t))\} dt,
\end{aligned}$$

whence

$$E(n^{-1/2}\hat{U}_L) = \sqrt{n} E\left\{ \int [Z - \mu(t)] Y(t) \lambda(t, 0, V) dt \right\} + o(\sqrt{n})$$

as $n \to \infty$. □

Define

$$B \equiv E\left\{\int [Z-\mu(t)]Y(t)\lambda(t,0,V)dt\right\}. \tag{2.3}$$

Lemma 2.1 says that the asymptotic bias of the logrank statistic $n^{-\frac{1}{2}}\hat{U}_L$ has top-order term $O(\sqrt{n}B)$ when $B \neq 0$. From the definition of $\mu(t)$, B can be written as

$$\begin{aligned} B &= \int E\{ZY(t)\lambda(t,0,V)\} - E\{Z \mid Y(t)\}E\{Y(t)\lambda(t,0,V)\}dt \\ &= \int E\{Y(t)\} \cdot \Big[E\big\{Z\lambda(t,0,V) \mid Y(t)=1\big\} \\ &\qquad -E\{Z \mid Y(t)=1\}E\{\lambda(t,0,V) \mid Y(t)=1\}\Big]dt \\ &= \int E\{Y(t)\} \cdot \mathrm{Cov}\big\{Z, \lambda(t,0,V) \mid Y(t)=1\big\}dt. \end{aligned} \tag{2.4}$$

Since (2.4) is generally not 0, unless Z and $\lambda(t,0,V)$ are conditionally uncorrelated given $Y(t)=1$ and not dependent on n, the bias of the numerator of the logrank test statistic is of the order of $\sqrt{n}B$. Hence for a clinical study with the biased logrank test, simply increasing the sample size will not correct the bias and may only make the problem worse.

2.2.2 The Variance

In this section we derive a formula for the variance of

$$n^{-\frac{1}{2}}\tilde{U}_L \;=\; n^{-\frac{1}{2}}\sum_{i}^{n}\int \{[Z_i-\mu(t)]dN_i(t) - \eta(t)dt\}$$

under Assumptions 1.1 and 1.3, where $\mu(t) = E\{ZY(t)\}/E\{Y(t)\}$ and $\eta(t) = E\{\lambda(t,0,V)Y(t)\}/E\{Y(t)\}$. As will be shown in Section 2.3.1, if Assumption 1.4

holds, $n^{-\frac{1}{2}}\hat{U}_L - n^{-\frac{1}{2}}\tilde{U}_L \xrightarrow{p} 0$ and $n^{-\frac{1}{2}}\tilde{U}_L$ is an iid sum with mean 0 under H_0. Along with the contiguous alternative theory we prove in Chapter 4, the variance formulae we provide here can be used to study the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_L$ under small violations of Assumption 1.4.

Define

$$V_L \equiv E_0\left\{\int [Z-\mu(t)]^2 Y(t)\lambda(t,0,V)dt\right\}; \tag{2.5}$$

$$V_C \equiv E_0\left\{\int_0^\infty \int_0^t [Z-\mu(t)][Z-\mu(s)][\lambda(t,0,V)-\eta(t)]Y(t)\eta(s)dsdt\right\} \tag{2.6}$$

Then the following results holds.

Lemma 2.2 *Under Assumptions 1.1 and 1.3, the asymptotic variance V_N of $n^{-\frac{1}{2}}\tilde{U}_L$ under H_0 is*

$$V_N = V_L - 2V_C - B^2$$

where V_L, V_C and B are defined in (2.5),(2.6) and (2.3), respectively.

Proof:

Define

$$V_B = E\left\{[\int [Z-\mu(t)][dN_1(t) - Y_1(t)\eta(t)dt]]^2\right\}.$$

The asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_L$ is

$$V_N = V_B - B^2.$$

In general, for a bounded integrating (signed) measure $d\sigma$ with atoms and a bounded symmetric function $g(\cdot,\cdot)$,

$$\begin{aligned}
&\int\int g(s,t)d\sigma(s)d\sigma(t) \\
= &\int g(t,t)\sigma(\{t\})d\sigma(t) + 2\int\int^{t-} g(s,t)d\sigma(s)d\sigma(t) \\
= &\sum_{t:\sigma(\{t\})>0} g(t,t)\ \sigma(\{t\})^2 \ +\ 2\int\int^{t-} g(s,t)d\sigma(s)d\sigma(t). \qquad (2.7)
\end{aligned}$$

In calculating the variance V_N we define $g(s,t) = [Z - \mu(t)][Z - \mu(s)]$ and $\sigma(t) = N(t) - \int_0^t Y(s)\eta(s)ds$. The only case where $\sigma(\{t\}) > 0$ is $\sigma(\{t\}) = 1$, when $T = t$ and $C \geq t$, that is, $\Delta N(t) = 1$.

From (2.7) we can rewrite variance V_B as:

$$\begin{aligned}
V_B &= E\left\{\sum_{t:\sigma(\{t\})>0} g(t,t)\ \sigma(\{t\})^2 \ +\ 2\int\int^{t-} g(s,t)d\sigma(s)d\sigma(t)\right\} \\
&\equiv V_1 \ +\ V_2 \qquad (2.8)
\end{aligned}$$

where

$$\begin{aligned}
V_1 &= E\left\{\int (Z-\mu(t))^2 dN(t)\right\} \\
&= E\left\{\int [Z-\mu(t)]^2 Y(t)\lambda(t,0,V)dt\right\} \\
&= V_L. \qquad (2.9)
\end{aligned}$$

The second term in V_B is

$$\begin{aligned}
V_2 &= 2E\{\int\int_0^{t-} [Z-\mu(t)][Z-\mu(s)][dN(t) - Y(t)\eta(t)dt][dN(s) - Y(s)\eta(s)ds]\} \\
&= 2E\{\int\int_0^{t-} [Z-\mu(t)][Z-\mu(s)] \times \Big[dM(t) + Y(t)(\lambda(t,0,V) - \eta(t))dt\Big] \\
&\qquad \times[dN(s) - Y(s)\eta(s)ds]\} \qquad (2.10)
\end{aligned}$$

where $M(t) = N(t) - \int_0^t Y(s)\lambda(s,0,V)ds$ is an $\mathcal{F}_t$ martingale. Since

$$\int_0^{t-} [Z-\mu(t)][Z-\mu(s)][dN(s) - Y(s)\eta(s)ds]$$

is an $\mathcal{F}_t$ predictable process when $s < t$, we have

$$E\Big\{\int\int_0^{t-}[Z-\mu(t)][Z-\mu(s)][dN(s)-Y(s)\eta(s)ds]dM(t)\Big\} = 0.$$

Hence, from (2.10),

$$\begin{aligned}
V_2 &= 2E\Big\{\int\int_0^{t-}[Z-\mu(t)][Z-\mu(s)][\lambda(t,0,V)-\eta(t)][dN(s)-Y(s)\eta(s)ds]Y(t)dt\Big\} \\
&= 2E\left\{\int[Z-\mu(t)][\lambda(t,0,V)-\eta(t)]E\{Y(t)\int_0^{t-}[Z-\mu(s)]dN(s)\mid Z,V\}dt\right\} \\
&\quad -2\int\int_0^t E\Big\{[Z-\mu(t)][Z-\mu(s)][\lambda(t,0,V)-\eta(t)]\eta(s)Y(t)Y(s)\Big\}dsdt. \qquad (2.11)
\end{aligned}$$

By definition, $Y(t)\int_0^{t-}[Z-\mu(s)]dN(s) = 0$ with probability 1; thus the first term of V_2 in (2.11) is 0. When $s < t$, $Y(s)Y(t) = Y(t)$, so that

$$\begin{aligned}
V_2 &= -2\int\int_0^t E\{[Z-\mu(t)][Z-\mu(s)][\lambda(t,0,V)-\eta(t)]\eta(s)Y(t)dsdt \\
&= -2V_C \qquad (2.12)
\end{aligned}$$

Thus from (2.8)-(2.12) , we conclude

$$V_N = V_L - 2V_C - B^2$$

where V_L is as in (2.14) and B in (2.3). □

2.3 Logrank Test under the Kong-Slud Assumption

In this section we study the large sample null distribution of the logrank rank test under the Kong-Slud Assumption defined in Assumption 1.4.

2.3.1 Kong-Slud Assumption

The Kong-Slud Assumption, defined in Assumption 1.4, assumes the conditional independence of the treatment group indicator and the covariates under H_0 given that the subject is still at risk. As shown in Lemma A.2, Assumption 1.6 implies Assumption 1.5 which implies Assumption 1.4.

From Lemma A.1, under Assumptions 1.1 and 1.3,

$$E_0\{Z \,|\, V, Y(t) = 1\} \quad = \quad E\{Z \mid V, C \geq t\}$$

and the Kong-Slud Assumption (Assumption 1.4) implies

$$\mu(t) \equiv E_0\{Z \mid Y(t) = 1\} \;=\; E\{Z \mid V, C \geq t\}. \tag{2.13}$$

Since the law $\mathcal{L}(Z, V, C)$ does not change under H_0 and H_A (for all null and alternative hypotheses defined in Section 1.3.3), the conditional expectation on the right hand side of (2.13) remains the same under both H_0 and H_A. Hence the Kong-Slud assumption can also be stated as saying that the conditional expectation of the treatment group indicator given the covariates and the censoring indicator at t equals the non-random function $\mu(t)$, the conditional expectation of the treatment group indicator under H_0 given that the subject is still at risk.

2.3.2 Large Sample Null Distribution

We derive the asymptotic distribution of the numerator of the logrank statistic under the null hypothesis.

Lemma 2.3 *When Assumptions 1.1, 1.3 and 1.4 hold, the numerator* $n^{-\frac{1}{2}}\hat{U}_L$ *of the logrank statistic is asymptotically normal distributed with mean 0 and variance* V_L *under the null hypothesis* $H_0 : \theta = 0$*, where* $n^{-\frac{1}{2}}\hat{U}_L$ *and* V_L *are defined in (1.3) and (2.5), respectively.*

Proof.

As a special case of formula (2.6) in Kong and Slud (1997), when Assumptions 1.1, 1.2 and 1.4 hold, the numerator $n^{-\frac{1}{2}}\hat{U}_L$ of the logrank statistic is asymptotically equal to an iid sum under $H_0 : \theta = 0$,

$$\begin{aligned} n^{-\frac{1}{2}}\hat{U}_L(t) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int \{Z_i - \mu(t)\}\big\{dN_i(t) - Y_i(t)E_0[\lambda(t,0,V) \mid Y(t)]\big\} \\ &\quad + o_p(1) \qquad (2.14) \end{aligned}$$

The independence of the terms within the sum on the right hand side of (2.14) is immediate.

Define the filtration

$$\mathcal{G}_t \equiv \sigma\{N_i(s), Y_i(s), Z_i;\ 0 \le s \le t,\ i = 1, 2, \dots\}.$$

Without covariates V_i being observed, $Y_i(t)E_0\{\lambda(t,0,V_i) \mid Y_i(t)\}$ is the intensity of $N_i(t)$ under H_0 with respect to the filtration $\sigma\{N_i(s), Y_i(s); 0 \le s \le t,\ i = 1, 2, \dots\}$, which is a subset of $\mathcal{G}_t$.

When Assumptions 1.1, 1.3 and 1.4 hold, also

$$E_0\{\lambda(t,0,V_i) \mid Y_i(t)\} = E_0\{\lambda(t,0,V_i) \mid Y_i(t), Z_i\}.$$

Therefore these expressions are equal to the $\mathcal{G}_t$ intensity for $N_i(t)$. Thus the process

$$N_i(t) - \int_0^s Y_i(s)E_0\{\lambda(s,V_i) \mid Y_i(s)\}ds$$

is also a $\mathcal{G}_t$ martingale.

Since $Z_i - \mu(t)$ is a $\mathcal{G}_t$ predicable process, from Section II.3.3 of Anderson et al (1992) the process

$$\int_0^t [Z_i - \mu(s)]\{dN_i(s) - Y_i(s)E_0[\lambda(s,0,V_i) \mid Y_i(s)]ds\}$$

is also a $\mathcal{G}_t$ martingale. We denote this martingale as $\tilde{M}_i(t)$. Thus the numerator $n^{-\frac{1}{2}}\hat{U}_L$ is asymptotically an iid sum of martingales with mean 0, from which we can write the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_L$ as

$$V_L = E\left\{\int [Z_1 - \mu(t)][dN_1(t) - Y_1(t)E_0(\lambda(t,0,V_1) \mid Y_1(t))]\right\}^2. \tag{2.15}$$

From Section II.3.2 of Anderson et al (1992), process $\{\tilde{M}_i^2 - \langle \tilde{M}_i, \tilde{M}_i\rangle\}(t)$ is also a $\mathcal{G}_t$ martingale, where

$$\langle \tilde{M}_i, \tilde{M}_i\rangle(t) = \int_0^t [Z_i - \mu(s)]^2 Y_i(s)E_0[\lambda(s,0,V_i) \mid Y_i(t)]dt$$

is the optional variation process of $\tilde{M}$. By Theorem II.3.1 of Anderson et al (1992), the asymptotic variance V_L in (2.15) can finally be simplified to

$$\begin{aligned} V_L &= E_0\{\int [Z - \mu(t)]^2 Y(t)E_0[\lambda(t,V) \mid Y(t), Z]dt\} \\ &= E_0\{\int [Z - \mu(t)]^2 Y(t)\lambda(t,0,V)dt\} \end{aligned} \tag{2.16}$$

which is the same as (2.5).

Finally we know that $n^{-\frac{1}{2}}\hat{U}_L$ is asymptotically an iid sum with mean 0 and finite variance V_L. By the Central Limit Theorem, the asymptotic distribution of $n^{-\frac{1}{2}}\hat{U}_L$ is $N(0, V_L)$. $\square$

Next we will show that $\hat{V}_L$, the square of the denominator of the logrank statistic, is a consistent estimator of V_L.

Lemma 2.4 *If Assumptions 1.1, 1.3 and 1.4 hold, then* $\hat{V}_L \xrightarrow{p} V_L$ *as* $n \to \infty$, *where* $\hat{V}_L$ *is defined in (1.5).*

Proof. From Section 1.3.1 and (1.2), we know that for each i, the process

$$M_i(t) = N_i(t) - \int_0^t Y(s)\lambda(s, 0, V_i)dt$$

is an $\mathcal{F}_t$ martingale under H_0. Recall that

$$\hat{V}_L \equiv \int \frac{1}{n} \cdot \frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)^2} d\bar{N}(t)$$

Based on the definition of $\hat{V}_L$ we can further define

$$\hat{V}_L^{[0,K]} = \int_{[0,K]} \frac{1}{n} \cdot \frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)^2} d\bar{N}(t)$$

and

$$\hat{V}_L^{(K,\infty)} = \int_{(K,\infty)} \frac{1}{n} \cdot \frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)^2} d\bar{N}(t).$$

Evidently $\hat{V}_L = \hat{V}_L^{[0,K]} + \hat{V}_L^{(K,\infty)}$. On the other hand,

$$\begin{aligned}
V_L &= \int E_0\{[Z - \mu(t)]^2 Y(t)\lambda(t, 0, V)\}dt \\
&= \int E_0\{E_0\{[Z - \mu(t)]^2 \mid Y(t) = 1, V\}Y(t)\lambda(t, 0, V)\}dt \\
&= \int E_0\{E_0\{[Z - \mu(t)]^2 \mid Y(t) = 1\}Y(t)\lambda(t, 0, V)\}dt \qquad \text{(by Assumption 1.4)} \\
&= \int E_0\{E_0\{Z^2 - 2Z\mu(t) + \mu^2(t) \mid Y(t) = 1\}Y(t)\lambda(t, 0, V)\}dt \\
&= \int E_0\{\mu(t)[1 - \mu(t)]Y(t)\lambda(t, 0, V)\}dt.
\end{aligned}$$

Similarly we can further define

$$V_L^{[0,K]} = \int_{[0,K]} E_0\{\mu(t)[1-\mu(t)]Y(t)\lambda(t,0,V)\}dt$$

and

$$V_L^{(K,\infty)} = \int_{(K,\infty)} E_0\{\mu(t)[1-\mu(t)]Y(t)\lambda(t,0,V)\}dt$$

with $V_L = V_L^{[0,K]} + V_L^{(K,\infty)}$.

Since $\hat{V}_L^{(K,\infty)} > 0$ and $\sup_{0\le t<\infty}\left|\bar{Y}_1(t)\bar{Y}_0(t)/\bar{Y}(t)^2\right| \le 1$ with probability one , we have

$$\begin{aligned} E|\hat{V}_L^{(K,\infty)}| &\le \frac{1}{n}E\left[\sum_{i=1}^{n}\int_{(K,\infty)} dN_i(t)\right] \\ &= E\left[\int_{(K,\infty)} dN_1(t)\right]. \end{aligned}$$

From (i) of Lemma A.6,

$$E|\hat{V}_L^{(K,\infty)}| \to 0 \text{ as } K\uparrow\infty. \tag{2.17}$$

Since $\mu(t)[1-\mu(t)] < 1$ we can similarly get from (ii) of Lemma A.6 that

$$E|V_L^{(K,\infty)}| \to 0 \text{ as } K\uparrow\infty. \tag{2.18}$$

From Lemma A.3 and the Uniform Law of Large Numbers over a compact set $[0,K]$,

$$\sup_{0\le t<\infty}\left|\frac{1}{n}\bar{N}(t) - E\{N(t)\}\right| \xrightarrow{L^2} 0,$$

and

$$\sup_{0\le t<K}\left|\frac{\bar{Y}_1(t)\bar{Y}_0(t)}{\bar{Y}(t)^2} - \frac{E\{ZY(t)\}E\{(1-Z)Y(t)\}}{[E\{Y(t)\}]^2}\right| \xrightarrow{L^2} 0.$$

Since

$$E_0\{N(t)\} = E_0\{\int_0^t Y(s)\lambda(s,0,V)\}$$

and

$$\frac{E_0\{ZY(t)\}E_0\{(1-Z)Y(t)\}}{[E_0\{Y(t)\}]^2} = \mu(t)[1-\mu(t)],$$

we have

$$\hat{V}_L^{[0,K]} \xrightarrow{p} V_L^{[0,K]} \text{ as } n \to \infty. \tag{2.19}$$

From (2.17), (2.18) and (2.19) we know that for any $\epsilon > 0$, there exist a real number $K > 0$ such that $P\{|\hat{V}_L^{(K,\infty)}| > \epsilon/3 < \epsilon/3$ and $P\{|V_L^{(K,\infty)}| > \epsilon/3\} < \epsilon/3$; for each K, there exist an integer $N > 0$ such that for any $n > N$, $P\{|\hat{V}_L^{[0,K]} - V_L^{[0,K]}| > \epsilon/3\} < \epsilon/3$. Therefore

$$\begin{aligned} P\{|\hat{V}_L - V_L| > \epsilon\} &\leq P\{|\hat{V}_L^{[0,K]} - V_L^{[0,K]}| + |\hat{V}_L^{(0,\infty)}| + |V_L^{(0,\infty)}| > \epsilon\} \\ &\leq P\{|\hat{V}_L^{[0,K]} - V_L^{[0,K]}| > \frac{\epsilon}{3}\} + P\{|\hat{V}_L^{(K,\infty)}| > \frac{\epsilon}{3}\} \\ &\quad + P\{|V_L^{(K,\infty)}| > \frac{\epsilon}{3}\} \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon. \end{aligned}$$

Thus $\hat{V}_L \xrightarrow{p} V_L$ as $n \to \infty$. $\square$

From Lemmas 2.3 and 2.4 we justify the large sample distribution of the logrank test statistic $n^{-\frac{1}{2}}\hat{U}_L$ under the null hypotheses of no treatment effect.

Theorem 2.1 *When Assumptions 1.1, 1.2 and the Kong-Slud Assumption I (Assumption 1.4) hold, the logrank test statistic* $n^{-\frac{1}{2}}\hat{U}_L/\hat{V}_L^{\frac{1}{2}}$ *is asymptotically distributed as standard normal under* $H_0 : \theta = 0$.

Proof. This theorem is a direct result of Lemma 2.3 and 2.4. $\square$

Although we have shown in previous sections that the logrank test with test statistic $n^{-\frac{1}{2}}\hat{U}_L/\hat{V}_L^{\frac{1}{2}}$ is generally biased under the more general Assumption 1.3 instead of 1.2, the above theorem guarantees a bias-free test once all assumptions, especially the important Kong-Slud Assumption (Assumption 1.4), can be verified. The problem of how to verify this assumption is also what we will study in future research. Since the statistic is asymptotically distributed as standard normal under H_0, the test with rejection region $\{|n^{-\frac{1}{2}}\hat{U}_L/\hat{V}_L^{\frac{1}{2}}| > z_\alpha\}$ is an asymptotically correct size α test, where z_α is the standard normal cutoff.

Remark 1

The bias correction "effect" of the Kong-Slud Assumption I (Assumption 1.4) on the logrank test with statistic $n^{-\frac{1}{2}}\hat{U}_L/\hat{V}_L^{\frac{1}{2}}$ can easily be verified here. From Lemma 2.1 we know that the bias of $n^{-\frac{1}{2}}\hat{U}_L$ under H_0 is:

$$E_0\{n^{-\frac{1}{2}}\hat{U}_L\} = \sqrt{n}B + o(\sqrt{n})$$

with

$$B = \int E_0\{[Z - \mu(t)]Y(t)\lambda(t, 0, V)\}dt.$$

and from (2.4),

$$\int E\{Y(t)\} \cdot \text{Cov}\{Z, \lambda(t, 0, V)\} dt.$$

The Kong-Slud Assumption assumes the conditional independence of Z and V given $Y(t) = 1$, hence

$$\text{Cov}\{Z, \lambda(t, 0, V)\} \;=\; 0.$$

Thus there is no bias. Hence there is no bias for $n^{-\frac{1}{2}}\hat{U}_L$ under Assumption 1.4.

Remark 2

DiRienzo and Lagakos (2001b) claim that the logrank test is "asymptotically valid" when the more restrictive Assumption 1.6 holds. From their context, the term "asymptotically valid" means that the test with rejection region $\{|n^{-\frac{1}{2}}\hat{U}_L/\hat{V}_L^{\frac{1}{2}}| > Z_\alpha\}$ can achieve the nominal significance level α. The support they give for this claim is incorrect.

In their Section 3, they claimed that T and C will be unconditionally independent under $H_0 : T \perp\!\!\!\perp Z|V$ when Assumptions 1.3 and 1.6 hold. This claim is wrong because in this setting only the conditional independence $T \perp\!\!\!\perp C|V$ holds, not that $T \perp\!\!\!\perp C$ when $C \perp\!\!\!\perp Z|V$, one of the two possibilities in Assumption 1.6. A quick counter example is as follows. Let the conditional survival function for T given V be $S(t, v) = e^{-\alpha v t}$. Under H_0 of no treatment effect, let the survival function for C be $S_C(t, v) = e^{-\gamma v t}$ so C is independent of Z given V. Then Assumption 1.6 holds. According to the authors' claim, T and C should be unconditionally independent. But from this example, They are not. Therefore the claim of the authors that the "validity" of the logrank test under condition Assumption 1.6 follows from the unconditional independence of T and C is not sufficient. Their assertion about the "validity" is correct since Assumption 1.6 is more restrictive and naturally implies Assumption 1.4, under which we proved the the correct asymptotic standard normal distribution of the logrank statistic.

Remark 3

Though not shown here, the asymptotic null distribution and consistency property of the logrank statistic can be extended to weighted logrank tests with nonrandom or predictable weights. Although usually not a fully efficient test, the logrank test will always be a safe choice, regarding the potential bias, as long as Assumption 1.4 holds.

Chapter 3

Stratified Logrank Tests under Two Types of Stratifications

We will study two types of stratified logrank tests in this chapter. The two tests differ in the degree of stratification. The first test is the one we defined in (1.6) and (1.8), where all components of V are discrete with finitely many values and we stratify on all levels of V. The second test is one applied in cases where not all values of V can can determine strata, such as (i) only part of V is discrete, (ii) to avoid the sample size problem in single stratum, extremely fine stratification is not allowed, (iii) covariate V is completely continuous. As defined in Section 1.3.6, we can stratify W, a function of V with discrete levels, and denote the stratified logrank test based on this stratification as the "W-stratified logrank test". In most of this chapter we assume only the general Assumptions 1.1 and 1.3. The Kong-Slud Assumption is used in studying the large sample null distribution of the W-stratified logrank test.

3.1 Stratified Logrank Test with Completely Discrete Covariate

In this section, we study the asymptotic distribution and consistency of the stratified logrank test defined in (1.6) and (1.8), where V is completely discrete with finitely many values. Note that the numerator $\hat{U}_S$ in (1.6) can be written as a sum of martingales:

$$\begin{aligned}
\hat{U}_S(t) &= \sum_v \Big[\int \frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)} d\bar{M}_{1v}(t) - \int \frac{\bar{Y}_{1v}(t)}{\bar{Y}_{\cdot v}(t)} d\bar{M}_{0v}(t)\Big] \\
&= \sum_v \left[\sum_{i=1}^{n} \frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)} \cdot \xi_i^v dM_{1i}(t) - \sum_{i=1}^{n} \frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)} \cdot \xi_i^v dM_{0i}(t)\right]
\end{aligned}$$

where

$$\bar{M}_{kv}(t) = \bar{N}_{kv}(t) - \int_0^t \bar{Y}_{kv}(s)\lambda_T(s,0,v)ds,$$

$$M_{1i}(t) = N_i(t) - \int_0^t Z_i Y_i(t)\lambda(s,0,V_i)dt,$$

$$M_{0i}(t) = N_i(t) - \int_0^t (1-Z_i)Y_i(t)\lambda(s,0,V_i)dt,$$

for $k = 1,2;\ i = 1,2,\ldots$ are all locally square integrable martingales under Assumptions 1.1 and 1.3. By the martingale central limit theorem, the numerator $\hat{U}_S$ is asymptotically normal with mean 0 under H_0.

Since subjects in different strata are independent, the variance of $n^{-\frac{1}{2}}\hat{U}_S$ can be calculated as:

$$\begin{aligned}
V_S &\equiv \frac{1}{n}\sum_v E\left\{\Big[\int \frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)} d\bar{M}_{1v}(t) - \int \frac{\bar{Y}_{1v}(t)}{\bar{Y}_{\cdot v}(t)} d\bar{M}_{0v}(t)\Big]^2\right\} \\
&= \frac{1}{n}\Big\{\sum_v E\Big[\int [\frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)}]^2 d\langle \bar{M}_{1v}, \bar{M}_{1v}\rangle(t) + \int [\frac{\bar{Y}_{1v}(t)}{\bar{Y}_{\cdot v}(t)}]^2 d\langle \bar{M}_{0v}, \bar{M}_{0v}\rangle(t) \\
&\qquad - 2\int \frac{\bar{Y}_{0v}(t)}{\bar{Y}_{\cdot v}(t)} \cdot \frac{\bar{Y}_{1v}(t)}{\bar{Y}_{\cdot v}(t)} d\langle \bar{M}_{1v}, \bar{M}_{0v}\rangle(t)\Big]\Big\},
\end{aligned}$$

where $\langle \bar{M}_{lv}, \bar{M}_{mv}\rangle$ is the compensator of process $\bar{M}_{lv}\bar{M}_{mv}$ for $l, m \in \{0,1\}$.

For each $V = v$, $M_{1v}(\cdot)$ is independent of $M_{0v}(\cdot)$, so $d\langle M_{1v}, M_{0v}\rangle(t) = 0$ for all t. Under H_0,

$$d\langle \bar{M}_{1v}, \bar{M}_{1v}\rangle(t) = \bar{Y}_{1v}(t)\lambda(t,0,V)dt,$$

$$d\langle \bar{M}_{0v}, \bar{M}_{0v}\rangle(t) = \bar{Y}_{0v}(t)\lambda(t, 0, V)dt.$$

Then

$$V_S = \frac{1}{n}\sum_v E\left\{\int \frac{\bar{Y}_{0v}(t)\bar{Y}_{1v}(t)}{\bar{Y}_v^2(t)} \cdot \bar{Y}_v(t)\lambda(t, 0, v)dt\right\}. \tag{3.1}$$

Hence a consistent estimator for V_S will be

$$\hat{V}_S = \frac{1}{n}\sum_v \int \frac{\bar{Y}_{0v}(t)\bar{Y}_{1v}(t)}{\bar{Y}_v^2(t)} d\bar{N}_v(t) \tag{3.2}$$

which is the squared denominator of the stratified logrank test as defined in (1.8). Thus the stratified logrank statistic $n^{-\frac{1}{2}}\hat{U}_S/(\hat{V}_S)^{\frac{1}{2}}$ is asymptotically standard normally distributed for large n under H_0.

Note that in each stratum $V = v$, the v terms of the stratified logrank statistic $\hat{U}_S$ can be considered as the logrank statistic discussed in Section 2.1. Thus we know that the logrank test with statistic $n^{-\frac{1}{2}}\hat{U}_S$ is consistent against the alternative of stochastic ordering $H_A : S(t|1, v) \leq S(t|1, v)$, for all t and v.

We summarize in the following proposition to conclude this section:

Proposition 3.1 *Assume V is completely discrete and finite valued. If Assumption 1.1 and 1.3 hold, then the stratified logrank test with statistic $n^{-\frac{1}{2}}\hat{U}_S/(\hat{V}_S)^{\frac{1}{2}}$ is (i) asymptotically standard normally distributed under the null hypothesis H_0 of no treatment effect and (ii) consistent against the alternative of stochastic ordering $H_A : S(t|1, v) \leq S(t|1, v)$ for all t and v.* □

3.2 The W-Stratified Logrank with the Kong-Slud Assumption

Though the difference in definitions between the W-stratified logrank (see Section 1.3.6) and stratified logrank statistic(see Section 1.3.5) is only at the subscript or superscript of v or w, for patients in a stratum with $W = w$, the hazard function becomes $E\{\lambda(t, 0, V_i)|W_i = w\}$, which is not homogeneous across this stratum. Note in this section, V_i need not be assumed discrete.

Next we derive the large sample null distribution of the W-stratified logrank test. Define

$$\lambda'(t, 0, v) = \frac{\partial}{\partial \theta}\lambda(t, \theta z, v)\Big|_{\theta=0, Z=1}$$

and

$$\Lambda'(t, 0, V) = \int_0^t \lambda'(s, 0, V)ds = \frac{\partial}{\partial \theta}\lambda(t, \theta z, V)\Big|_{\theta=0, z=1}.$$

Lemma 3.1 *When Assumptions 1.1, 1.3 and 1.4 hold, the numerator of the W-stratified logrank test with test statistic $n^{-\frac{1}{2}}\hat{U}_W$ defined in Section 1.3.6 is asymptotically normal distributed with mean 0 and variance V_W under the null hypothesis of no treatment effect, where $V_W = V_L$ and V_L is the asymptotic variance of the numerator $n^{-\frac{1}{2}}\hat{U}_L$ of the logrank statistic.*

Proof.

(i) In the first part of the proof, we derive the null distribution of the numerator of the logrank statistic under H_0.

Under the contiguous alternative $H_n : \theta_n = b/\sqrt{n}$,

$$M_i(t) \;=\; N(t) - \int_0^t Y_i(s)\lambda(s, \frac{b}{\sqrt{n}}Z_i, V_i)ds$$

is a locally square integrable martingale, for $i = 1, 2, \ldots, n$. Therefore, for a fixed stratum $W = w$, one has

$$\begin{aligned} n^{-\frac{1}{2}}\hat{U}_W^{w,[0,K]} &= n^{-\frac{1}{2}}b\int_{[0,K]} \xi_i^w \left\{ Z_i - \frac{\sum_{i=1}^n \xi_i^w Z_i Y_i(t)\lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)} \right\} dM_i(t) \\ &\quad + n^{-\frac{1}{2}}\int_{[0,K]} \left\{ \frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)} - \frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i}{\sum_{i=1}^n \xi_i^w Y_i(t)} \right\} d\bar{N}_w(t) \\ &\equiv A_1 + A_2 \end{aligned} \tag{3.3}$$

where $\bar{N}_w(t) \equiv \sum_{i=1}^n \xi_i^w \cdot N_i(t)$.

Under Assumption 1.4, Z and V are conditionally independent given $Y(t) = 1$ under $H_0 : \theta = 0$. One shows by Law of Large Numbers that for each t

$$\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i}{\sum_{i=1}^n \xi_i^w Y_i(t)} \xrightarrow{p} \mu(t)$$

and

$$\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t, 0, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t, 0, V_i)} \xrightarrow{p} \mu(t) \tag{3.4}$$

under H_0. As a consequence, under contiguous alternatives we can also show that (3.4) holds when we replace 0 by $n^{-\frac{1}{2}}bZ_i$.

From Lemmas A.3 and A.12, for $K > 0$,

$$\sup_{0 \le t \le K} \left| \frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)} - \mu(t) \right| \xrightarrow{p} 0,$$

Therefore by Lemma A.13, we have

$$n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^K \xi_i^w \left\{ \frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t, n^{-\frac{1}{2}}bZ_i, V_i)} - \mu(t) \right\} dM_i(t) \xrightarrow{p} 0.$$

Thus, A_1 can be rewritten as

$$A_1 = n^{-\frac{1}{2}} \sum_{i=1}^n \int_{[0,K]} \xi_i^w \{Z_i - \mu(t)\} dM_i(t) + o_p(1) \tag{3.5}$$

To consider A_2, note that

$$\lambda(t, n^{-\frac{1}{2}} b Z_i, V_i) = \lambda(t, 0, V_i) + n^{-\frac{1}{2}} b Z_i \lambda'(t, 0, V_i) + R_{1i}^{(n)}(t)$$

with

$$R_{1i}^{(n)}(t) = \frac{b^2}{2} n^{-\frac{1}{2}} Z_i [\lambda'(t, \epsilon_i Z_i, V_i) - \lambda'(t, 0, V_i)],$$

where $\epsilon_i = \epsilon_i^{(n)}$ is such that $\epsilon_i^{(n)} b > 0$ and $0 < |\epsilon_i^{(n)}| < n^{-\frac{1}{2}}|b|$. By the continuity of and uniform integrability of function $\lambda'(t, \cdot, V)$, we have

$$E\{\sup_{0 \le t < K} |\lambda'(t, \epsilon_i^{(n)} Z_i, V_i) - \lambda'(t, 0, V_i)|\} \to 0.$$

Hence we have

$$R_{1i}^{(n)}(t) = o_{p,L1}(n^{-\frac{1}{2}}), \text{uniformly over } t \in [0, K].$$

With this expansion, we have

$$\begin{aligned}
& \frac{\sum_{i=1}^n \xi_i^w Z_i Y_i(t) \lambda(t, \frac{b}{\sqrt{n}} Z_i, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t) \lambda(t, n^{-\frac{1}{2}} b Z_i, V_i)} \\
= \quad & \frac{\sum_{i=1}^n \xi_i^w Y_i(t) \lambda(t, 0, V_i)}{\sum_{i=1}^n \xi_i^w Z_i Y_i(t) \lambda(t, 0, V_i)} + b n^{-\frac{1}{2}} \Big[\frac{\sum_{i=1}^n \xi_i^w Z_i^2 Y_i(t) \lambda'(t, 0, V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t) \lambda(t, 0, V_i)} \\
& - \frac{\{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t, 0, V_i)\}\{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda'(t, 0, V_i)\}}{\{\sum_{i=1}^n \xi_i^w Y_i(t) \lambda(t, 0, V_i)\}^2} \Big] + o_p(n^{-\frac{1}{2}}) \\
= \quad & B_{21} + B_{22} \ + \ o_p(n^{-\frac{1}{2}}) \qquad (3.6)
\end{aligned}$$

where the $o_p(n^{-\frac{1}{2}})$ in (3.6) is uniform over the compact set $[0, K]$.

From Lemma A.3, we have that under H_0

$$\sup_{0 \le t < \infty} |\frac{1}{n} \bar{N}_w(t) - E_0\{\xi_i^w N_1(t)\}| \xrightarrow{p_0} 0$$

where $E_0\{\xi_i^w N_1(t)\} = \int_0^t E_0[\xi_1^w Y(s) \lambda(s, 0, V)] ds$.

Since the probability sequence P_{θ_n} under H_n is contiguous with respect to P_0, we also have

$$\sup_{0\le t<\infty}\left|\frac{1}{n}\bar{N}_w(t)-\int_0^t E_0[\xi_1^w Y(s)\lambda(s,0,V)]ds\right| \xrightarrow{p_{\theta_n}} 0$$

That is, over $t\in[0,K]$,

$$\frac{1}{n}\bar{N}_w(t)=\int_0^t E_0[\xi_1^w Y(s)\lambda(s,0,V)]ds+o_p(1). \tag{3.7}$$

By the Law of Large Numbers, the first term in B_{22}

$$\frac{\sum_{i=1}^n \xi_i^w Z_i^2 Y_i(t)\lambda'(t,0,V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t,0,V_i)} \xrightarrow{P_0,P_{\theta_n}} \frac{E_0[\xi^w Z^2 Y(t)\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda(t,0,V)]}. \tag{3.8}$$

Similarly, for the second term in B_{22}, we have

$$\begin{aligned}&\frac{\{\sum_{i=1}^n \xi_i^w Y_i(t)Z_i\lambda(t,0,V_i)\}\{\sum_{i=1}^n \xi_i^w Y_i(t)Z_i\lambda'(t,0,V_i)\}}{\{\sum_{i=1}^n \xi^w Y_i(t)\lambda(t,0,V_i)\}^2}\\ &\xrightarrow{P_0,P_{\theta_n}} \frac{E_0[\xi^w Y(t)Z\lambda(t,0,V)]E_0[\xi^w Y(t)Z\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda(t,0,V)]^2}\end{aligned} \tag{3.9}$$

Thus, from (3.7), (3.8) and (3.9), we have

$$\begin{aligned}
& n^{-\frac{1}{2}}\int_0^K B_{22}d\bar{N}_w(t)\\
= \;& b\int_0^K \Big[\frac{E_0[\xi^w Z^2 Y(t)\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda(t,0,V)]}-\frac{E_0[\xi^w Y(t)Z\lambda(t,0,V)]E_0[\xi^w Y(t)Z\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda(t,0,V)]^2}\Big]\\
& \qquad \times E_0[\xi^w Y(t)\lambda(t,0,V)]dt\; + o_p(1)\\
= \;& b\int_0^K E_0[\xi^w Y(t)\lambda'(t,0,V)]\Big[\frac{E_0[\xi^w Z^2 Y(t)\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda'(t,0,V)]}-\frac{E_0[\xi^w ZY(t)\lambda(t,0,V)]}{E_0[\xi^w Y(t)\lambda(t,0,V)]}\\
& \qquad \times\frac{E_0[\xi^w ZY(t)\lambda'(t,0,V)]}{E_0[\xi^w Y(t)\lambda'(t,0,V)]}\Big]dt+o_p(1)\\
= \;& b\int_0^K \mu(t)[1-\mu(t)]E_0[\xi^w Y(t)\lambda'(t,0,V)]dt+o_p(1).
\end{aligned} \tag{3.10}$$

For the rest of A_2, we have

$$\begin{aligned}
& n^{-\frac{1}{2}} \int_0^K \left[\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t,0,V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t,0,V_i)} - \frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i}{\sum_{i=1}^n \xi_i^w Y_i(t)} \right] d\bar{N}_w(t) \\
= \; & n^{-\frac{1}{2}} \int_0^K \xi_i^w \Big[\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t,0,V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t,0,V_i)} - \mu(t) \Big] dM_i(t) \\
& - n^{-\frac{1}{2}} \int_0^K \xi_i^w \Big[\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i}{\sum_{i=1}^n \xi_i^w Y_i(t)} - \mu(t) \Big] dM_i(t) \\
& + n^{-\frac{1}{2}} \int_0^K \xi_i^w \Big[\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i \lambda(t,0,V_i)}{\sum_{i=1}^n \xi_i^w Y_i(t)\lambda(t,0,V_i)} - \mu(t) \Big] Y_i(t)\lambda(t,0,V_i) dt \\
& - n^{-\frac{1}{2}} \int_0^K \xi_i^w \Big[\frac{\sum_{i=1}^n \xi_i^w Y_i(t) Z_i}{\sum_{i=1}^n \xi_i^w Y_i(t)} - \mu(t) \Big] Y_i(t)\lambda(t,0,V_i) dt \\
= \; & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^K \xi_i^w \Big[Z_i - \mu(t) \Big] \Big[\lambda(t,0,V_i) - \frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]} \Big] dt \\
& + o_p(1) \qquad (3.11)
\end{aligned}$$

Thus, from (3.3) to (3.11),

$$\begin{aligned}
n^{-\frac{1}{2}} \hat{U}_W^{w,[0,K]} \; = \; & n^{-\frac{1}{2}} \sum_i^n \int_0^K \xi_i^w (Z_i - \mu(t)) \left[dN_i(t) - Y_i(t) \frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]} dt \right] \\
& + b \int_0^K \sigma_Z^2(t) E_0 \Big[\xi^w Y(t) \lambda'(t,0,V) \Big] dt + o_p(1)
\end{aligned}$$

Since $\sum_w \xi_i^w = 1$ by its definition, for each i, the numerator of the W-stratified logrank statistic under H_n becomes

$$\begin{aligned}
n^{-\frac{1}{2}} \hat{U}_W^{[0,K]} \; = \; & n^{-\frac{1}{2}} \sum_w \hat{U}_W^{w,[0,K]} \\
= \; & \sum_w n^{-\frac{1}{2}} \sum_i^n \int_0^K \xi_i^w (Z_i - \mu(t)) \left[dN_i(t) - Y_i(t) \frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]} dt \right] \\
& + b \int_0^K \sigma_Z^2(t) E_0[Y(t)\lambda'(t,0,V)] dt + o_p(1) \qquad (3.12)
\end{aligned}$$

Hence the numerator of the logrank statistic under H_0 is

$$
\begin{aligned}
n^{-\frac{1}{2}}\hat{U}_W^{[0,K]} &= \sum_w n^{-\frac{1}{2}} \sum_i^n \int_0^K \xi_i^w [Z_i - \mu(t)] \left[dN_i(t) - Y_i(t) \frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]} dt \right] \\
&\qquad + o_p(1) \\
&\equiv n^{-\frac{1}{2}} U_W^{[0,K]} + o_p(1) \qquad (3.13)
\end{aligned}
$$

From Lemma A.8 we know that

$$
n^{-\frac{1}{2}}\hat{U}_W^{(K,\infty)} \overset{p,L_1}{\longrightarrow} 0 \text{ as } K \uparrow \infty
$$

$$
n^{-\frac{1}{2}}U_W^{(K,\infty)} \overset{p,L_1}{\longrightarrow} 0 \text{ as } K \uparrow \infty.
$$

Thus for any $\epsilon > 0$, there exist a large number $K > 0$ such that

$$
P\{\left| n^{-\frac{1}{2}}\hat{U}_W^{(K,\infty)} \right| > \frac{\epsilon}{3}\} \;<\; \frac{\epsilon}{3},
$$

$$
P\{\left| n^{-\frac{1}{2}}U_W^{(K,\infty)} \right| > \frac{\epsilon}{3}\} \;<\; \frac{\epsilon}{3}
$$

uniformly over all n. Then from (3.13) we also know that for each such K, there exist an integer $N > 0$ such that for all $n \geq N$ we have

$$
P\left\{ \left| n^{-\frac{1}{2}}\hat{U}_W^{[0,K]} - n^{-\frac{1}{2}}U_W^{[0,K]} \right| > \frac{\epsilon}{3} \right\} \;<\; \frac{\epsilon}{3}.
$$

Therefore

$$
\begin{aligned}
& P\left\{ \left| n^{-\frac{1}{2}}\hat{U}_W - n^{-\frac{1}{2}}U_W \right| > \epsilon \right\} \\
\leq\;& P\left\{ \left| n^{-\frac{1}{2}}\hat{U}_W^{[0,K]} - n^{-\frac{1}{2}}U_W^{[0,K]} \right| + \left| n^{-\frac{1}{2}}\hat{U}_W^{(K,\infty)} \right| + \left| n^{-\frac{1}{2}}U_W^{(K,\infty)} \right| > \epsilon \right\} \\
\leq\;& P\left\{ \left| n^{-\frac{1}{2}}\hat{U}_W^{[0,K]} - n^{-\frac{1}{2}}U_W^{[0,K]} \right| > \frac{\epsilon}{3} \right\} + P\left\{ \left| n^{-\frac{1}{2}}\hat{U}_W^{(K,\infty)} \right| > \frac{\epsilon}{3} \right\} + P\left\{ \left| n^{-\frac{1}{2}}U_W^{(K,\infty)} \right| > \frac{\epsilon}{3} \right\} \\
<\;& \frac{\epsilon}{3} \;+\; \frac{\epsilon}{3} \;+\; \frac{\epsilon}{3}. \\
=\;& \epsilon
\end{aligned}
$$

Hence we have

$$n^{-\frac{1}{2}}\hat{U}_W \ - \ n^{-\frac{1}{2}}U_W \ \xrightarrow{p} \ 0. \tag{3.14}$$

Under the Kong-Slud Assumption 1.4,

$$E_0\Big\{ \int \xi_i^w (Z_i - \mu(t))[dN_i(t) - Y_i(t)\frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]}dt]\Big\} \ = \ 0.$$

Hence under H_0, $\hat{U}_W^w$ is asymptotically the sum of iid distributed random variables with mean 0, provided the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_W^w$ is finite. Then by the Central Limit Theorem, $n^{-\frac{1}{2}}\hat{U}_W^w$ is asymptotically normally distributed with mean 0 under H_0.

Since $n^{-\frac{1}{2}}\hat{U}_W$ is the finite sum of $n^{-\frac{1}{2}}\hat{U}_W^w$, we also have that the numerator of the logrank statistic is asymptotically normally distributed with mean 0 and variance V_W under H_0.

(ii) In the second part of the proof, we show $V_W = V_L$.

A formula for the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_W$ based on the sum of iid terms with mean 0 in $n^{-\frac{1}{2}}U_W$ is:

$$E_0\left\{\Big[\int \xi_i^w [Z - \mu(t)]\{dN(t) - Y(t)E[\lambda(t,0,V)|Y(t), W = w]dt\}\Big]^2\right\}. \tag{3.15}$$

From Assumption 1.4 and the uniqueness of the Doob-Meyer decomposition, we know that $Y(t)E[\lambda(t,0,V)|Y(t), W \ = \ w]$ is the intensity of $N(t)$ under the filtration

$$\mathcal{G}_t^w \equiv \sigma\{N_i(s), Y_i(s), W_i, Z_i; 0 \le s \le t, i = 1, 2, \dots\}.$$

Thus the process

$$\int_0^t \xi^w [Z - \mu(s)]\{dN(s) - Y(t)E[\lambda(t,0,V)|Y(t), W = w]dt\}$$

is a $\mathcal{G}_t^w$ martingale. So the asymptotic variance (3.15) can be simplified to

$$\begin{aligned} & E_0\left\{\int \xi^w[Z-\mu(t)]^2 Y(t) E_0[\lambda(t,0,V)|Y(t),Z,W=w]\right\} \\ = & \int E_0\{\xi^w[Z-\mu(t)]^2 Y(t)\lambda(t,0,V)\}dt \end{aligned} \tag{3.16}$$

Now V_W is the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_W$, and by the independence over strata, V_W is the sum of (3.16) over different strata $W = w$. Recalling that $\xi_i^w = I_{[W_i=w]}$ and $\sum_w \xi_i^w = 1$, we have

$$\begin{aligned} V_W &= \int \sum_w E\{\xi_i^w \cdot [Z-\mu(t)]^2 Y(t)\lambda(t,0,V)\}dt \\ &= \int E\{[Z-\mu(t)]^2 Y(t)\lambda(t,0,V)\}dt \\ &= V_L. \end{aligned} \tag{3.17}$$

Hence the second part of this lemma is also proved. □

Equation (3.17) shows that numerators of the logrank $n^{-\frac{1}{2}}\hat{U}_L$ and of the W-stratified logrank $n^{-\frac{1}{2}}\hat{U}_W$ have the same asymptotic variance under the Kong-Slud Assumption. Thus to show the W-stratified logrank test is asymptotically valid in this case, it is sufficient to know that square of the denominator of the W-stratified logrank

$$\hat{V}_W = \frac{1}{n}\sum_w \frac{\bar{Y}_{0w}(t)\bar{Y}_{1w}(t)}{\bar{Y}_w(t)^2} d\bar{N}_w(t)$$

is asymptotically equivalent to that of the logrank statistic. We will prove this in the following lemma.

Lemma 3.2 *If Assumptions 1.1, 1.3 and 1.4 all hold, then* $\hat{V}_L - \hat{V}_W = o_p(1)$ *under* $H_0 : \theta = 0$.

Proof.

By the Uniform Law of Large Numbers over a compact set , for each stratum $W = w$ and $K > 0$, under Assumption 1.4 and H_0,

$$\sup_{0 \le t < K} \left| \frac{\bar{Y}_{0w}(t)\bar{Y}_{1w}(t)}{\bar{Y}_w(t)^2} - \mu(t)[1-\mu(t)] \right| \overset{p,L^2}{\to} 0.$$

Define

$$\sigma_Z^2(t) = \mu(t)[1-\mu(t)].$$

Thus

$$\begin{aligned} \hat{V}_W^{[0,K]} &= \int_0^K \sum_w \frac{1}{n}\sigma_Z^2(t) d\bar{N}_w(t) + \int_0^K \frac{1}{n} \sum_w \sum_{j \in I(w)} \left[\frac{\bar{Y}_{1w}(t)\bar{Y}_{0w}(t)}{\bar{Y}_w(t)^2} - \sigma_Z^2(t) \right] dN_j(t) \\ &= \int_0^K \sigma_Z^2(t) \frac{d\bar{N}(t)}{n} + o_p(1). \end{aligned}$$

Since $\sup_{0<t<\infty}\{\bar{Y}_{1w}(t)\bar{Y}_{0w}(t)/\bar{Y}_w(t)^2\} < 1$ with probability one and $\sup_{0<t<\infty}\sigma_Z^2(t) < 1$, from Lemma A.6 it follows that

$$\hat{V}_W^{(K,\infty)} \overset{p,L_1}{\to} 0 \text{ as } K \uparrow \infty$$

and

$$\int_K^\infty \sigma_Z^2(t) \frac{d\bar{N}(t)}{n} \overset{p,L_1}{\to} 0 \text{ as } K \uparrow \infty.$$

Hence using similar reasoning as in the proof of Lemma 3.1 leading to (3.14), we have

$$\hat{V}_W = \int \sigma_Z^2(t) \frac{d\bar{N}(t)}{n} + o_p(1),$$

while from the proof of lemma 2.4, we already know that

$$\hat{V}_L^{[0,K]} = \int_0^K \sigma_Z^2(t) \frac{d\bar{N}(t)}{n} + o_p(1).$$

Finally, we have the asymptotic equivalence of $\hat{V}_L$ and $\hat{V}_G$:

$$\hat{V}_L - \hat{V}_W = o_p(1).$$

This lemma is proved. $\square$.

Corollary 3.1 *Under Assumptions 1.1, 1.3 and 1.4,* $\hat{V}_W \xrightarrow{p} V_W$*.*

Proof.

This corollary is an immediate result of Lemma 2.4, Lemma 3.2 and Equation (3.17). Hence we can conclude that the square of the denominator of the W-stratified logrank statistic is a consistent estimator of the asymptotic variance for its numerator. $\square$

Lemma 3.1 and Corollary 3.1 provide the large sample null distribution of the W-Stratified logrank statistic:

Theorem 3.1 *When Assumptions 1.1, 1.3 and 1.4 hold, the W-stratified logrank statistic* $n^{-\frac{1}{2}}\hat{U}_W/\hat{V}_W^{\frac{1}{2}}$ *is asymptotically standard normally distributed.*

Thus the W-Stratified logrank test with rejection region $\{\left|n^{-\frac{1}{2}}\hat{U}_W/\hat{V}_W^{\frac{1}{2}}\right| > Z_{\alpha/2}\}$ can achieve the nominal significance level α under H_0.

3.3 Comparisons

In this section we will compare the three test statistics, logrank $n^{-\frac{1}{2}}\hat{U}_L$, stratified logrank $n^{-\frac{1}{2}}\hat{U}_S$ and W-stratified logrank $n^{-\frac{1}{2}}\hat{U}_W$ in terms of alternative mean and asymptotic relative efficiencies. Unless mentioned separately, all comparisons are made under Assumptions 1.1, 1.3 and 1.4 and assume covariate V to be discrete with finitely many values.

First we give two formulas for the asymptotic means of the numerators of the logrank and stratified logrank statistic under the contiguous alternative $H_n : \theta = b/\sqrt{n}$.

From (3·1) in Kong and Slud (1997), the asymptotic mean of the numerator of $n^{-\frac{1}{2}}\hat{U}_L$ the logrank statistic under H_n can be written as:

$$\begin{aligned} bE^L_{alt} = \;& b \int \sigma^2_Z(t) E_0\big\{Y(t)\lambda'(t,0,V)\big\}dt \\ -\;& b\int \sigma^2_Z(t)E_0\left\{Y(t)\Lambda'(t,0,V)\left[\lambda(t,0,V) - \frac{E_0\{Y(t)\lambda(t,0,V)\}}{E_0\{Y(t)\}}\right]\right\}dt. \end{aligned} \tag{3.18}$$

From the proof of Lemma 3.1, Equation (3.12) and the result of Lemma A.14 (Lemma A·1 of Kong and Slud, 1997), the asymptotic alternative mean of the numerator $n^{-\frac{1}{2}}\hat{U}_W$ of the W-stratified logrank statistic is:

$$\begin{aligned} bE^W_{alt} = \;& b\int \sigma^2_Z(t)E_0\{Y(t)\lambda'(t,0,V)\}dt \\ -\;& b\int \sigma^2_Z(t)E_0\{Y(t)\lambda(t,0,V)\Lambda'(t,0,V)\}dt \\ +\;& \sum_w b\int \sigma^2_Z(t)E_0\{\xi^w Y(t)\Lambda'(t,0,V)\}\frac{E_0[\xi^w Y(t)\lambda(t,0,V)]}{E_0[\xi^w Y(t)]}dt. \end{aligned} \tag{3.19}$$

As a special case of the W-stratified logrank with $W = V$, we can derive the alternative mean of the numerator $n^{-\frac{1}{2}}\hat{U}_W$ of the stratified logrank statistic from

(3.19) as:

$$bE_{alt}^S = b \cdot \int \sigma_Z^2(t) E_0\{Y(t)\lambda'(t,0,V)\} dt. \tag{3.20}$$

From (3.18), (3.20) and (3.19) we can find that the difference of alternative mean for the logrank and stratified logrank is

$$\begin{aligned} & bE_{alt}^S - bE_{alt}^L \\ = \; & b \int \sigma_Z^2(t) \cdot E_0\{Y(t)\} \cdot \mathrm{Cov}\Big\{ \lambda(t,0,V), \Lambda'(t,0,V) \;\Big|\; Y(t) = 1 \Big\} dt. \end{aligned} \tag{3.21}$$

and the one for W-stratified logrank and stratified logrank is

$$\begin{aligned} & bE_{alt}^S - bE_{alt}^W \\ = \; & \sum_w b \int \sigma_Z^2(t) E_0\{\xi^w Y(t)\} \cdot \mathrm{Cov}\Big\{ \lambda(t,0,V), \Lambda'(t,0,V) \;\Big|\; W = w, Y(t) = 1 \Big\} dt \end{aligned} \tag{3.22}$$

3.3.1 Homogeneous Model

We say a sample is homogeneous if all patients in the study have the same level of covariates, so the hazard function can be written as $\lambda(t, \theta z, v_0)$ with a non-random vector v_0 for all patients. If the true model is homogenous and the stratification is not necessary, and from (3.18),(3.20) and (3.19) we can find that

$$E_{alt}^L = E_{alt}^S = E_{alt}^W \;=\; \int \sigma_Z^2(t) \lambda'(t,0,v_0) E_0\{Y(t)\} dt$$

Hence these three test statistics have equal alternative means when the true model is a homogeneous model. From Section (3.2) we also know the three test statistics have equal asymptotic variances, and hence they are equally efficient.

3.3.2 Cox Proportional Hazard Model

When the true model is a Cox proportional hazard model,

$$\lambda(t, \theta z, v) = \lambda_0(t) h(\beta, v) e^{\theta z} \tag{3.23}$$

where $\lambda_0(t)$ is a nonrandom nuisance hazard-intensity and β is a q-dimensional vector. Assume V is discrete with finite values. Then taking $z = 1$, we have

$$\lambda'(t, 0, V) = \lambda(t, 0, V) = \lambda_0(t) h(\beta, V)$$

and

$$\Lambda'(t, 0, V) = \Lambda_0(t) h(\beta, V)$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$.

Then the conditional covariance in (3.21),

$$\text{Cov}\Big\{\lambda(t, 0, V), \Lambda'(t, 0, V) \;\Big|\; Y(t) = 1\Big\} \;=\; \lambda_0(t)\Lambda_0(t)\text{Var}\Big\{h(\beta, V)\Big|Y(t) = 1\Big\},$$

is nonnegative or strictly positive, if V is nondegenerate. Thus

$$E^S_{alt} \;>\; 0 \;\text{ and }\; E^S_{alt} - E^L_{alt} \;>\; 0$$

and hence

$$\{bE^S_{alt}\}^2 \;>\; \{bE^L_{alt}\}^2.$$

Similarly, we also have

$$\{bE^S_{alt}\}^2 \;>\; \{bE^W_{alt}\}^2.$$

Finally it can be concluded that under the Cox proportional hazard model (3.23), the stratified logrank is most efficient among the three test statistics.

3.3.3 Accelerated Failure Model

Suppose the true model is an accelerated life model,

$$\log T = \theta z + \gamma v + \epsilon,$$

where both γ and v are q-dimensional vectors and ϵ is an log-logistic distributed random variable. The hazard function under this model becomes

$$\lambda(t, \theta z, v) = \frac{\exp\{\theta z + \gamma v\}}{1 + t\exp\{\theta z + \gamma v\}} \tag{3.24}$$

Note that

$$\begin{aligned}
\lambda'(t, 0, v) &= \frac{e^{\gamma v}}{\{1 + t\exp\{\gamma v}\}^2\}, \\
\Lambda'(t, 0, v) &= \frac{te^{\gamma v}}{1 + t\exp\{\gamma v\}} = t \cdot \lambda(t, 0, v)
\end{aligned}$$

Thus the conditional covariance

$$\mathrm{Cov}\Big\{\lambda(t, 0, V), \Lambda'(t, 0, V) \;\Big|\; Y(t) = 1\Big\} \;=\; t \cdot \mathrm{Var}\Big\{\lambda(t, 0, V)\Big| Y(t) = 1\Big\}$$

is non negative. If V is nondegerate, again we have

$$E^S_{alt} > 0 \;\text{ and }\; E^S_{alt} - E^L_{alt} > 0$$

and hence

$$\{bE^S_{alt}\}^2 > \{bE^L_{alt}\}^2.$$

Similarly,

$$\{bE^S_{alt}\}^2 > \{bE^G_{alt}\}^2.$$

Thus we know that under the accelerated failure model (3.24), the stratified logrank is still the most efficient one among the three tests.

From the above three examples and formula (3.18)-(3.22), we can find a sufficient condition for the stratified logrank to be more efficient than both the logrank and W-stratified logrank.

Proposition 3.2 *Denote η_1 and η_2 as the asymptotic relative efficiency of stratified logrank versus logrank and stratified logrank versus W-stratified logrank, respectively. Both η_1 and η_2 are with respect to the contiguous alternative $H_n : \theta = b/\sqrt{n}$. Assume Assumptions 1.1, 1.3 and 1.4 hold and V is discrete with finite values. If $\lambda'(t, 0, V) \geq 0$ and $\Lambda'(t, 0, V)$ is positively correlated with $\lambda(t, 0, V)$, or if $\lambda'(t, 0, V) \leq 0$ and $\Lambda'(t, 0, V)$ is negatively correlated to $\lambda(t, 0, V)$, for all t conditionally given that $Y(t) = 1$, that is, for all t,*

$$\lambda'(t, 0, V) \cdot Cov\{\lambda(t, 0, V), \Lambda'(t, 0, V)|Y(t) = 1\} \;\geq\; 0,$$

then

$$\eta_1 \geq 1\ ;\ \eta_2 \geq 1$$

with strict $>$ if V is non-degenerate.

Note that the above theorem is only true under Assumption 1.4 with large sample size n and fixed number of strata n_v. The efficiency of the stratified logrank will be undermined if n_v is very large. We will discuss this topic in Chapter 6.

3.4 Remark

Both the logrank and the stratified or W-stratified logrank test statistics are asymptotically distributed as standard normal under Assumption 1.4 and H_0. We found that the denominators of both the logrank and the W-stratified logrank test statistics are asymptotically equivalent, but the two statistics may not be equivalent because the difference of numerators may not be neglected:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}(\hat{U}_L - \hat{U}_S) \\
= \quad &\int \sum_v \left[\left\{ \lambda(t,0,V) - \frac{E_0\{Y(t)\lambda(t,0,V)\}}{E_0\{Y(t)\}} \right\} \frac{1}{\sqrt{n}} \sum_{j \in I(v)} \{Z_j - \mu(t)\} Y_j(t) \right] dt + o_p(1),
\end{aligned}
$$

where the integral is a strictly positive random variable , even in the limit as $n \to \infty$. When the true model (with covariate V) is misspecified as model-free, neither the plain logrank test nor the stratified logrank test is the optimum test. Thus the fact that the two numerators have difference $O_p(1)$ while the two denominators are asymptotic equivalent does not violate the Hajek convolution theorem. When the true model has no covariate V at all, these logrank tests are the optimum tests, then from the above formula we easily have $E\{I_{[V=v]} \cdot [Z - \mu(t)]^2 Y(t)\lambda(t,V)\}dt = o_p(1)$, which agrees with the convolution theorem.

Chapter 4

A New Class of Contiguous Alternatives

In Chapters 2 and 3 we have showed that under Assumptions 1.1, 1.3 and the Kong-Slud Assumption I (Assumption 1.4), the logrank and W-stratified logrank statistics are all centered and have the asymptotic standard normal distribution under the null hypothesis of no treatment effect. However, when Assumption 1.4 does not hold, these two test statistics are biased.

In this chapter, we introduce a new class of alternatives within which Assumption 1.4 is violated and the treatment effect is small, and we prove that they are a sequence of contiguous alternatives with respect to a fixed "null" probability under which Assumption 1.4 holds and there is no treatment effect. Here the violation of Assumption 1.4 is represented by an interaction term in the log conditional survival function for censoring such that the term has specified rate behavior with respect to n. It is not difficult to show that if Assumption 1.5 does not hold, that is, if there is an interaction term of treatment and covariate inside the survival function for censoring, the Kong-Slud Assumption fails to hold.

Given $Z = z$ and $V = v$, denote the conditional hazard intensity functions of T and C as $\lambda(t, \theta z, v)$, and $\lambda_C(t, \psi, z, v)$, their cumulative conditional hazard functions as $\Lambda(t, \theta z, v)$ and $\Lambda_C(t, \psi, z, v)$ and their conditional density functions as $f_T(t, \theta z, v)$ and $f_C(t, \psi, z, v)$, respectively.

Theorem 4.1 *Assume*

1. *The survival function for the censoring time r.v. C satisfies*

$$-\log[S_C(t|z,v)] = a(t,z) + b(t,v) + \psi \cdot c(t, zh(v))$$

for some positive functions $a(\cdot), b(\cdot)$ and $c(\cdot)$, where $\psi \in \mathbb{R}$ is constant in t, z, v.

2. *For $i = 1, \ldots, n$ and $r = 0, 1, 2$, the following terms*

$$\frac{\partial^r}{\partial \theta^r} \log f_T(t, \theta Z_i, V_i), \quad \frac{\partial^r}{\partial \psi^r} \log f_C(t, \psi, Z_i, V_i)$$

are all continuous and uniformly integrable, with respect to dt, over θ and ψ in a sufficiently small neighborhood of 0, and under the null hypothesis $H_0 : \theta = \psi = 0$;

3. *For $i = 1, \ldots, n$, the density functions satisfy*

$$\begin{aligned} \frac{d}{d\theta} E_\theta \left\{ \frac{\partial}{\partial \theta} \log f_T(T_i, \theta z, v) \right\} &= \int \frac{\partial}{\partial \theta} \left\{ \left[\frac{\partial}{\partial \theta} \log f_T(t, \theta z, v) \right] f_T(t, \theta z, v) \right\} dt; \\ \frac{d}{d\psi} E_\psi \left\{ \frac{\partial}{\partial \psi} \log f_C(C_i, z, v, \psi) \right\} &= \int \frac{\partial}{\partial \psi} \left\{ \left[\frac{\partial}{\partial \psi} \log f_C(t, z, v, \psi) \right] f_C(t, z, v, \psi) \right\} dt \end{aligned} \tag{4.1}$$

Then the hypotheses $H_A:\ \theta_n = b/\sqrt{n}; \psi_n = c/\sqrt{n}$ and $H_0 : \theta = \psi = 0$ are mutually contiguous.

Proof.

The tool we will use for this proof is part of Le Cam's first lemma (Van der Vaart, 1998, Lemma 6.4, (i),(ii),(iii)). Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. In order to show the sequence Q_n is contiguous with respect to the sequence P_n, we just need to show that $dP_n/dQ_n \xrightarrow{\mathcal{D}(Q_n)} U$ such that $P(U > 0) = 1$, where $\mathcal{D}(Q_n)$ means "in distribution under Q_n". If we can further show that $E(U) = 1$, then P_n is contiguous with respect to Q_n.

Here we can let the sequence P_n be the probability measure of the alternative hypothesis H_A and the sequence Q_n be the probability measure of the null hypothesis H_0. Then the sequence dP_n/dQ_n will be the likelihood ratio of the alternative and null hypothesis. This theorem states that P_n and Q_n are mutually contiguous.

The joint density function of (T, C, Z, V) can be written as

$$f_{T,C|Z,V}(t, c|z, v) \cdot f_{Z,V}(z, v) = f_T(t, z\theta, v) f_C(t, z, v, \psi) dF_z dF_v$$

Since the difference between P_n and Q_n relates only to θ and ψ, the factors not dependent on these two parameters in the likelihood ratio will cancel. The likelihood ratio is

$$\frac{dP_n}{dQ_n} = \frac{dP_n^T}{dQ_n^T} \cdot \frac{dP_n^C}{dQ_n^C}$$

where

$$\frac{dP_n^T}{dQ_n^T} = \frac{\Pi_{i=1}^n f_T(T_i, \theta_n Z_i, V_i)}{\Pi_{i=1}^n f_T(T_i, 0, V_i)}$$

and

$$\frac{dP_n^C}{dQ_n^C} = \frac{\Pi_{i=1}^n f_C(C_i, Z_i, V_i, \psi_n)}{\Pi_{i=1}^n f_C(C_i, Z_i, V_i, 0)}$$

Taking logarithms in these expressions, we obtain the log-likelihood ratios:

$$\log \frac{dP_n^T}{dQ_n^T} = \sum_{i=1}^n \{\log f_T(T_i, \theta_n Z_i, V_i) - \log f_T(T_i, 0, V_i)\}$$

$$\log \frac{dP_n^C}{dQ_n^C} = \sum_{i=1}^n \{\log f_C(C_i, Z_i, V_i, \psi_n) - \log f_C(C_i, Z_i, V_i, 0)\}$$

Since $\theta_n = b/\sqrt{n}$ and $\psi_n = a/\sqrt{n}$, for each i, we Taylor expand $\log f_T(T_i, \theta_n Z_i, V_i)$ and $\log f_C(C_i, Z_i, V_i, \psi_n)$ about $\theta = 0$ and $\psi = 0$, respectively, leading to

$$\begin{aligned} & \log f_T(T_i, \frac{b}{\sqrt{n}} Z_i, V_i) \\ = \quad & \log f_T(T_i, 0, V_i) + \frac{b}{\sqrt{n}} \cdot logL_{Ti}^{(1)}(0) + \frac{1}{n}\frac{b^2}{2} \cdot logL_{Ti}^{(2)}(0) + \frac{1}{n} R_{2i}^T \end{aligned}$$

and

$$\begin{aligned} & \log f_C(C_i, \frac{a}{\sqrt{n}}, Z_i, V_i) \\ = \quad & \log f_C(C_i, 0, Z_i, V_i) + \frac{a}{\sqrt{n}} \cdot logL_{Ci}^{(1)}(0) + \frac{1}{n}\frac{a^2}{2} \cdot logL_{Ci}^{(2)}(0) + \frac{1}{n} R_{2i}^C. \end{aligned}$$

where

$$logL_{Ti}^{(r)}(x) = \frac{\partial^r}{\partial \theta^r} \log f_T(T_i, \theta Z_i, V_i)\Big|_{\theta = x},$$

$$logL_{Ci}^{(r)}(x) = \frac{\partial^r}{\partial \theta^r} \log f_C(C_i, Z_i, V_i, \psi)\Big|_{\psi = x}, \; r = 0, 1, 2, \tag{4.2}$$

$$R_{2i}^T = \frac{b^2}{2} \cdot \{logL_{Ti}^{(2)}(\theta_i^*) - logL_T^{(2)}(0)\},$$

$$R_{2i}^C = \frac{a^2}{2} \cdot \{logL_{Ci}^{(2)}(\psi_i^*) - logL_C^{(2)}(0)\};$$

$$0 \le \theta_i^* = g_1(T_i, Z_i, V_i) \le \frac{b}{\sqrt{n}},$$

$$0 \le \psi_i^* = g_2(C_i, Z_i, V_i) \le \frac{a}{\sqrt{n}}.$$

Thus the log-likelihood ratio becomes

$$\begin{aligned} \log \frac{dP_n}{dQ_n} = & \frac{1}{\sqrt{n}} \sum_{i=1}^n \{b \cdot logL_{Ti}^{(1)}(0) + a \cdot logL_{Ci}^{(1)}(0)\} \\ & + \frac{1}{n} \cdot \frac{1}{2} \sum_{i=1}^n \{b^2 \cdot logL_{Ti}^{(2)}(0) + a^2 \cdot logL_{Ci}^{(2)}(0)\} + \frac{1}{n} \sum_{i=1}^n \{R_{2i}^T + R_{2i}^C\} \\ \equiv & \; A_{1n} \; + \; A_{2n} \; + \; A_{3n}. \end{aligned} \tag{4.3}$$

From the definition (4.2) we know that for each $i = 1, \ldots, n$, $logL_{Ti}^{(1)}(0)$ are iid distributed under H_0 with mean

$$\mu_T = E_0\{\frac{\partial}{\partial\theta}\log f_T(T_i, \theta Z_i, V_i)\Big|_{\theta=0}\} = 0$$

and variance

$$\sigma_T^2 = E_0\{[logL_{T1}^{(1)}(0)]^2\};$$

Similarly, $logL_{Ci}^{(1)}(0)$, $i = 1, \ldots, n$ are iid with mean $\mu_T = 0$ and variance

$$\sigma_C^2 = E_0\{[logL_{C1}^{(1)}(0)]^2\}$$

Furthermore, since T and C are conditionally independent given Z and V,

$$E_0\{logL_{T1}^{(1)}(0) \cdot logL_{C1}^{(1)}(0)\} = E_0\{E_0[logL_{T1}^{(1)}(0)|Z, V] \cdot E_0[logL_{C1}^{(1)}(0)]\} = 0$$

Thus

$$E_0\{[logL_{T1}^{(1)}(0) + logL_{C1}^{(1)}(0)]^2\} = \sigma_T^2 \ + \ \sigma_C^2 \tag{4.4}$$

From Lemma 7.3.11 in Casella and Berger (2001) and the third assumption in this theorem,

$$\begin{aligned}
\sigma_T^2 = E_0\{[logL_{T1}^{(1)}(0)]^2\} \quad &= \quad - \ E_0\{logL_{T1}^{(2)}(0)\}, \\
\sigma_C^2 = E_0\{[logL_{C1}^{(1)}(0)]^2\} \quad &= \quad - \ E_0\{logL_{C1}^{(2)}(0)\}.
\end{aligned} \tag{4.5}$$

Then by the second assumption of this theorem,

$$\sigma_T^2 < \infty \text{ and } \sigma_C^2 < \infty.$$

By the Central Limit Theorem, in (4.3)

$$A_{1n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{b \cdot logL_{Ti}^{(1)}(0) + a \cdot logL_{Ci}^{(1)}(0)\} \overset{\mathcal{D}(Q_n)\cdot H_0}{\longrightarrow} N(0, b^2\sigma_T^2 + a^2\sigma_C^2). \tag{4.6}$$

By the Law of Large Numbers and (4.5),

$$A_{2n} = \frac{1}{n} \cdot \frac{1}{2} \sum_{i=1}^{n} \{b^2 \cdot logL_{Ti}^{(2)}(0) + a^2 \cdot logL_{Ci}^{(2)}(0)\} \overset{\mathcal{D}(Q_n)\cdot H_0}{\longrightarrow} -\frac{1}{2}(b^2\sigma_T^2 + a^2\sigma_C^2) \quad (4.7)$$

By the continuity of $logL_{Ti}^{(2)}(\cdot)$ and $logL_{Ci}^{(2)}(\cdot)$ around the small neighborhood of 0,

$$logL_{Ti}^{(2)}(\theta_i^*) - logL_{Ti}^{(2)}(0) = o_p(1), \text{ in probability and } L^2;$$

$$logL_{Ci}^{(2)}(\psi_i^*) - logL_{Ci}^{(2)}(0) = o_p(1) \text{ in probability and } L^2.$$

Thus

$$A_{3n} = \sum_{i=1}^{n} \{R_{2i}^T + R_{2i}^C\} = o_p(1) \text{ and} in\ L^2. \quad (4.8)$$

Finally, from (4.6) - (4.8) and Slutsky's Lemma, the asymptotic distribution of the log-likelihood ratio under the null hypothesis H_0 is

$$\log \frac{dP_n}{dQ_n} \overset{\mathcal{D}(Q_n)\cdot H_0}{\longrightarrow} W \overset{d}{\sim} N(\mu, \sigma^2) \quad (4.9)$$

where

$$\mu = -\frac{1}{2}\sigma^2 \text{ and } \sigma^2 = b^2\sigma_T^2 + a^2\sigma_C^2$$

Thus the likelihood ratio converges in distribution to a lognormal distributed r.v. $U = e^W$:

$$\frac{dP_n}{dQ_n} \overset{\mathcal{D}(Q_n)\cdot H_0}{\longrightarrow} U \overset{d}{\sim} e^{-N(\mu,\sigma^2)}.$$

Therefore, since $\mu = -\frac{1}{2}\sigma^2$,

$$P(U > 0) = 1 \text{ and } E(U) = 1.$$

By Le Cam's first lemma(Van der Vaart, 1998, Lemma 6.4,(i)(ii)), the sequence Q_n is contiguous with respect to the sequence P_n, and by part (iii) of the same Lemma with P_n and Q_n reversed, the sequence P_n is also contiguous with respect to the sequence Q_n.

Therefore the sequences P_n and Q_n are mutually contiguous. $\square$

From Theorem 4.1 and its proof, the following corollary is immediate:

Corollary 4.1 *Under the same assumptions in Theorem 4.1, the null hypothesis* $H_0 : \theta = 0.$ *and the alternative hypothesis* $H_A : \theta_n = \frac{b}{\sqrt{n}}$ *are mutually contiguous.*

Corollary 4.2 *Under the same assumptions in Theorem 4.1, the null hypothesis* $H_0^* : \psi = 0.$ *and the alternative hypothesis* $H_A^* : \psi_n = \frac{c}{\sqrt{n}}$ *are mutually contiguous.*

Remark

When the violation of the Kong-Slud Assumption is small, Theorem 4.1 enables us, under certain regularity conditions, to calculate the asymptotic distribution of the logrank statistic $n^{-\frac{1}{2}}\hat{U}_L$ or the stratified logrank statistic $n^{-\frac{1}{2}}\hat{U}_W$ under the contiguous alternatives to $\theta = 0, \psi = 0$. Simulation studies on the application of this theorem will be provided in Chapter 6.

Chapter 5

A Bias-corrected Logrank Test

In this chapter we study and extend a bias correction method proposed in Dirienzon and Lagakos (2001a) and apply it to the logrank statistic to get a "bias corrected" logrank test. In Section 5.1 we describe how this bias-correction method works and what is its limitation. In Section 5.2, we prove several useful lemmas first and then prove the theorem that assures the asymptotic normal distribution of the bias-corrected test with unknown and Kaplan-Meier estimated conditional distribution function of the censoring. A correct consistent variance estimator is also found within our asymptotic framework.

5.1 The $\varphi(\cdot)$ Function and the Weighted at Risk Indicator

The bias-correction method proposed in DiRienzo and Lagakos (2001a) uses information obtained from the censoring distribution to weight each subject at risk. The binary at risk indicator function $Y(t)$ is replaced by a continuous variable taking values in a unit interval. In their proposal, the weighted at risk indicator function is

$$Y_i^*(t) \equiv \varphi(t, Z_i, V_i) \cdot Y_i(t)$$

with

$$\varphi(t, Z_i, V_i) \equiv \frac{S_C(t, 1, V_i) \wedge S_C(t, 0, V_i)}{S_C(t, Z_i, V_i)}.$$

Then the conditional expectation of $Y^*(t)$ given Z and V is independent of Z under H_0 because by Assumption 1.1 and 1.3, for each i,

$$\begin{aligned} E\{Y_i^*(t)|Z_i, V_i\} &= E\{\varphi(t, Z_i, V_i)Y_i(t)|Z_i, V_i\} \\ &= \frac{S_C(t,0,V_i) \wedge S_C(t,0,V_i)}{S_C(t,Z_i,V_i)} \cdot S_C(t, Z_i, V_i)S(t,0,V_i) \\ &= \{S_C(t,0,V_i) \wedge S_C(t,0,V_i)\} \cdot S(t,0,V_i) \end{aligned}$$

does not depend upon Z_i.

The essential part of the weighting function is placing $S_C(t, Z_i, V_i)$, the conditional survival function of the censoring time, in the denominator of $\varphi(t, Z_i, V_i)$. The numerator in the definition of $\varphi(t, Z_i, V_i)$ must be a function that does not depend upon Z_i. Thus we suggest to define the weighting function in a general way:

$$\varphi(t, Z_i, V_i) = \frac{g(t, V_i)}{S_C(t, Z_i, V_i)}.$$

Here g would be chosen so that $g(t, V_i) \leq S_C(t, Z_i, V_i)$ with probability one. To comply with the technical requirements of all theorems regarding the choice of $\varphi(\cdot)$, we further restrict $g(\cdot)$ by assuming the following:

Assumption 5.1 *For any z and v,*

(i) $g(t, v)$ is a function of $S_C(t, 1, v)$ and $S_C(t, 0, v)$ and $\hat{g}$ is the same function evaluated at the Kaplan-Meier estimators $\hat{S}_C(t, z, v)$;

(ii) $g(t, v) \leq S_C(t, z, v)$;

(iii) $\inf_{0 \leq t \leq K} E\{g(t, V)\} > 0$ *for fixed $K > 0$ so that* $E\{S_C(t, Z, V)\} > 0$.

(iv)Let $\hat{g}$ and $\hat{S}_C$ be the Kaplan-Meier estimators of g and S_C, then

$$|\hat{g}(t, v) - g(t, v)| \leq c_1 \left|\hat{S}_C(t, 1, v) - S_C(t, 1, v)\right| + c_0 \left|\hat{S}_C(t, 0, v) - S_C(t, 0, v)\right|$$

for some constants c_1 and c_0.

It is clear that minimum, $g(t,v) = S_C(t,1,v) \wedge S_C(t,0,v)$ and the product, $g(t,v) = S_C(t,1,v)S_C(t,0,v)$ are two direct examples that satisfy the assumption above.

A bias-corrected logrank statistic $n^{-\frac{1}{2}}\hat{U}_\varphi$ is obtained by replacing each $Y_i(t)$ in the statistic $n^{-\frac{1}{2}}\hat{U}_L$ by $Y_i^*(t)$. Then $n^{-\frac{1}{2}}\hat{U}_\varphi$ can be written as

$$n^{-\frac{1}{2}}\hat{U}_\varphi \equiv n^{-\frac{1}{2}} \sum_{i=1}^{n} \int \left[Z_i - \frac{\sum_{j=1}^{n} \varphi(t, Z_j, V_j) Y_j(t) Z_j}{\sum_{j=1}^{n} \varphi(t, Z_j, V_j) Y_j(t)} \right] \varphi(t, Z_i, V_i) dN_i(t). \tag{5.1}$$

As shown in DiRienzo and Lagakos (2001a), $n^{-\frac{1}{2}}\hat{U}_\varphi$ is asymptotically a sum of iid terms with mean 0 and thus asymptotically bias free under H_0. Along with the consistent sample variance estimator constructed based on this asymptotic sum, a bias corrected statistic that is asymptotically standard normally distributed under the null hypothesis can thus be found.

A major limitation of this method is that the weighting function $\varphi(\cdot)$ depends on the distribution functions of the censoring time. The authors who proposed this method found from simulations that their bias corrected test with estimated weighting function is very similar to the one with known weighting function $\varphi(\cdot)$.

We are interested in studying the asymptotic properties of the bias corrected logrank test when the censoring distribution is unknown and an estimated function is then substituted. Regarding different types of estimators of the conditional distribution of the censoring time, we find that using a stratumwise nonparametric Kaplan-Meier estimate or a proper parametric estimator will result in an asymptotically valid bias corrected logrank test. A simulation study also suggests that a valid bias-corrected test can be obtained with φ estimated under a semi-parametric model, Aalen's additive model.

5.2 Bias Correction with Kaplan-Meier Estimation

As one of the major contributions of this thesis, we show in this section that when the φ function is estimated by $\hat{\varphi}$ using the stratified Kaplan-Meier method, the corrected test based on $\hat{\varphi}$ is asymptotically normally distributed with mean zero under the null hypothesis of no randomized treatment effect. We also derive a consistent variance estimator for the test statistic.

Define

$$n^{-\frac{1}{2}}\tilde{U}_\varphi \equiv n^{-\frac{1}{2}}\sum_{i=1}^{n}\int [Z_i-\pi]\varphi(t,Z_i,V_i)\Big\{dN_i(t)-Y_i(t)\frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]}dt\Big\}.$$

Recall that under H_0 the corrected test statistic $n^{-\frac{1}{2}}\hat{U}_\varphi$ with known φ function is asymptotically an iid sum with mean 0. This result has been proved by DiRienzo and Lagakos (2001) and is an important reference for our work here. We provide an alternative proof for the similar result in the following lemma.

Lemma 5.1 *When Assumptions 1.1 and 1.3 hold,*

$$n^{-\frac{1}{2}}\hat{U}_\varphi - n^{-\frac{1}{2}}\tilde{U}_\varphi \xrightarrow{p} 0 \ as\ n \to \infty$$

under H_0.

Proof.

For $K > 0$ fixed in such a way that $E\{S_C(K,Z,V)\} > 0$ and $E\{S(K,0,V)\} > 0$, define $n^{-\frac{1}{2}}\hat{U}_\varphi^{[0,K]} = n^{-\frac{1}{2}}\hat{U}_\varphi - n^{-\frac{1}{2}}\hat{U}_\varphi^{(K,\infty)}$, where $n^{-\frac{1}{2}}\hat{U}_\varphi^{(K,\infty)}$ is defined as in Lemma

A.9 by restricting the integral in the definition of $n^{-\frac{1}{2}}\hat{U}_\varphi$ to (K,∞). Adding and subtracting $\pi = E(Z_i)$ within $\hat{U}_\varphi^{[0,K]}$ we obtain

$$n^{-\frac{1}{2}}\hat{U}_\varphi^{[0,K]} = -S_1 \; + \; S_2,$$

where

$$\begin{aligned}
S_1 \;&\equiv\; n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^K \varphi(t,Z_i,V_i)\Big(\frac{n^{-1}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big)dN_i(t)\\
&=\; n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^K \varphi(t,Z_i,V_i)\Big(\frac{n^{-1}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big)dM_i(t)\\
&\quad + \int_0^K \Big(\frac{n^{-1}\sum_i \varphi(s,Z_i,V_i)Y_i(t)\lambda(t,0,V_i)}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big)\\
&\qquad\qquad \cdot \Big(n^{-\frac{1}{2}}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t)\Big)dt\\
&\equiv\; S_{1a}+S_{1b} \qquad\qquad (5.2)
\end{aligned}$$

and

$$S_2 \;\equiv\; n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^K \varphi(t,Z_i,V_i)(Z_i-\pi)dN_i(t). \qquad (5.3)$$

By the uniform Law of Large Numbers, which also shows that the denominator is uniformly bounded away from 0, and Lemma (A.4),

$$\sup_{0\le t\le K}\Big|\varphi(t,Z_i,V_i)\frac{n^{-1}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t)}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big| \;\xrightarrow{p} 0$$

Then from Lemma A.13,

$$S_{1a} \xrightarrow{p} 0, \text{ as } n\to\infty.$$

Also by the uniform Law of Large Numbers and Lemma A.4,

$$\sup_{0\le t\le K}\Big|\frac{n^{-1}\sum_i \varphi(t,Z_i,V_i)Y_i(t)\lambda(t,0,V_i)}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)} - \frac{E\{g(t,V)e^{-\Lambda(t,0,V)}\lambda(t,0,V)\}}{E\{g(t,V)e^{-\Lambda(t,0,V)}\}}\Big| \xrightarrow{p} 0.$$

Note that by independence of Z_i and V_i,

$$\begin{aligned} E\{(Z_i - \pi)\varphi(t, Z_i, V_i)Y_i(t)\} &= E\{(Z_i - \pi)g(t, V_i)e^{-\Lambda(t,0,V_i)}\} \\ &= E\{Z_i - \pi\} \cdot E\{g(t, V_i)e^{-\Lambda(t,0,V_i)}\} \\ &= 0. \end{aligned}$$

Then because

$$\frac{E\{g(t, V)e^{-\Lambda(t,0,V)}\lambda(t, 0, V)\}}{E\{g(t, V)e^{-\Lambda(t,0,V)}\}} = \frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]}, \tag{5.4}$$

by the Donsker theorem

$$n^{-\frac{1}{2}} \sum_i (Z_i - \pi)\varphi(t, Z_i, V_i)Y_i(t) = O_p(1)$$

uniformly over t in the compact set $[0, K]$. Therefore,

$$S_{1b} = n^{-\frac{1}{2}} \int_0^K \sum_i (Z_i - \pi)\varphi(t, Z_i, V_i)Y_i(t) \frac{E\{g(t, V)e^{-\Lambda(t,0,V)}\lambda(t, 0, V)\}}{E\{g(t, V)e^{-\Lambda(t,0,V)}\}} dt + o_p(1).$$

Thus

$$\begin{aligned} n^{-\frac{1}{2}} \hat{U}_\varphi^{[0,K]} &= S_{1a} + S_{1b} + S_2 \\ &= n^{-\frac{1}{2}} \tilde{U}_\varphi^{[0,K]} + o_p(1). \end{aligned} \tag{5.5}$$

Therefore

$$n^{-\frac{1}{2}} \hat{U}_\varphi^{[0,K]} - n^{-\frac{1}{2}} \tilde{U}_\varphi^{[0,K]} \xrightarrow{p} 0$$

as $n \to \infty$. From Lemma A.9 we also have

$$n^{-\frac{1}{2}} \hat{U}_\varphi^{(K,\infty)} \xrightarrow{p,L_1} 0 \text{ and } n^{-\frac{1}{2}} \tilde{U}_\varphi^{(K,\infty)} \xrightarrow{p,L_1} 0 \text{ as } K \uparrow \infty.$$

Thus by Lemma A.5,

$$n^{-\frac{1}{2}} \hat{U}_\varphi - n^{-\frac{1}{2}} \tilde{U}_\varphi \xrightarrow{p} 0$$

as $n \to \infty$. $\square$

Next we study the analogous asymptotic representation of the corrected test with stratified Kaplan-Meier estimated φ function. Define

$$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} \equiv n^{-\frac{1}{2}} \sum_{i=1}^{n} \int [Z_i - \frac{\sum_{i=1}^{n} \hat{\varphi}(t, Z_i, V_i) Y_i(t) Z_i}{\sum_{i=1}^{n} \hat{\varphi}(t, Z_i, V_i) Y_i(t)}] \hat{\varphi}(t, Z_i, V_i) dN_i(t)$$

$$n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}} \equiv n^{-\frac{1}{2}} \sum_{i=1}^{n} \int [Z_i - \pi] \hat{\varphi}(t, Z_i, V_i) \Big\{ dN_i(t) - Y_i(t) \frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]} dt \Big\}.$$

Lemma 5.2 *Let Assumptions 1.1, 1.3 hold and V_i be finite-valued. Let $\hat{\varphi}(t, Z_i, V_i)$ be the stratified Kaplan-Meier estimator for $\varphi(t, Z_i, V_i)$. Then under H_0,*

$$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}} \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Proof.

Similar to the proof of Lemma 5.1, we first consider the convergence of the two statistics if the integrals are restricted to $[0, K]$. Define $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{[0,K]} = n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{(K,\infty)}$. Adding and subtracting $\pi = E(Z_i)$ within $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{[0,K]}$ we obtain

$$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{[0,K]} = S_1^* + S_2^*.$$

where

$$\begin{aligned}
S_1^* &\equiv n^{-\frac{1}{2}} \sum_{i}^{n} \int_0^K \hat{\varphi}(t, Z_i, V_i) \Big(\frac{n^{-1} \sum_j (Z_j - \pi) \hat{\varphi}(t, Z_j, V_j) Y_j(t))}{n^{-1} \sum_j \hat{\varphi}(t, Z_j, V_j) Y_j(t)} \Big) dN_i(t) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^K \hat{\varphi}(t, Z_i, V_i) \Big(\frac{n^{-1} \sum_j (Z_j - \pi) \hat{\varphi}(t, Z_j, V_j) Y_j(t))}{n^{-1} \sum_j \hat{\varphi}(t, Z_j, V_j) Y_j(t)} \Big) dM_i(t) \\
&\quad + \int_0^K \Big(\frac{n^{-1} \sum_i \hat{\varphi}(s, Z_i, V_i) Y_i(t) \lambda(t, 0, V_i)}{n^{-1} \sum_j \hat{\varphi}(t, Z_j, V_j) Y_j(t)} \Big) \\
&\qquad \cdot \Big(n^{-\frac{1}{2}} \sum_j (Z_j - \pi) \hat{\varphi}(t, Z_j, V_j) Y_j(t) \Big) dt \\
&\equiv S_{1a}^* + S_{1b}^* \qquad (5.6)
\end{aligned}$$

and

$$S_2^* \;\equiv\; n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^K \hat{\varphi}(t, Z_i, V_i)(Z_i - \pi) dN_i(t). \tag{5.7}$$

From Theorem 6.3.1 of Fleming and Harrington (1990) we can derive that the stratified Kaplan-Meyer estimator $\hat{\varphi}(t, z, v)$ satisfies

$$\sup_{0 \le t \le K} \sup_{z,v} \sqrt{n} |\hat{\varphi}(t, z, v) - \varphi(t, z, v)| = O_p(1), \tag{5.8}$$

which then implies

$$\max_i \sup_{0 \le t \le K} \sqrt{n} |\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)| \;=\; O_p(1).$$

This immediately shows that

$$\sup_{0 \le t \le K} n^{-\frac{1}{2}} \big| \sum_i (Z_i - \pi)[\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)] Y_i(t) \big| \;=\; O_p(1) \tag{5.9}$$

$$\sup_{0 \le t \le K} n^{-1} \big| \sum_i (Z_i - \pi)[\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)] Y_i(t) \big| \;=\; o_p(1) \tag{5.10}$$

and

$$\sup_{0 \le t \le K} n^{-1} \big| \sum_i [\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)] Y_i(t) \big| \;=\; o_p(1) \tag{5.11}$$

$$\sup_{0 \le t \le K} n^{-1} \big| \sum_i \lambda(t, 0, V_i)[\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)] Y_i(t) \big| \;=\; o_p(1). \tag{5.12}$$

Therefore by (5.10) and (5.11),

$$\sup_{0 \le t \le K} \big| n^{-1} \sum_i (Z_i - \pi) \hat{\varphi}(t, Z_i, V_i) Y_i(t) \big| \;\xrightarrow{p}\; 0$$

$$\sup_{0 \le t \le K} \big| n^{-1} \sum_i \hat{\varphi}(t, Z_i, V_i) Y_i(t) - E\{Y(t) \varphi(t, Z, V)\} \big| \;\xrightarrow{p}\; 0.$$

Then we use Lemma A.4 to get

$$\frac{n^{-1}\sum_i (Z_i - \pi)\hat{\varphi}(s, Z_i, V_i)Y_i(t)}{n^{-1}\sum_j \hat{\varphi}(t, Z_j, V_j)Y_j(t)} \xrightarrow{p} 0$$

uniformly on the compact set $[0, K]$. Finally we apply Lemma A.13 and get

$$S^*_{1a} \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Similarly, from (5.11), (5.12), the uniform law of large numbers and Lemma A.4 we have

$$\frac{n^{-1}\sum_i \hat{\varphi}(s, Z_i, V_i)Y_i(t)\lambda(t, 0, V_i)}{n^{-1}\sum_j \hat{\varphi}(t, Z_j, V_j)Y_j(t)} - \frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]} \xrightarrow{p} 0$$

uniformly on the compact set $[0, K]$, and from (5.9),

$$\sup_{0\le t\le K} n^{-\frac{1}{2}}\Big|\sum_i (Z_i - \pi)\hat{\varphi}(t, Z_i, V_i)Y_i(t)\Big| = O_p(1).$$

Therefore,

$$S^*_{1b} = n^{-\frac{1}{2}}\int_0^K \sum_i (Z_i - \pi)\hat{\varphi}(t, Z_i, V_i)Y_i(t)\frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]}dt + o_p(1).$$

Thus

$$\begin{aligned} n^{-\frac{1}{2}}\hat{U}^{[0,K]}_{\hat{\varphi}} &= S^*_{1a} + S^*_{1b} + S^*_2 \\ &= n^{-\frac{1}{2}}\tilde{U}^{[0,K]}_{\hat{\varphi}} + o_p(1), \end{aligned}$$

Therefore

$$n^{-\frac{1}{2}}\hat{U}^{[0,K]}_{\hat{\varphi}} - n^{-\frac{1}{2}}\tilde{U}^{[0,K]}_{\hat{\varphi}} \xrightarrow{p} 0$$

as $n \to \infty$. From Lemma A.11 we also have

$$n^{-\frac{1}{2}}\hat{U}^{(K,\infty)}_{\hat{\varphi}} \xrightarrow{p,L_1} 0 \text{ and } n^{-\frac{1}{2}}\tilde{U}^{(K,\infty)}_{\hat{\varphi}} \xrightarrow{p,L_1} 0 \text{ as } K \uparrow \infty.$$

Thus by Lemma A.5,

$$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}} \xrightarrow{p} 0$$

as $n \to \infty$. $\square$

With the two asymptotic terms $n^{-\frac{1}{2}}\tilde{U}_{\varphi}$ and $n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}$, we study the asymptotic property of $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_{\varphi}$ in Lemma 5.3 and find that it is asymptotically a sum of iid terms with mean zero.

Define

$$\bar{\lambda}(t) \equiv \frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]}. \tag{5.13}$$

$$N_i^C(t) = I_{[C_i \le T]} \cdot I_{[C_i \le t]}$$

$$M_i^C(t) = N_i^C(t) - \int_0^t Y_i(t) d\Lambda_C(t, Z_i, V_i)$$

Lemma 5.3 *When Assumptions 1.1 and 1.3 hold and $\hat{\varphi}(t, Z_i, V_i)$ is a stratified Kaplan-Meier estimator of $\varphi(t, Z_i, V_i)$, then*

$$n^{-\frac{1}{2}}\hat{U}_d - n^{-\frac{1}{2}}U_d \xrightarrow{p} 0 \text{ as } n \to \infty$$

with

$$n^{-\frac{1}{2}}\hat{U}_d \equiv n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_{\varphi}$$

and

$$\begin{aligned} n^{-\frac{1}{2}}U_d \equiv & -n^{-\frac{1}{2}}\int\sum_{i=1}^n (Z_i - \pi) g(t, V_i) e^{-\Lambda(t,0,V_i)}[\bar{\lambda}(t) - \lambda(t,0,V_i)] \\ & \cdot\{\int_0^t \frac{dM_i^C(s)}{S_C(s, Z_i, V_i)e^{-\Lambda(s,0,V_i)}}\}dt. \end{aligned}$$

Proof.

As in the proof of the previous two lemmas, we will first prove the convergence of $n^{-\frac{1}{2}}\hat{U}_d^{[0,K]} - n^{-\frac{1}{2}}U_d^{[0,K]}$ to 0 for any $K > 0$. Define

$$n^{-\frac{1}{2}}\tilde{U}_d \equiv n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\tilde{U}_{\varphi}.$$

Since V_i is finite-valued, $n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{[0,K]}$ can be written as

$$n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{[0,K]} \;\equiv\; R_1 \;-\; R_2,$$

where

$$\begin{aligned}
R_1 &= \sum_{z,v}\int_0^K \hat{\varphi}(t,z,v)(z-\pi)n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}dM_i(t), \\
R_2 &= \sum_{z,v}\int_0^K \varphi(t,z,v)(z-\pi)[\bar{\lambda}(t)-\lambda(t,0,v)]n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}Y_i(t)dt \\
&= \sum_{z,v}\int_0^K \hat{g}(t,v)\frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)}(z-\pi)[\bar{\lambda}(t)-\lambda(t,0,v)] \\
&\quad \cdot n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}\{\frac{Y_i(t)}{S_C(t,z,v)} - e^{-\Lambda(t,0,v)} + e^{-\Lambda(t,0,v)}\}dt \\
&\equiv \; R_{2a} \;+\; R_{2b}
\end{aligned}$$

with

$$R_{2a} \;=\; \sum_{z,v}\int_0^K \hat{g}(t,v)e^{-\Lambda(t,0,v)}\frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)}(z-\pi)[\bar{\lambda}(t)-\lambda(t,0,v)]n^{-\frac{1}{2}}\{\sum_{i=1}^n I_{[Z_i=z,V_i=v]}\}dt$$

and

$$\begin{aligned}
R_{2b} &= \sum_{z,v}\int_0^K \hat{\varphi}(t,z,v)(z-\pi)[\bar{\lambda}(t)-\lambda(t,0,v)] \\
&\quad \cdot n^{-\frac{1}{2}}\{\sum_{i=1}^n I_{[Z_i=z,V_i=v]}\{Y_i(t) - S_C(t,z,v)e^{-\Lambda(t,0,v)}\}\}dt.
\end{aligned}$$

By (5.8) and Lemma A.13,

$$\begin{aligned} R_1 &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^K (Z_i - \pi)\varphi(t, Z_i, V_i) dM_i(t) \; + o_p(1) \\ &\equiv R_1^* + \; o_p(1). \end{aligned} \tag{5.14}$$

Next consider R_{2a}. By Theorem 6.3.1, for each fixed v and z,

$$\sup_{0\le t\le K} \left| \sqrt{n}\{\frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)} - 1\} \right| \;=\; O_p(1).$$

By Assumption 5.1 we also have

$$\max_v \sup_{0\le t\le K} \sqrt{n}\big|\hat{g}(t,v) - g(t,v)\big| \;=\; O_p(1),$$

which implies

$$\max_v \sup_{0\le t\le K} \left| [\hat{g}(t,v) - g(t,v)] \left\{ \frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)} - 1 \right\} n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{n} I_{[Z_i=z, V_i=v]} \right\} \right| = o_p(1)$$

and

$$\max_v \sup_{0\le t\le K} \left| [\hat{g}(t,v) - g(t,v)] n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{n} I_{[Z_i=z, V_i=v]} \right\} \right| = o_p(1).$$

Let $p_{z,v} \;\equiv\; P\{Z_i = z, V_i = v\}$. Then by the Law of Large Numbers

$$n^{-1} \sum_{i=1}^{n} I_{[Z_i=z, V_i=v]} \;\overset{p,L_1}{\longrightarrow}\; p_{z,v}.$$

Finally R_{2a} becomes

$$R_{2a} = R_{2a}^{(1)} \;+\; R_{2a}^{(2)} \;+\; o_p(1) \tag{5.15}$$

with

$$R_{2a}^{(1)} = \sum_{z,v} \int_0^K g(t,v) e^{-\Lambda(t,0,v)} (z - \pi)[\bar{\lambda}(t) - \lambda(t,0,v)] \sqrt{n} \left\{ \frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)} - 1 \right\} p_{z,v} dt$$

and

$$R_{2a}^{(2)} = \sum_{z,v} \int_0^K g(t,v) e^{-\Lambda(t,0,v)} (z-\pi)[\bar{\lambda}(t) - \lambda(t,0,v)] n^{-\frac{1}{2}} \left\{ \sum_{i=1}^n I_{[Z_i=z,V_i=v]} \right\} dt.$$

Next consider R_{2b}. By Lemma A.3,

$$\max_v \sup_{0 \le t \le K} \left| n^{-1} \sum_{i=1}^n I_{[Z_i=z,V_i=v]} \{Y_i(t) - S_c(t,z,v) e^{-\Lambda(t,0,v)}\} \right| = o_p(1).$$

Then use (5.9) to get

$$\begin{aligned} R_{2b} &= \sum_{z,v} \int_0^K \varphi(t,z,v)(z-\pi)[\bar{\lambda}(t) - \lambda(t,0,v)] \\ &\quad \cdot n^{-\frac{1}{2}} \Big\{ \sum_{i=1}^n I_{[Z_i=z,V_i=v]} \{Y_i(t) - S_C(t,z,v) e^{-\Lambda(t,0,v)}\} \Big\} dt \; + \; o_p(1). \\ &\equiv R_{2b}^* \; + \; o_p(1). \end{aligned} \tag{5.16}$$

Then adding all the right hand side terms from (5.14) to (5.16) leads to

$$\begin{aligned} n^{-\frac{1}{2}} \tilde{U}_{\hat{\varphi}}^{[0,K]} &= R_1^* - R_{2a}^{(1)} - R_{2a}^{(2)} - R_{2b}^* + o_p(1) \\ &= n^{-\frac{1}{2}} \tilde{U}_{\varphi}^{[0,K]} \; - \; R_{2a}^{(1)} \; + \; o_p(1). \end{aligned}$$

By Corollary 3.2.1 and Theorem 6.3.1 in Fleming and Harrington (1990), we derive that for each fixed z and v and uniformly over the compact set $[0,K]$,

$$\sqrt{n} \left\{ \frac{S_C(t,z,v)}{\hat{S}_C(t,z,v)} - 1 \right\} \;=\; \sqrt{n} \sum_{i=1}^n \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{\sum_j^n Y_j(s) I_{[Z_j=z,V_j=v]}} dM_i^C(s) \; + \; o_p(1)$$

Note that the right hand said of the above is asymptotically the following,

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{n^{-1} \sum_j^n Y_j(s) I_{[Z_j=z,V_j=v]}} dM_i^C(s) \\ =\; & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{p_{z,v} \cdot E\{Y_1(s) | Z_1 = z, V_1 = v\}} dM_i^C(s) \; + \; o_p(1) \\ =\; & n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{p_{z,v} \cdot S_C(s,z,v) e^{-\Lambda(s,0,v)}} dM_i^C(s) \; + \; o_p(1). \end{aligned}$$

Therefore $R_{2a}^{(1)}$ becomes

$$\begin{aligned} R_{2a}^{(1)} &= \sum_{z,v} \int_0^K g(t,v) e^{-\Lambda(t,0,v)} (z-\pi)[\bar{\lambda}(t) - \lambda(t,0,v)] \\ &\qquad \cdot n^{-\frac{1}{2}} \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{S_C(s,z,v)e^{-\Lambda(s,0,v)}} dM_i^C(s)dt \ + \ o_p(1). \end{aligned}$$

Hence

$$\begin{aligned} & n^{-\frac{1}{2}} \tilde{U}_{\hat{\varphi}}^{[0,K]} - n^{-\frac{1}{2}} \tilde{U}_{\varphi}^{[0,K]} \\ = & \ - \sum_{z,v} \int_0^K g(t,v) e^{-\Lambda(t,0,v)} (z-\pi)[\bar{\lambda}(t) - \lambda(t,0,v)] \\ & \quad \cdot n^{-\frac{1}{2}} \int_0^t \frac{I_{[Z_i=z,V_i=v]}}{S_C(s,z,v)e^{-\Lambda(s,0,v)}} dM_i^C(s)dt + \ o_p(1) \\ = & \ - n^{-\frac{1}{2}} \int_0^K \sum_{i=1}^n (Z_i - \pi) g(t,V_i) e^{-\Lambda(t,0,V_i)} [\bar{\lambda}(t) - \lambda(t,V_i)] \\ & \quad \cdot \{ \int_0^t \frac{dM_i^C(s)}{S_C(s,Z_i,V_i)e^{-\Lambda(s,0,V_i)}} \} dt + o_p(1) \\ \equiv & \ n^{-\frac{1}{2}} U_d^{[0,K]} \ + \ o_p(1) \end{aligned}$$

Thus

$$n^{-\frac{1}{2}} \tilde{U}_d^{[0,K]} - n^{-\frac{1}{2}} \tilde{U}_d^{[0,K]} \xrightarrow{p} 0.$$

Furthermore, Lemma A.9 and A.11 show that uniformly for all n,

$$n^{-\frac{1}{2}} \tilde{U}_{\varphi}^{(K,\infty)} \overset{p,L_1}{\to} 0 \text{ as } K \uparrow \infty,$$

$$n^{-\frac{1}{2}} \tilde{U}_{\hat{\varphi}}^{(K,\infty)} \overset{p,L_1}{\to} 0 \text{ as } K \uparrow \infty,$$

and

$$n^{-\frac{1}{2}} \tilde{U}_{d}^{(K,\infty)} \overset{p,L_1}{\to} 0 \text{ as } K \uparrow \infty.$$

Then by Lemma A.5,

$$n^{-\frac{1}{2}} \tilde{U}_d - n^{-\frac{1}{2}} U_d \ \xrightarrow{p} \ 0 \text{ as } n \to \infty. \tag{5.17}$$

Since

$$n^{-\frac{1}{2}}\hat{U}_d - n^{-\frac{1}{2}}U_d \;=\; (n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\tilde{U}_{\varphi}) - (n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\tilde{U}_{\varphi}) \;+\; (n^{-\frac{1}{2}}\tilde{U}_d - n^{-\frac{1}{2}}U_d),$$

applying (5.17) and Lemmas 5.1 and 5.2 will lead to

$$n^{-\frac{1}{2}}\hat{U}_d - n^{-\frac{1}{2}}U_d \;\xrightarrow{p}\; 0$$

as $n \to \infty$. □

Next we prove the asymptotic unbiasedness and normality of the corrected test statistic $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$. We also find its asymptotic variance Σ and a consistent variance estimator $\hat{\Sigma}$.

Theorem 5.1 *When Assumptions 1.1 and 1.3 hold and $\hat{\varphi}(t, Z_i, V_i)$ is a stratified Kaplan-Meyer estimator of $\varphi(t, Z_i, V_i)$, the random variable $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ is asymptotically normally distributed with mean zero and variance Σ, the latter being consistently estimated under H_0 by $\hat{\Sigma} = \hat{\Sigma}_1 - \hat{\Sigma}_2$, where $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are defined in (5.22) and (5.23), respectively.*

Proof.

By Lemma 5.1 and 5.2,

$$\begin{aligned} n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} &= n^{-\frac{1}{2}}\hat{U}_{\varphi} \;+\; n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_{\varphi} \\ &= n^{-\frac{1}{2}}\tilde{U}_{\varphi} \;+\; n^{-\frac{1}{2}}U_d \;+\; o_p(1). \end{aligned} \tag{5.18}$$

From DiRienzo and Lagakos (2001), we already know that $n^{-\frac{1}{2}}\tilde{U}_{\varphi}$ is a sum of iid terms with mean zero and hence is asymptotically normally distributed with mean zero and variance Σ_1, which is the probability limit of $n^{-1}\sum_{i=1}^{n} A_i^2$ with

$$A_i = \int \varphi(t, Z_i, V_i)(Z_i - \pi)\{dN_i(t) - Y_i(t)\bar{\lambda}(t)dt\}.$$

A consistent estimator of Σ_1 is

$$\Sigma_1^{(n)} \equiv n^{-1}\sum_{i}^{n}[A_i^{(n)} - \bar{A}^{(n)}]^2$$

with

$$A_i^{(n)} = \int \varphi(t, Z_i, V_i)(Z_i - \bar{Z}) \left\{ dN_i(t) - \frac{Y_i(t)}{\sum_j^n Y_j(t)\varphi(t, Z_j, V_j)} \sum_j^n \varphi(t, Z_j, V_j)dN_j(t) \right\}$$

and $\bar{A}^{(n)} = n^{-1}\sum_{i=1}^{n} A_i^{(n)}$.

From Lemma 5.3, $n^{-\frac{1}{2}}U_d$ is also an iid sum with mean

$$\begin{aligned}
& E\{\int (Z_1 - \pi)g(t, V_1)e^{-\Lambda(t,0,V_1)}[\bar{\lambda(t)} - \lambda(t, 0, V_1)]\{\int_0^t \frac{dM_1^C(s)}{S_C(s, Z_i, V_i)e^{-\Lambda(s,0,V_1)}}\}dt\} \\
= & \sum_{z,v} P_{z,v}(z - \pi)g(t, v)e^{-\Lambda(t,0,v)}E\{\int_0^t \frac{dM_1^C(s)}{S_C(s, z, v)e^{-\Lambda(s,0,v)}} \mid Z_1 = z, V_1 = v\}dt \\
= & \; 0
\end{aligned}$$

because when given z and v $\int_0^t 1/\{S_C(s, z, v)e^{-\Lambda(s,0,v)}\}dM_1^C(s)$ is an $\mathcal{F}_t$ martingale. Hence by the Central limit theorem $n^{-\frac{1}{2}}U_d$ is also asymptotically normally distributed with mean zero under H_0. Thus from (5.18) we can conclude that $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ is asymptotically normally distributed with mean zero under H_0.

Next we show how to find Σ, the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$. Note that $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} = n^{-\frac{1}{2}}\hat{U}_{\varphi} + n^{-\frac{1}{2}}\hat{U}_d$. Then

$$Var(n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}) = Var(n^{-\frac{1}{2}}\hat{U}_{\varphi}) + Var(n^{-\frac{1}{2}}\hat{U}_d) + 2Cov(n^{-\frac{1}{2}}\hat{U}_{\varphi}, n^{-\frac{1}{2}}\hat{U}_d).$$

Let Σ_2 be the asymptotic variance of $n^{-\frac{1}{2}}\hat{U}_d$, then we have $\Sigma_2 = Var(n^{-\frac{1}{2}}U_d)$. Denote

$$h_1(s,v) \equiv \int_s^\infty g(t,v)e^{-\Lambda(t,v)}[\bar{\lambda}(t) - \lambda(t,v)]dt$$
$$h_2(s,z,v) \equiv 1/\{S_C(s,z,v)e^{-\Lambda(s,0,v)}\}.$$

Then by changing the order of s and t integrals within $n^{-\frac{1}{2}}U_d$,

$$\begin{aligned} n^{-\frac{1}{2}}U_d &= -\sum_{z,v}(z-\pi)\int_0^\infty \int_s^\infty g(t,v)e^{-\Lambda(t,0,v)}[\bar{\lambda}(t) - \lambda(t,v)]dt \\ &\quad \cdot n^{-\frac{1}{2}}\sum_{i=1}^n \frac{I_{[Z_i=z,V_i=v]}}{S_C(s,z,v)e^{-\Lambda(s,0,v)}}dM_i^C(s) \\ &\equiv -\sum_{z,v}(z-\pi)\int_0^\infty h_1(s,v)n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}h_2(s,z,v)dM_i^C(s). \end{aligned}$$

which now is a sum of iid martingales with mean zero. Therefore the variance

$$\begin{aligned} \Sigma_2 &= \sum_{z,v} p_{z,v}(z-\pi)^2\int_0^\infty h_1^2(s,v)h_2^2(s,z,v)S_C(s,z,v)e^{-\Lambda(s,v)}\lambda_C(s,z,v)ds \\ &= \sum_{z,v} p_{z,v}(z-\pi)^2\int_0^\infty h_1^2(s,v)h_2(s,z,v)\lambda_C(s,z,v)ds. \end{aligned} \quad (5.19)$$

Let Σ_3 be the asymptotic covariance of $n^{-\frac{1}{2}}\hat{U}_\varphi$ and $n^{-\frac{1}{2}}\hat{U}_d$. Then we know $\Sigma_3 = Cov(n^{-\frac{1}{2}}\tilde{U}_\varphi, n^{-\frac{1}{2}}U_d)$. It is also true that

$$\Sigma_3 = E\{n^{-\frac{1}{2}}\tilde{U}_\varphi \times n^{-\frac{1}{2}}U_d\}$$

because both $n^{-\frac{1}{2}}\tilde{U}_\varphi$ and $n^{-\frac{1}{2}}U_d$ have mean zero. Recall that $n^{-\frac{1}{2}}\tilde{U}_\varphi$ can be written as

$$\begin{aligned} n^{-\frac{1}{2}}\tilde{U}_\varphi &= \sum_{z,v}(z-\pi)\int_0^\infty \varphi(t,z,v)n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}dM_i(t) \\ &\quad - \sum_{zv}(z-\pi)\int_0^\infty \varphi(t,z,v)[\bar{\lambda}(t) - \lambda(t,0,v)]n^{-\frac{1}{2}}\sum_{i=1}^n I_{[Z_i=z,V_i=v]}Y_i(t)dt \\ &\equiv B_1 + B_2. \end{aligned}$$

Therefore

$$\Sigma_3 \;=\; E\{B_1 \cdot n^{-\frac{1}{2}} U_d\} \;+\; E\{B_2 \cdot n^{-\frac{1}{2}} U_d\}.$$

The first of these two expectations is 0 because $N_i(t)$ and $N_i^C(t)$ will not jump at the same time. That is, $\langle M_1, M_1^C \rangle(t) = 0$. Hence

$$\begin{aligned}
&E\{B_1 \cdot n^{-\frac{1}{2}} U_d\} \\
= \;& \sum_{z,v} p_{z,v}(z-\pi)^2 E\left\{ \int \varphi(t,z,v) dM_1(t) \cdot \int h_1(s,v) h_2(s,z,v) dM_1(s) \right\} \\
= \;& \sum_{z,v} p_{z,v}(z-\pi)^2 E\left\{ \int \varphi(t,z,v) h_1(t,v) h_2(t,z,v) d\langle M_1, M_1^C \rangle(t) \right\} \\
= \;& 0.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&E\{B_2 \cdot n^{-\frac{1}{2}} U_d\} \\
= \;& \sum_{z,v} p_{z,v}(z-\pi)^2 E\{ \int \varphi(t,z,v)[\bar{\lambda}(t) - \lambda(t,v)] Y_1(t) dt \\
& \qquad \times \int h_1(s,v) h_2(s,z,v) dM_1^C(s)\} \\
= \;& \sum_{z,v} p_{z,v} \int \int \varphi(t,z,v)[\bar{\lambda}(t) - \lambda(t,v)] h_1(s,v) h_2(s,z,v) \\
& \qquad \times E\{Y_1(t) - Y_1(t)\lambda_C(s,v) ds\} dt \qquad (5.20)
\end{aligned}$$

Given z and v, the conditional expectation

$$\begin{aligned}
&E\{Y_1(t)[dN_1^C(s) - Y_1(s)\lambda_C(s,0,v)ds]\} \\
= \;& E\{Y_1(t) dN_1^C(s)\} \;-\; E\{Y_1(t) Y_1(s) \lambda_C(s,z,v) ds\} \\
= \;& I_{[s \geq t]} E\{Y_1(s) Y_1(t) \lambda_C(s,z,v)\} ds - E\{Y_1(s \vee t) \lambda_C(s,z,v) ds\} \\
= \;& -\, I_{[s<t]} \cdot E\{Y_1(t) \lambda_C(s,z,v) ds\} \\
= \;& -\, I_{[s<t]} S_C(t,z,v) e^{-\Lambda(t,0,v)} \lambda_C(s,z,v) ds
\end{aligned}$$

Then substitute this expression into the last line of (5.20), finding

$$
\begin{aligned}
& E\{B_2 \cdot n^{-\frac{1}{2}} U_d\} \\
= & -\sum_{z,v} p_{z,v}(z-\pi)^2 \int_0^\infty \int_0^t g(t,v) e^{-\Lambda(t,0,v)} [\bar{\lambda}(t) - \lambda(t,v)] \\
& \qquad \times h_1(s,v) h_2(s,z,v) \lambda_C(s,z,v) \cdot ds \cdot dt \\
= & -\sum_{z,v} p_{z,v}(z-\pi)^2 \int_0^\infty \int_s^\infty g(t,v) e^{-\Lambda(t,v)} [\bar{\lambda}(t) - \lambda(t,v)] dt \\
& \qquad \times h_1(s,v) h_2(s,z,v) \lambda_C(s,z,v) ds \\
= & -\sum_{z,v} p_{z,v}(z-\pi)^2 \int_0^\infty h_1^2(s,v) h_2(s,z,v) \lambda_C(s,z,v) ds \\
= & -\Sigma_2.
\end{aligned}
$$

Therefore,

$$\Sigma_3 \; = \; -\Sigma_2.$$

Thus the asymptotic variance of $n^{-\frac{1}{2}} \hat{U}_{\hat{\varphi}}$ becomes

$$\Sigma \; = \; \Sigma_1 \; + \; \Sigma_2 \; - \; 2\Sigma_2 \; = \; \Sigma_1 \; - \Sigma_2. \tag{5.21}$$

Next we find a consistent estimator $\hat{\Sigma}$ for Σ based on (5.21) and without knowing $\varphi(t,z,v)$. First, define $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ and

$$\hat{A}_i^{(n)} = \int \hat{\varphi}(t, Z_i, V_i)(Z_i - \bar{Z}) \left\{ dN_i(t) - Y_i(t) \frac{\sum_j^n \hat{\varphi}(t, Z_j, V_j) dN_j(t)}{\sum_j^n Y_j(t) \hat{\varphi}(t, Z_j, V_j)} \right\}. \tag{5.22}$$

Then

$$\hat{\Sigma}_1 \; \equiv \; n^{-1} \sum_{i=1}^n \{\hat{A}_i^{(n)} - n^{-1} \sum_j \hat{A}_j^{(n)}\}^2$$

is a consistent estimator of Σ_1.

The consistent estimator of Σ_2 can be defined from (5.19). Note that the consistent estimator of $h_1(s,v)$ for fixed v is $\hat{h}_1(s,v) =$

$$n^{-1}\sum_{i=1}^{n}\int_s^{\infty} I_{[Z_i=z,V_i=v]}\hat{\varphi}(t,z,v)\left\{dN_i(t) - Y_i(t)\frac{\sum_j \hat{\varphi}(t,Z_j,V_j)dN_j(t)}{\sum_j Y_j(t)\hat{\varphi}(t,Z_j,V_j)}\right\},$$

and a consistent estimator of $h_2(s,z,v)$ for fixed z,v is

$$\hat{h}_2(s,z,v) = \frac{1}{\sum_i Y_i(s)I_{[Z_i=z,V_i=v]}/\sum_i I_{[Z_i=z,V_i=v]}}.$$

The natural consistent estimator for $\int S_C(s,z,v)e^{-\Lambda(s,0,v)}\lambda_C(s,z,v)ds$ is

$$n^{-1}\sum_{i=1}^{n}\int I_{[Z_i=z,V_i=v]}dN_i^C(t).$$

Thus a consistent estimator for Σ_2 is

$$\hat{\Sigma}_2 = \sum_{zv}\frac{\sum_{i=1}^n I_{[Z_i=z,V_i=v]}}{n}(z-\bar{Z})^2 n^{-1}\sum_i\int \hat{h}_1^2(s,v)\hat{h}_2^2(s,z,v)I_{[Z_i=z,V_i=v]}dN_i^C(s) \tag{5.23}$$

Finally, a consistent estimator for Σ is

$$\hat{\Sigma} = \hat{\Sigma}_1 - \hat{\Sigma}_2$$

and the random variable $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}/\sqrt{\hat{\Sigma}}$ is asymptotically distributed as standard normal. □

Corollary 5.1 $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ *is asymptotically independent of* $n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$.

Proof.

Denote the asymptotic variances of $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$, $n^{-\frac{1}{2}}\hat{U}_{\varphi}$ and $n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ by Σ, Σ_1 and σ_2, respectively. In the proof of Theorem 5.1, we found that

$$\Sigma_1 = \Sigma + \Sigma_2.$$

Since $n^{-\frac{1}{2}}\hat{U}_{\varphi} = n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} + (n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}})$, so $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ is asymptotically uncorrelated with $n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$. Because $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ and $n^{-\frac{1}{2}}\hat{U}_{\varphi} - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ are asymptotically normally distributed, they are also asymptotically independent. $\square$

5.3 Aalen's Additive Model

Aalen's additive model (Aalen, 1980) may give us a way to extend the bias-correction method to a statistic with substituted semi-parametric estimates of conditional survival functions for censoring. Assume the conditional hazard function of the censoring has an additive form:

$$\lambda_C(t, Z, V) = a(t)Z + b(t, V) + c(t)ZV.$$

Denote $A(t) = \int_0^t a(s)ds$, $B(t, V) = \int_0^t b(s, V)ds$ and $C(t) = \int_0^t c(s)ds$. The weighting function φ becomes

$$\varphi(t, Z_i, V_i) = \frac{\exp\{-B(t, V_i)\} \wedge \exp\{-A(t) - B(t, V_i) - C(t)V_i\}}{\exp\{-A(t)Z_i - B(t, V_i) - C(t)Z_iV_i\}},$$

which can be further simplified as

$$\varphi(t, Z_i, V_i) = \frac{1 \wedge \exp\{-A(t) - C(t)V_i\}}{\exp\{-A(t)Z_i - C(t)Z_iV_i\}}.$$

Here $A(t)$ and $C(t)$ can all be estimated by the Ordinary Least Square estimators $\hat{A}(t)$ and $\hat{C}(t)$(Aalen, 1980) or by the Weighted Least Square estimators $\hat{A}_w(t)$ and $\hat{C}_w(t)$ (McKeague, 1988). Then the estimated φ function, as a function of the OLS or WLS estimators of $A(t)$ and $C(t)$, can be used to construct a bias-corrected logrank test. Though appeared to be a little more complicated comparing to the stratified

Kaplan-Meier estimated $\varphi(\cdot)$ function, this Aalen's Additive Model based approach does give another form of estimated $\varphi(\cdot)$ function from which a promising statistic $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ with tractable asymptotic distribution can be constructed. Especially when the numbers of levels of V gets large, this Aalen's additive model based approach does not seem to suffer from the same degradation of performance as the stratified Kaplan-Meier estimator based method. Preliminary simulation studies have shown promising results and we would like to continue studying this approach in future research.

Chapter 6

Simulations and Examples

In this chapter we are interested in analyzing and assessing the following through simulations and numerical calculations:

1. The bias of the logrank test with different types of violations of the Kong-Slud Assumption.
2. The asymptotic validity of the logrank test under the Kong-Slud Assumption.
3. The asymptotic validity of the stratified logrank test.
4. The asymptotic approximation to the distribution of the bias-corrected logrank test statistic introduced in Chapter 5.

6.1 The Bias

In this section we use both numerical calculation and simulation to study the bias of the logrank when Kong-Slud Assumption does not hold.

The example given here is based on an simple setting. Let Z be a binary random variables with values 0 or 1 and $P(Z = 1) = \frac{1}{2}$. Let V be discrete with two values 1 and 2 with $P(V = 1) = \frac{1}{2}$. Furether define he hazard function of T under H_0 is a scalar $\lambda_T(t, 0, j) = \lambda_j; j = 1, 2$ and the hazard function of C is $\lambda_C(t, Z = i, v_j) = \lambda^C_{ij}; i = 0, 1$ and $j = 1, 2$. Then the asymptotic mean and

Table 6.1: The bias of a logrank test in Numerical Calculations and Simulations

Censoring Dependence I: $\lambda_{11}^C \neq \lambda_{12}^C = \lambda_{01}^C = \lambda_{02}^C$

λ_1	λ_2	λ_{11}^C	λ_{12}^C	$\mu_0(100)$	$\hat{\mu}_0(100)$	$\mu_0(400)$	$\hat{\mu}_0(400)$	$\alpha^*(400)$	$\hat{\alpha}(400)$
1.0	2.5	3.0	1.5	.2976	.2800	.5952	.5790	.092	.092
1.0	2.5	1.0	3.7	-.2977	-.2998	-.5954	-.5960	.092	.091
3.0	0.5	1.0	3.7	.6005	.5349	1.2009	1.1553	.225	.217
1.0	2.5	0.8	3.2	-.3007	-.2958	-.6015	-.6035	.092	.099

Censoring Dependence II: $\lambda_{11}^C = 2\lambda_{01}^C$; $\lambda_{12}^C = 2\lambda_{02}^C$

λ_1	λ_2	λ_{11}^C	λ_{12}^C	$\mu_0(100)$	$\hat{\mu}_0(100)$	$\mu_0(400)$	$\hat{\mu}_0(400)$	$\alpha^*(400)$	$\hat{\alpha}(400)$
3.0	0.5	1.0	3.0	.3707	.3433	.7414	.7524	.115	.114
2.5	1.0	1.2	2.8	.1860	.1785	.3720	.3526	.065	.063
2.5	1.0	0.8	3.2	.280	0.2716	.560	0.5597	.081	.081
3.0	0.5	1.0	3.7	.4683	0.4314	.9367	0.9204	.155	.148

variance of the logrank test statistic $n^{-\frac{1}{2}}\hat{U}_L$ are calculated using the general formula in Section 2.1 and estimated by simulation with $R = 5000$ realizations under two types of dependent censoring. The actual type I errors for the tests using the logrank statistic under these assumptions are calculated.

In Table 6.1, $\mu_0(100)$ and $\mu_0(400)$ are the numerical values of $E\{n^{-\frac{1}{2}}\hat{U}_L\}$ based on Lemma 2.1 with sample sizes 100 and 400; $\hat{\mu}_0(100)$ and $\hat{\mu}_0(400)$ are the empirical estimats of the bias with sample sizes 100 and 400 from a simulation with 10000 realizations ; $\alpha^*(400)$ is the calculated size (type I error) of the test from the numerical results and $\hat{\alpha}(400)$ is the estimated size (type I error) from the simulations. Here the nominal type I error α is 0.05.

Table 6.1 confirms several results: (i) The logrank statistic is generally biased when the distribution of censoring time depends upon both the treatment group and covariates. The bias, whether positive or negative, does inflate the type I error of the test. (ii) The bias is proportional to the square root of the sample size. Here $\hat{\mu}_0(400)/\hat{\mu}_0(100) \approx 2$. (iii) The result of Lemma 2.1 is also confirmed. There is a high probability to get $\hat{\mu}_0(\cdot) \ < \ \mu_0(\cdot)$, which is true in many other simulations. This is implied by Lemma 2.1. Recall that μ_0 is calculated from $\sqrt{n} \cdot B$ and $\hat{\mu}$ is an empirical estimator of $E_0(n^{-\frac{1}{2}}\hat{U}_L)$. Lemma 2.1 shows the mean of $n^{-\frac{1}{2}}\hat{U}_L$ has an $o(\sqrt{n})$ term entering as an additive factor

$$E\{n^{-\frac{1}{2}}\hat{U}_L\} \ = \ \sqrt{n}B + o(\sqrt{n}).$$

From the proof of Lemma 2.1, we see that this $o(\sqrt{n})$ term generally enters with a sign opposite to B, which implies $\left|E\{n^{-\frac{1}{2}}\hat{U}_L\}\right| \ < \ \left|\sqrt{n}B\right|$.

6.2 Comparisons between the Logrank and Stratified Logrank Test

It would be interesting to compare the classical logrank test with the stratified logrank test when the Kong-Slud Assumption holds. We can confirm from these examples that (i) the logrank and stratified logrank statistics are both asymptotically valid under the Kong-Slud assumption; (ii) the two statistics have the same asymptotic variance under the Kong-Slud assumption; (iii) the stratified logrank statistic is asymptotically valid when Kong-Slud assumption fails and the logrank statistic is biased; (iv) the power decreases when the number of strata increases within the same set of data and (v) the contiguous theory we proved in Chapter 4 appears to hold approximately in finite samples.

In these examples, denote

$\hat{\mu}$: empirical means of $n^{-\frac{1}{2}}\hat{U}_L$ or $n^{-\frac{1}{2}}\hat{U}_S$

$\hat{S}$: empirical standard deviation of $n^{-\frac{1}{2}}\hat{U}_L$ or $n^{-\frac{1}{2}}\hat{U}_S$

$\hat{S}_\mu$: the square root of the empirical averages of $\hat{V}_L$ and $\hat{V}_S$.

N : number of realizations in simulation

n : sample size, number of subjects in one simulation realization.

Case 1: Under Kong-Slud Assumption and H_0

Assume the conditional survival function for T and C are

$$S(t, z, v) = \exp\{-e^{\theta_1 z + \beta_1 v} t\}$$

and

$$S_C(t, z, v) = \exp\{-(e_2^{\theta} + e_2^{\beta})t\}$$

respectively. Let V be discrete with values

$$(v_1, \ldots, v_{10}) = (1, 3, 4, 6, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5)$$

and $\theta_1 = 0$; $\beta_1 = 0.1$; $\theta_2 = -.8$; $\beta_2 = -1$; $P(V = v_i) = p_i = .1$; $i = 1, \ldots, 10$, $N = 10000$. The result can be found at Table 6.2.

Case 2: Under Kong-Slud Assumption and H_n

Here we keep the same setting as in case 1 except for assigning $\theta_1 = -4/\sqrt{n}$. Table 6.3 shows an increase trend of the Asymptotic Relative Efficiency (ARE) of

Table 6.2: Logrank and Stratified Logrank Under KS and H_0, $N = 10000$

$n = 400$	$\hat{\mu}$	95% CI for μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	-0.0041	(-0.0118, 0.0036)	0.3937	0.3958	(0.3957, 0.3960)
Str. LG	-0.0065	(-0.014, 0.0009)	0.3769	0.3781	(0.3779, 0.3782)
$n = 900$	$\hat{\mu}$	95% CI for μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	0.0034	(-0.0044, 0.0112)	0.3984	0.3973	(0.3973, 0.3974)
Str. LG	0.0044	(-0.0032, 0.0121)	0.3879	0.3883	(0.3882, 0.3884)
$n = 1600$	$\hat{\mu}$	95% CI for μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	0.0018	(-0.0060, 0.0096)	0.3985	0.3979	(0.3979, 0.3980)
Str. LG	0.0007	(-0.0070, 0.0084)	0.3912	0.3923	(0.3923, 0.3924)

the stratified logrank statistic to the logrank statistic when Kong-Slud Assumption holds and under $H_n : \theta_1 = b/\sqrt{n}$.

Case 3: When Kong-Slud Assumption Fails and under H_0

Let $\lambda_C(t, 1, v_4) = 1.2$ and $\lambda_C(t, z, v) = 1.6 for z \neq 1$ and $v \neq v_4$, $\theta_1 = 0; \beta_1 = 0.1$ and $(v_1, \ldots, v_{10}) = (c, 3, 4, 6, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5)$. Table 6.4 shows the relative unbiasedness of the stratified logrank statistic and an increase proportional to $\sqrt{n}$ of bias for the logrank statistic.

6.3 Simulations on the Bias-corrected Logrank Test

In this section we use simulations to study properties of the bias-corrected test statistics introduced in Chapter 5.

Table 6.3: Logrank and Stratified Logrank Under KS and H_n, $N = 10000$

$n = 400$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	ARE
LG	-0.58379	(-0.59119, -0.57638)	0.37779	0.38306	
Str. LG	-0.54263	(-0.54971, -0.53555)	0.36121	0.36776	0.94511
$n = 900$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	ARE
LG	-0.60170	(-0.60928, -0.59412)	0.38683	0.38886	
Str. LG	-0.58477	(-0.59219, -0.57736)	0.37844	0.38077	0.98686
$n = 1600$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	ARE
LG	-0.60117	(-0.60888, -0.59346)	0.39343	0.39175	
Str. LG	-0.59550	(-0.60311, -0.58789)	0.38808	0.38666	1.00845
$n = 2500$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	ARE
LG	-0.61118	(-0.61875, -0.60361)	0.38637	0.39322	
Str. LG	-0.61068	(-0.61819, -0.60318)	0.38297	0.38975	1.0162

Table 6.4: Logrank and Stratified Logrank Without KS and under H_0, $N = 10000$

$n = 400$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	0.00775	(0.00101, 0.01449)	0.34383	0.34381	(0.34368, 0.34393)
Str. LG	0.00036	(-0.00605, 0.00678)	0.32738	0.32750	(0.32737, 0.32762)
$n = 900$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	0.01499	(0.00823, 0.02174)	0.34476	0.34537	(0.34529, 0.34545)
Str. LG	0.00327	(-0.00330, 0.00985)	0.33556	0.33672	(0.33664, 0.33680)
$n = 1600$	$\hat{\mu}$	95% CI of μ	$\hat{S}$	$\hat{S}_\mu$	95% CI of S_μ
LG	0.01727	(0.01043, 0.0241)	0.34878	0.34601	(0.34595, 0.34607)
Str. LG	0.00317	(-0.00355, 0.0099)	0.34304	0.34056	(0.34049, 0.34062)

6.3.1 Bias-corrected Logrank Test Under H_0

Let $\lambda(t,z,v) = \exp(\theta z)\exp(\beta v)$ and $\lambda_C(t,z,v) = \gamma_1 zv + \gamma_2 v + \gamma_3 z$ be the hazard functions of T and C, respectively. According to Assumption 1.5, the logrank test statistic $n^{-\frac{1}{2}}\hat{U}_L$ in this example is biased. Then several bias-corrected test statistic are considered: $n^{-\frac{1}{2}}\hat{U}_\varphi$, $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$, and $n^{-\frac{1}{2}}\hat{U}_S$, where $n^{-\frac{1}{2}}\hat{U}_\varphi$ is the modified logrank statistic with known distribution function of censoring and $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$ is the one based on Kaplan-Meier estimated distribution function of censoring time.

In the first example (Table 6.5), let $\theta = 0, \beta = -.4$, $\gamma_1 = .1$, $\gamma_2 = .2$ and $\gamma_3 = .1$. Here $P(V = v) = 0.5$ with $v \in \{1,2\}$ and $P(Z = 1) = 0.5$. The number of simulations $R = 2500$. Let $V_1 = \mathrm{Var}(n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_\varphi), V_2 = \mathrm{Cov}(n^{-\frac{1}{2}}\hat{U}_\varphi, n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_\varphi)$ and $V_3 = \mathrm{Cov}(n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}, n^{-\frac{1}{2}}\hat{U}_\varphi - n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}})$ are three variance or covariance quantities that are of interest to us to confirm our findings in Chapter 5. According to the proof of Theorem 5.1, the covariances $V_2 = -V_1$ and $V_3 = 0$. In Table 6.5 - 6.7, $\hat{V}_1$, $\hat{V}_2$ and $\hat{V}_3$ are their empirical estimator and V_1 is the numerical calculation based on Theorem 5.1.

Due to the choice of covariates in the above example, the magnitude of V_1 and V_2, contrasting to $\mathrm{Var}\{n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}\}$ and $\mathrm{Var}\{n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}\}$, is very small. In the second example (Table 6.6), let $\theta = 0, \beta = .4$, $\gamma_1 = -.1$, $\gamma_2 = .2$ and $\gamma_3 = .1$. Here $P(V = v) = 0.5$ with $v \in \{1,9\}$ and $P(Z = 1) = 0.5$. The number of simulations is still $R = 2500$. Again Table 6.6 confirms Theorem 5.1. In addition, Table 6.6 shows that the variance of $n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}} - n^{-\frac{1}{2}}\hat{U}_\varphi$ is quite substantial when difference between levels of covariates is large.

Table 6.5: Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_0, Example 1

Mean	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	0.0010	0.0006	0.0020	0.0126
n=200	-0.0041	-0.0038	-0.0029	0.0132
n=300	-0.0055	-0.0040	-0.0028	0.0163
n=400	0.0028	0.0023	0.0031	0.0283

StDev	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	0.3334	0.3310	0.3608	0.3646
n=200	0.3376	0.3370	0.3680	0.3708
n=300	0.3426	0.3423	0.3729	0.3766
n=400	0.3427	0.3425	0.3765	0.3786

	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$
n=100	0.00269	-0.00214	-0.00055
n=200	0.00199	-0.00119	-0.00080
n=300	0.00171	-0.00095	-0.00077
n=400	0.00168	-0.00091	-0.00077

Table 6.6: Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_0, Example 2

Mean	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	-.0411	-.0023	-.0005	-.2033
n=400	-.0354	.0015	0	-.4243
StDev	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	0.3260	0.2736	0.2992	0.2976
n=400	0.3405	0.2882	0.3111	0.3042
	$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	V_1
n=100	.0304	-.0309	0	.0310
n=400	.0417	-.0373	0	.0310

6.3.2 Bias-corrected Logrank Test under H_A

In this example (Table 6.7), let $\theta = 2/\sqrt{n}, \beta = .4$, $\gamma_1 = 2/\sqrt{n}$, $\gamma_2 = .2$ and $\gamma_3 = .1$. Here $P(V = v) = 0.5$ with $v \in \{1, 4\}$ and $P(Z = 1) = 0.5$. Table 6.7 shows efficacy of these bias-corrected tests under the alternatives $H_n : \theta = 2/\sqrt{n}, \gamma_1 = 2\sqrt{n}$ that are contiguous to $H_0 : \theta = \gamma_1 = 0$ as proved in Theorem 4.1. In this example it turns out that the stratified logrank test is the most efficient among the three bias-corrected tests while different result showed in other simulation examples. At this moment there is no conclusive result that which test is the most powerful. The result of an ongoing study on power comparisons among these Bias-corrections will be provided in the near future.

Table 6.7: Bias-corrected Tests When Kong-Slud Assumption Fails and Under H_A

Mean	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	-.1271	-.1272	-.1478	-.0857
n=400	-.1482	-.1515	-.1785	-.0815
StDev	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	0.2441	0.2421	0.2748	0.2772
n=400	0.2713	0.2680	0.2945	0.2987
Efficacy	$n^{-\frac{1}{2}}\hat{U}_{\varphi}$	$n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}$	$n^{-\frac{1}{2}}\hat{U}_S$	$n^{-\frac{1}{2}}\hat{U}_L$
n=100	.2713	.2761	.2894	.0955
n=400	.2982	.3196	.3672	.0744

Chapter 7

Summary of Results and Future Research Problems

In this chapter we give a summary of the results of this thesis and point out future research problems that have arisen from this thesis.

1. We have proved that the logrank statistic is asymptotically distributed as a standard normal under the null hypothesis of no treatment effect if the Kong-Slud Assumption holds. However, if the assumption fails, the logrank rank test is generally biased and the potential bias can cause serious validity problems in clinical trials. Therefore how to practically verify the Kong-Slud Assumption is always an interesting and meaningful future research problem for us.

2. The classical results of the stratified logrank statistic have been reviewed in this thesis. Under suitable regularity conditionas, a complete stratified logrank statistic, as defined in 1.3.5, is still valid when Kong-Slud Assumptions fails. Therefore we would suggest that investigators use the stratified logrank test because the potential bias is a major concern to the validity of the clinical trials . At this moment we only make such a suggestion for moderately stratified studies because of the widespread belief that a heavily stratified statistic may have substantial loss in efficiency (Green and Byar, 1978; Schoenfeld and Tsiatis, 1987). Hence we are interested in doing a systematic study on the relation between power and stratification. Properties of a stratified logrank

test for which the number of strata increases with respect to the sample size will be of interest to us in the future.

3. We have studied good properties of the partially stratified logrank statistic, such as $n^{-\frac{1}{2}}\hat{U}_W$ defined in Section 1.3.6, under the Kong-Slud Assumption in Chapter 3. We also admit the fact that this test is also generally biased if the Kong-Slud Assumption fails. Remember that we have the dilemma of "Bias-free" v.s "Heavily-stratified" above. A possible way to step out of this dilemma might to use a "hybrid" approach: do a preliminary analysis to detect covariates which may have interactions with the treatment group on the conditional distribution of the censoring and then carry out a stratified logrank test that only stratifies on the "detected" covariates or "suspected" important covariates. It is very possible to get a moderately stratified test by using this approach. And in theory, with suitable regularity conditions, we are able to prove this test is asymptotically valid. The key component of this approach is the preliminary analysis, in which not only proper statistical analytical tools are important, but also experience and expert opinions are crucial. We would like to have some input in this approach.

4. We have established a contiguous alternative theory regarding small violations of the Kong-Slud Assumption in Chapter 4. We will carry out numerical large-sample power studies using theoretical formulas to contrast logrank or stratified-logrank tests, especially W-stratified ones, with tests based on estimated phi-functions.

5. We have discussed and extended the bias-correction method proposed by DiRenzo and Lagakos (2001a) in Chapter 5. Our work solidifies the practical use of this method when the distribution function for censoring is unknown and estimated. Due to our work a correct consistent variance estimator has been found. Furthermore, we find that a parametric model estimation for the distribution of the censoring will also give a valid bias-corrected logrank test. We will extend the theorems about $\hat{U}_{\hat{\varphi}}$ to the case where estimation of the conditional censoring distributions is done by a parametric model. Another interesting approach is to assume Aalen's additive model. A kernel smoothed weighted least square estimator for the distribution of censoring has been found, partially by simulation, to be good to use. How to fully show the applicability, either by theory or simulations, of this new direction with Aalen's additive model will be an interesting future research problem to work with.

Chapter A

Appendix: Lemmas and Proofs

Lemma A.1 *If Assumption 1.1 and 1.3 hold, we have*

(i) $E_0\{Z \mid V, Y(t) = 1\} \ = \ E\{Z|V, C \geq t\}$.

(ii) The Kong-Slud Assumption I (Assumption 1.4) implies $E_0\{Z|Y(t) = 1\} = E\{Z|V, C \geq t\}$.

Proof:

First we prove (i) without Assumption 1.4. Assumption 1.3 implies that under the null hypothesis of no treatment effect,

$$E_0\{Y(t) = 1|Z, V\} = P_0\{T \geq t|V\}P\{C \geq t|Z, V\}.$$

Then we have

$$\begin{aligned}
& E_0\{Z \mid V, Y(t) = 1\} \\
= \ & \frac{P_0\{Z = 1, Y(t) = 1|V\}}{P_0\{Y(t) = 1|V\}} \\
= \ & \frac{P_0\{Y(t) = 1 \,|Z = 1, V\}P\{Z = 1 \,|V\}}{E_0\{E_0[Y(t)|V]|Z, V\}} \\
= \ & \frac{P_0\{T \geq t|V\}P\{C \geq t|Z = 1, V\}P\{Z = 1\}}{P_0\{T \geq t|V\}[P\{Z = 1\}P\{C \geq t|Z = 1, V\} + P\{Z = 0\}P\{C \geq t|Z = 0, V\}]} \\
= \ & \frac{P\{Z = 1, C \geq t|V\}}{P\{C \geq t|V\}} \\
= \ & P\{Z = 1|C \geq t, V\} \\
= \ & E\{Z|V, C \geq t\}.
\end{aligned}$$

Hence (i) is proved. (ii) is a direct result (i) and Assumption 1.4. $\square$

Lemma A.2 *If Assumptions 1.1 and 1.3 hold, then*

(i) Assumption 1.5 implies Assumption 1.4;

(ii) Assumption 1.6 implies Assumption 1.5.

Proof.

(i) Assumptions 1.1 and 1.5 imply that Z and V are conditionally independent given $C \geq t$, and we also know that Z and V are conditionally independent given $T \geq t$ under H_0. Conditional independence of Z and V given $Y(t) = 1$ under H_0 in Assumption 1.4 follows. The following is the proof with details.

Assumption 1.5 expresses the survival function of the censoring is

$$S_C(t|z, v) = e^{-a(t,z)} \cdot e^{-b(t,v)}.$$

Then from Assumptions 1.1 and 1.3,

$$\mu(t) = \frac{E_0[ZY(t)]}{E_0[Y(t)]} = \frac{E[Ze^{-a(t,Z)}]}{E[e^{-a(t,Z)}]}.$$

From Lemma A.1,

$$\begin{aligned} E_0\{Z \,|\, V, Y(t) = 1\} &= \frac{P(Z = 1)P(C \geq t|Z = 1, V)}{P(C \geq t|V)} \\ &= \frac{P(Z = 1)e^{-a(t,1)} \cdot e^{-b(t,V)}}{E[e^{-a(t,Z)}|V] \cdot e^{-b(t,V)}} \\ &= \frac{E[Ze^{-a(t,Z)}]}{E[e^{-a(t,Z)}]} \end{aligned}$$

hence we have

$$E_0\{Z \,|\, V, Y(t) = 1\} = \mu(t),$$

which is Assumption 1.4.

(ii) Assumption 1.6 says that the distribution of censoring depends only on covariates or only on treatment group. Both of these are special cases of Assumption 1.5.□

Lemma A.3 *For $i = 1, \ldots, n$ with $N_i(t), Y_i(t), Z_i$ are defined as in Section 1.1.*

$$\sup_{0\le t<\infty} |\frac{1}{n}\sum_{i=1}^{n} N_i(t) - E[N_1(t)]| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0; \tag{A.1}$$

$$\sup_{0\le t<\infty} |\frac{1}{n}\sum_{i=1}^{n} Y_i(t) - E[Y_1(t)]| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0; \tag{A.2}$$

$$\sup_{0\le t<\infty} |\frac{1}{n}\sum_{i=1}^{n} Z_iY_i(t) - E[Z_1Y_1(t)]| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0; \tag{A.3}$$

Proof.

1. Since $E[N_1(t)] = Pr[T_1 \le t, C_1 \le T_1]$ is a subdistribution function and $N_i(t) = I_{[T_i\le t]}I_{[T_i\le C_i]}$ are iid indicator functions that are monotone, nondecreasing and right-continuous in t, by the Glivenko-Cantelli Theorem (Section 19.1, van der Vaart, 1998),

$$\sup_{0\le t<\infty} |\frac{1}{n}\sum_{i=1}^{n} N_i(t) - E[N_1(t)]| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0$$

2. For $i = 1, \ldots, n$, Let $X_i = \min\{T_i, C_i\}$. Then

$$E[1 - Y(t+)] = P[X \le t]$$

is the distribution function of X. By the Glivenko-Cantelli Theorem,

$$\sup_{0\le t<\infty} |\frac{1}{n}[1 - \sum_{i=1}^{n} Y_i(t)] - E[1 - Y_1(t)]| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0,$$

proving (A.2).

3. Since

$$E[Z(1 - Y(t))] = P[Z = 1, X \le t]$$

is a sub-distribution function, by the Glivenko-Cantelli Theorem,

$$\sup_{0\le t<\infty} |\frac{1}{n}[\sum_{i=1}^{n} Z_i(1-Y_i(t))] - E[Z(1-Y_1(t))]| \overset{P,L^2(\Omega)}{\longrightarrow} 0.$$

By the Law of Large Numbers,

$$|\frac{1}{n}\sum_{i=1}^{n} Z_i - E(Z)| \overset{P,L^2(\Omega)}{\longrightarrow} 0.$$

Thus, (A.3) is proved through the previous lines and the triangle inequality:

$$\begin{aligned}
& \sup_{0\le t<\infty} |\frac{1}{n}\sum_{i=1}^{n} Z_iY_i(t) - E[Z_1Y_1(t)]| \\
= \quad & \sup_{0\le t<\infty} |-\frac{1}{n}\sum_{i=1}^{n} Z_i(1-Y_i(t)) + \frac{1}{n}\sum_{i=1}^{n} Z_i + E[Z(1-Y_1(t))] - E[Z_1]| \\
\le \quad & \sup_{0\le t<\infty} |\frac{1}{n}[\sum_{i=1}^{n} Z_i(1-Y_i(t))] - E[Z(1-Y_1(t))]| + |\frac{1}{n}\sum_{i=1}^{n} Z_i - E(Z)| \\
\overset{P,L^2(\Omega)}{\longrightarrow} \quad & 0. \qquad\qquad (A.4)
\end{aligned}$$

□

Lemma A.4 *Given $K > 0$ and $0 < c_1 < \infty$, let $\hat{A}^{(n)}(t)$ and $\hat{B}^{(n)}(t)$ be sequences of stochastic processes with $A(t) = E\{\hat{A}^{(n)}(t)\}$ and $B(t) = E\{\hat{B}^{(n)}(t)\}$. If $\sup_{0\le t\le K} |\hat{A}^{(n)}(t)/\hat{B}^{(n)}(t)| \le c_1$, $\sup_{0\le t\le K} |1/B(t)| < \infty$ and*

$$\sup_{0\le t\le K} |\hat{A}^{(n)}(t) - A(t)| \overset{p,L_1}{\to} 0$$

$$\sup_{0\le t\le K} |\hat{B}^{(n)}(t) - B(t)| \overset{p,L_1}{\to} 0$$

as $n \to \infty$. Then

$$\sup_{0\le t\le K} \left|\frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} - \frac{A(t)}{B(t)}\right| \overset{p,L_1}{\to} 0$$

as $n \to \infty$.

Proof.

Since

$$\frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} - \frac{A(t)}{B(t)} \;=\; \frac{1}{B(t)}[\hat{A}^{(n)}(t) - A(t)] \;-\; \frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} \cdot \frac{1}{B(t)}[\hat{B}^{(n)}(t) - B(t)],$$

it follows that

$$\begin{aligned}
&\sup_{0\le t\le K} \Big| \frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} - \frac{A(t)}{B(t)} \Big| \\
\le\;& \sup_{0\le t\le K} \Big| \frac{1}{B(t)}[\hat{A}^{(n)}(t) - A(t)] \Big| + \sup_{0\le t\le K} \Big| \frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} \cdot \frac{1}{B(t)}[\hat{B}^{(n)}(t) - B(t)] \Big| \\
\le\;& \sup_{0\le t\le K} \Big| \frac{1}{B(t)} \Big| \cdot \Big\{ \sup_{0\le t\le K} \Big| \hat{A}^{(n)}(t) - A(t) \Big| + c_1 \cdot \sup_{0\le t\le K} \Big| \hat{B}^{(n)}(t) - B(t) \Big| \Big\}.
\end{aligned} \tag{A.5}$$

From (A.5), the hypotheses of the Lemma immediately imply that as $n \to \infty$,

$$\sup_{0\le t\le K} \Big| \frac{\hat{A}^{(n)}(t)}{\hat{B}^{(n)}(t)} - \frac{A(t)}{B(t)} \Big| \overset{p,L_1}{\to} 0$$

$\square$.

Lemma A.5 *Let $G_n(t)$ and $\tilde{G}_n(t)$ be two sequences of stochastic processes such that*

$$\int_0^K \{G_n(t) - \tilde{G}_n(t)\}dt \overset{p}{\to} 0 \text{ as } n \to \infty$$

for any $K > 0$. If both

$$\int_K^\infty G_n(t)dt \overset{p}{\to} 0 \text{ and} \int_K^\infty \tilde{G}_n(t)dt \overset{p}{\to} 0 \text{ as } K \uparrow \infty$$

uniformly in n, then

$$\int_0^\infty \{G_n(t) - \tilde{G}_n(t)\}dt \overset{p}{\to} 0 \text{ as } n \to \infty.$$

Proof.

For any $\epsilon > 0$, there exists a real number $K > 0$ such that, for all n,

$$P\{\Big|\int_K^\infty G_n(t)dt\Big| > \frac{\epsilon}{3}\} < \frac{\epsilon}{3}, \;\; P\{\Big|\int_K^\infty \tilde{G}_n(t)dt\Big| > \frac{\epsilon}{3}\} < \frac{\epsilon}{3}.$$

Then for fixed K, there exists an integer $N > 0$ such that for any integer $n > N$,

$$P\{\Big|\int_0^K \{G_n(t) - \tilde{G}_n(t)\}dt\Big| > \frac{\epsilon}{3}\} \;<\; \frac{\epsilon}{3}.$$

Therefore

$$\begin{aligned}
& P\{\Big|\int_0^\infty \{G_n(t) - \tilde{G}_n(t)\}dt\Big| \;>\; \epsilon\} \\
\leq \;& P\{\Big|\int_K^\infty G_n(t)dt\Big| > \frac{\epsilon}{3}\} + P\{\Big|\int_K^\infty G_n(t)dt\Big| > \frac{\epsilon}{3}\} \\
& \qquad\qquad + P\{\Big|\int_0^K \{G_n(t) - \tilde{G}_n(t)\}dt\Big| > \frac{\epsilon}{3}\} \\
< \;& \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon
\end{aligned}$$

Hence

$$\int_0^\infty \{G_n(t) - \tilde{G}_n(t)\}dt \xrightarrow{p} 0$$

as $n \to \infty$. □

Lemma A.6 *For $i = 1, \ldots, n$, let $N_i(t), Y_i(t), Z_i$ be defined as in Section 1.1. Then under Assumptions 1.1 and 1.3,*

$$\textit{(i)} \int_K^\infty dN_1(t) \xrightarrow{p,L_1} 0 \textit{ as } K \uparrow 0;$$

$$\textit{(ii)} \int_K^\infty Y_1(t)\lambda(t, 0, V_1)dt \xrightarrow{p,L_1} 0 \textit{ as } K \uparrow 0.$$

Proof.

By definition $N_1(t) = I_{[T_1 \le C_1]} \cdot I_{[T_1 \le t]}$, so that $N_1(\infty) = I_{[T_1 \le C_1]}$. Thus,

$$
\begin{aligned}
E\{|\int_K^\infty dN_1(t)|\} &= E\{N_1(\infty) - N_1(K)\} \\
&= E\{I_{[T_1 \le C_1]} \cdot I_{[T_1 > K]}\} \\
&\le E\{I_{[T_1 > K]}\} \\
&= P\{T_1 > K\} \ \to 0 \text{ as } K \uparrow \infty.
\end{aligned}
$$

Hence (i) is proved.

Since $M_1(t) = N_1(t) - \int Y_1(t)\lambda(t, 0, V_1)dt$ is an $\mathcal{F}_t$ martingale with mean 0,

$$
\begin{aligned}
& E\{|\int_K^\infty Y_1(t)\lambda(t, 0, V_1)dt|\} \\
= \ & -E\{\int_K^\infty dM_1(t)\} + E\{\int_K^\infty dN_1(t)\} \\
\to \ & 0 \text{ as } K \uparrow \infty.
\end{aligned}
$$

Hence (ii) is proved. □

Lemma A.7 *For $K > 0$, Define*

$$
n^{-\frac{1}{2}}\hat{U}_L^{(K,\infty)} \equiv n^{-\frac{1}{2}} \sum_{i=1}^n \int_K^\infty [Z_i - \frac{\sum_{i=1}^n Z_i Y_i(t)}{\sum_{i=1}^n Y_i(t)}]dN_i(t)
$$

and

$$
n^{-\frac{1}{2}}U_L^{(K,\infty)} \equiv n^{-\frac{1}{2}} \sum_{i=1}^n \int_K^\infty [Z_i - \mu(t)][dN_i(t) - Y_i(t)\frac{E[Y(t)\lambda(t, 0, V)]}{E[Y(t)]}dt].
$$

Then under Assumptions 1.1, 1.3 and 1.4, uniformly over all n,

$$
\textit{(i)}\ n^{-\frac{1}{2}}\hat{U}_L^{(K,\infty)} \overset{p, L_1}{\longrightarrow} 0 \ \textit{as}\ K \uparrow \infty
$$

$$\textit{(ii)}\ n^{-\frac{1}{2}}U_L^{(K,\infty)} \stackrel{p,L_1}{\longrightarrow} 0 \textit{ as } K \uparrow \infty.$$

Proof.

(i) Define

$$\bar{Z}(t) \equiv \frac{\sum_{i=1}^n Z_i Y_i(t)}{\sum_{i=1}^n Y_i(t)},$$

then $n^{-\frac{1}{2}}\hat{U}_L^{(K,\infty)}$ can be rewritten as

$$\begin{aligned} n^{-\frac{1}{2}}\hat{U}_L^{(K,\infty)} &= n^{-\frac{1}{2}}\sum_{i=1}^n \int_K^\infty [Z_i - \mu(t)]dN_i(t) \ + \ n^{-\frac{1}{2}}\sum_{i=1}^n \int_K^\infty [\mu(t) - \bar{Z}(t)]dM_i(t) \\ &\quad + n^{-\frac{1}{2}}\sum_{i=1}^n \int_K^\infty [\mu(t) - \bar{Z}(t)]Y_i(t)\lambda(t,0,V_i)dt \\ &\equiv A_1 \ + \ A_2 \ + \ A_3 \end{aligned} \tag{A.6}$$

Note that A_1 is an iid sum with mean 0 and A_2 is a sum of martingales. Therefore,

$$\begin{aligned} E\{A_1^2\} &= E\{\int_K^\infty [Z_1 - \mu(t)]dN_1(t)\}^2 \\ &= E\{\int_K^\infty [Z_1 - \mu(t)]^2 dN_1(t)\} \\ &\leq E\{\int_K^\infty dN_1(t)\} \rightarrow 0 \text{ as } K \uparrow \infty. \end{aligned}$$

Hence $A_1 \stackrel{L^2}{\rightarrow} 0$ as $K \uparrow \infty$ uniformly for all n. Furthermore, because when $i \neq j$

$$E\{\int_K^\infty [\mu(t) - \bar{Z}(t)]dM_i(t) \int_K^\infty [\mu(t) - \bar{Z}(t)]dM_j(t)\} = 0,$$

$$\begin{aligned} E\{A_2^2\} &= E\{\int_K^\infty [\mu(t) - \bar{Z}(t)]dM_1(t)\}^2 \\ &= E\{\int_K^\infty [\mu(t) - \bar{Z}(t)]^2 d\langle M_1\rangle(t)\} \\ &= E\{\int_K^\infty [\mu(t) - \bar{Z}(t)]^2 Y_1(t)\lambda(t,0,V_1)dt\} \\ &\leq E\{\int_K^\infty Y_1(t)\lambda(t,0,V_1)dt\} \\ &= E\{\int_K^\infty dN_1(t)\} \rightarrow 0 \text{ as } K \uparrow \infty, \end{aligned}$$

where $\langle M_i \rangle$ is the predictable process of M_i as defined in Section 1.3.1. Theorem II.3.1 of Andersen et al (1992) was applied in above derivations. Hence $A_2 \xrightarrow{L^2} 0$ as $K \uparrow \infty$ uniformly for all n.

Next consider A_3, which can be rewritten as

$$A_3 = -n^{-\frac{1}{2}} \int_K^\infty \frac{\sum_{i=1}^n [Z_i - \mu(t)] Y_i(t)}{\bar{Y}(t)} \sum_{j=1}^n Y_j(t) \lambda(t, 0, V_j) dt.$$

Hence

$$\begin{aligned} E\{A_3^2\} &= E\{\frac{1}{n} \frac{\sum_{i=1}^n [Z_i - \mu(t)] Y_i(t)}{\bar{Y}(t)} \frac{\sum_{k=1}^n [Z_k - \mu(s)] Y_k(s)}{\bar{Y}(s)} \sum_{j=1}^n Y_j(t) \lambda(t, 0, V_j) \\ &\quad \times \sum_{l=1}^n Y_l(s) \lambda(s, 0, V_l) \end{aligned} \tag{A.7}$$

Let

$$\mathbf{Y}(t) \equiv \{Y_1(t), Y_2(t), \ldots, Y_n(t))\}$$

$$\mathbf{V} \equiv \{V_1, V_2, \ldots, V_n\}.$$

By Assumption 1.4,

$$E\{Z_i \mid \mathbf{Y}(t), \mathbf{V}\} = E\{Z_i \mid Y_i(t) = 1\} = \mu(t).$$

Hence

$$E\left\{ \sum_i [Z_i - \mu(t) Y_i(t)] \Big| \mathbf{Y}(t), \mathbf{V} \right\} = 0$$

and

$$E\left\{ \sum_{i \neq k} [Z_i - \mu(t)] Y_i(t) [Z_k - \mu(s)] Y_k(s) \right\} = 0.$$

Therefore (A.7) can be further simplified as

$$\begin{aligned} E\{A_3^2\} &= E\Big\{ \frac{1}{n} \int_K^\infty \int_K^\infty \frac{\sum_{i=1}^n [Z_i - \mu(s)][Z_i - \mu(t)] Y_i(s) Y_i(t)}{\bar{Y}(t) \bar{Y}(s)} \\ &\quad \sum_{j,k} Y_j(t) \lambda(t, 0, V_j) Y_k(s) \lambda(s, 0, V_k) dt ds \Big\} \end{aligned} \tag{A.8}$$

Recall that by bringing the absolute value into the expectation on the right hand side of (A.8), we estimated

$$\sum_{i=1}^{n} \big|[Z_i - \mu(s)][Z_i - \mu(t)]\, Y_i(s)\, Y_i(t)\big| \;\leq\; \sum_{i=1}^{n} Y_i(\max(s,t)) \;=\; \bar{Y}(s \vee t).$$

Also, by symmetry in s, t, we write the double integral as twice the integral over the region with $s < t$, so from (A.8) we know that for fixed K,

$$E\{A_3^2\} \leq \frac{2}{n}\, E \sum_{j,k=1}^{n} \int_K^\infty \int_s^\infty \frac{Y_j(t)\, Y_k(s)}{\bar{Y}(s)}\, \lambda(t, 0, V_j)\, \lambda(s, 0, V_k)\, dt\, ds. \qquad \text{(A.9)}$$

To simplify this expression further , we observe that when $s < t$ and $Y_j(t) = Y_k(s) = 1$,

$$\bar{Y}(s) \geq 1 + \sum_{i:\ i \neq j,k} Y_i(s).$$

So that using independence of $(Y_i(\cdot), V_i)$ for different i, the right hand side of (A.9) is bounded above by

$$\frac{2}{n}\, E \sum_{j,k=1}^{n} \int_K^\infty \int_s^\infty Y_j(t)\, Y_k(s)\, E\{[1 + \sum_{i:i \neq j,k} Y_i(s)]^{-1}\}\, \lambda(t, 0, V_j)\, \lambda(s, 0, V_k)\, dt\, ds \qquad \text{(A.10)}$$

Next we separate the double sum over j and k of (A.10) into two terms, one including all the terms with $j = k$, and the other consisting of all terms with $j \neq k$. The first such sum is

$$\frac{2}{n}\, E \sum_{j=1}^{n} \int_K^\infty \int_K^t Y_j(t)\, E\{[1 + \sum_{i:i \neq j} Y_i(s)]^{-1}\}\, \lambda(t, 0, V_j)\, \lambda(s, 0, V_j)\, ds\, dt,$$

which is crudely bounded above by the normalized expected *iid* sum

$$
\begin{aligned}
& \frac{2}{n} E \sum_{j=1}^{n} \int_K^{\infty} Y_j(t)\; \Lambda(t,0,V_j)\, \lambda(t,0,V_j)\, dt \\
= \quad & 2\, E \int_K^{\infty} S_{C|V}(t|V_1)\, e^{-\Lambda(t,0,V_1)}\, \Lambda(t,0,V_1)\, \lambda(t,0,V_1)\, dt \\
\leq \quad & 2\, E \int_K^{\infty} e^{-\Lambda(t,0,V_1)}\, \Lambda(t,0,V_1)\, \lambda(t,0,V_1)\, dt \\
= \quad & 2\, E \int_{\Lambda(K,0,V_1)}^{\infty} xe^{-x}\, dx \to 0 \text{ as } K \uparrow \infty. \qquad\qquad (A.11)
\end{aligned}
$$

When $j \neq k$, the conditional expectation is bounded as follows:

$$
\begin{aligned}
& E\Big\{ \int_s^{\infty} Y_j(t)\, \lambda(t,0,V_j)\, dt \,\Big|\, V_j, V_k, Y_i(\cdot)_{i\neq j} \Big\} \\
= \quad & E\Big\{ \int_s^{\infty} e^{-\Lambda(t,0,V_j)} S_{C|V}(t|V_j) \lambda(t,0,V_j) dt \,\Big|\, V_j \Big\} \\
\leq \quad & E\Big\{ S_{C|V}(s|V_j) \int_s^{\infty} e^{-\Lambda(t,0,V_j)} d\Lambda(t,0,V_j) \,\Big|\, V_j \Big\} \\
= \quad & E\{ S_{C|V}(s|V_j) e^{-\Lambda(s,0,V_j)} \mid V_j \} \;=\; E\{Y_1(s) \mid V_j\} \qquad\qquad (A.12)
\end{aligned}
$$

Also, observe that for $n \geq 2$ and $j \neq k$, the random variable $\sum_{i:\, i\neq j,k} Y_i(s)$ is distributed as $\text{Binom}(n-2,\, EY_1(s))$, and for a $\text{Binom}(m,a)$ random variable W,

$$
E(\frac{1}{1+W}) \;=\; \frac{1}{(m+1)a} \sum_{k=0}^{m} \frac{(m+1)!}{(k+1)!\,(m-k)!}\, a^{k+1}\, (1-a)^{m-k} \;\leq\; \frac{1}{(m+1)a}
$$

Therefore, with $m = n-2$ and $a = EY_1(s)$,

$$
E\{[1 + \sum_{i:i\neq j,k} Y_i(s)]^{-1}\} \;\leq\; \frac{1}{(n-1)EY_1(s)} \qquad\qquad (A.13)
$$

Therefore substituting (A.12) and (A.13) yields an upper bound for the sum of terms in (A.10) with $j \neq k$ as

$$
\begin{aligned}
& \frac{2}{n(n-1)}\, E \int_K^{\infty} \frac{1}{EY_1(s)} \Big[\sum_{j\neq k} E(Y_j(s)|V_j)\, Y_k(s)\, \lambda(s,0,V_k) \Big]\, ds \\
= \quad & \frac{2}{n(n-1)}\, E\Big[\int_K^{\infty} \sum_{j,k:\, j\neq k} Y_k(s)\, \lambda(s,0,V_k)\, ds \Big] \\
= \quad & 2\, EY_1(K) \;\to\; 0 \text{ as } K \uparrow \infty. \qquad\qquad (A.14)
\end{aligned}
$$

Thus from (A.7) to (A.14),

$$E\{A_3^2\} \;\to\; 0 \text{ as } K \uparrow \infty.$$

Finally, by (A.6) and convergence of A_1, A_2 and A_3, (i) is proved.

(ii) As shown in the proof of Lemma 2.3: when Assumptions 1.1, 1.3 and 1.4 hold, the process

$$M_i^*(t) \;\equiv\; N_i(t) - \int_0^s Y_i(s) E_0\{\lambda(s,0,V_i)\,|\,Y_i(s)\}$$

is a $\sigma\{N_i(s), Y_i(s), Z_i; 0 \le s \le t), i = 1, 2, \dots\}$ adapted martingale. Hence $U_L^{(K,\infty)}$ is an iid sum of martingales with mean 0. Therefore,

$$\begin{aligned}
& E_0\{[n^{-\frac{1}{2}} U_L^{(K,\infty)}]^2\} \\
= \;& n^{-1} \sum_{i=1}^n E_0 \left\{ \left[\int_K^\infty [Z_i - \mu(t)] \big[dN_i(t) - Y_i(t) E[\lambda(t,0,V)|Y(t)=1] dt \big] \right]^2 \right\} \\
= \;& n^{-1} \sum_{i=1}^n E_0 \left\{ \left[\int_K^\infty [Z_i - \mu(t)] dM_i^*(t) \right]^2 \right\} \\
= \;& E_0 \left\{ \int_K^\infty [Z_1 - \mu(t)]^2 \; d\langle M_1^* \rangle(t) \right\} \\
= \;& E_0 \left\{ \int_K^\infty [Z_1 - \mu(t)]^2 Y_1(t) E[\lambda(t,0,V_1)|Y_1(t)=1] dt \right\} \\
= \;& E_0 \left\{ \int_K^\infty [Z_1 - \mu(t)]^2 Y_1(t) E[\lambda(t,0,V_1)|Z_1, Y_1(t)=1] dt \right\} \\
= \;& E_0 \left\{ \int_K^\infty [Z_1 - \mu(t)]^2 Y_1(t) \lambda(t,0,V_1) dt \right\} \\
\le \;& E_0 \left\{ \int_K^\infty dN_1(t) \right\} \;\to\; 0 \text{ as } K \uparrow \infty.
\end{aligned}$$

Hence (ii) is also proved. □

Lemma A.8 *For $K > 0$, define*

$$n^{-\frac{1}{2}} \hat{U}_W^{(K,\infty)} \;\equiv\; n^{-\frac{1}{2}} \sum_w \sum_{i=1}^n \int_K^\infty \left\{ Z_i - \frac{\sum_{i=1}^n \xi_i^w Z_i Y_i(t)}{\sum_{i=1}^n \xi_i^w Y_i(t)} \right\} \cdot \xi_i^w dN_i(t)$$

and

$$n^{-\frac{1}{2}}U_W^{(K,\infty)} \equiv n^{-\frac{1}{2}}\sum_w\sum_i^n\int_K^\infty \xi_i^w[Z_i-\mu(t)][dN_i(t)-Y_i(t)\frac{E[\xi^w Y(t)\lambda(t,0,V)]}{E[\xi^w Y(t)]}dt].$$

Then under Assumptions 1.1, 1.3 and 1.4, we have for W a discrete finite-valued random variable,

$$(i)\ n^{-\frac{1}{2}}\hat{U}_W^{(K,\infty)} \overset{p,L_1}{\rightarrow} 0 \text{ as } K\uparrow\infty$$

$$(ii)\ n^{-\frac{1}{2}}U_W^{(K,\infty)} \overset{p,L_1}{\rightarrow} 0 \text{ as } K\uparrow\infty$$

uniformly over all n.

Proof.

This Lemma is a direct result of applying Lemma A.7 on each stratum with $W_i = w$.

□

Lemma A.9 *For $K > 0$, Define*

$$n^{-\frac{1}{2}}\hat{U}_\varphi^{(K,\infty)} \equiv n^{-\frac{1}{2}}\sum_{i=1}^n\int_K^\infty [Z_i - \frac{\sum_{j=1}^n \varphi(t,Z_j,V_j)Y_j(t)Z_j}{\sum_{j=1}^n \varphi(t,Z_j,V_j)Y_j(t)}]\varphi(t,Z_i,V_i)dN_i(t)$$

and

$$n^{-\frac{1}{2}}U_\varphi^{(K,\infty)} \equiv n^{-\frac{1}{2}}\sum_{i=1}^n\int_K^\infty [Z_i-\pi]\varphi(t,Z_i,V_i)\Big\{dN_i(t)-Y_i(t)\frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]}dt\Big\}$$

where V is discrete with finite value. By definition $\varphi(t,Z,V) = g(t,V)/S_C(t,Z,V)$ with $g(\cdot)$ satisfies Assumption 5.1. Further assume

$$\max_{v,w}\int \exp\{-\Lambda(t,0,w)\}\lambda(t,0,v)dt < \infty\ ,$$

$$\max_{v,w}\int \exp\{-\Lambda(t,0,w)/2\}\lambda(t,0,v)dt < \infty$$

$$\max_v \int \exp\{-\Lambda(t,0,v)\}\lambda^2(t,0,v)dt < \infty.$$

Then under Assumptions 1.1 and 1.3 ,

$$\textit{(i)}\ n^{-\frac{1}{2}}\hat{U}_{\varphi}^{(K,\infty)} \stackrel{p,L_1}{\longrightarrow} 0 \textit{ as } K \uparrow \infty$$

$$\textit{(ii)}\ n^{-\frac{1}{2}}\tilde{U}_{\varphi}^{(K,\infty)} \stackrel{p,L_1}{\longrightarrow} 0 \textit{ as } K \uparrow \infty$$

uniformly over all n.

Proof.

From (5.2) and (5.3) we have

$$\begin{aligned} S_1^{(K,\infty)} &\equiv n^{-\frac{1}{2}}\hat{U}_{\varphi}^{(K,\infty)} - n^{-\frac{1}{2}}\int_K^\infty \sum_{i=1}^n \varphi(t,Z_i,V_i)(Z_i-\pi)dN_i(t) \\ &= -n^{-\frac{1}{2}}\int_K^\infty \sum_{i=1}^n \varphi(t,Z_i,V_i)\Big(\frac{n^{-1}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big)dN_i(t). \end{aligned}$$

Hence

$$\begin{aligned} & E\{|S_1^{(K,\infty)}|\} \\ \leq\ & E\left\{n^{-\frac{1}{2}}\int_K^\infty \varphi(t,Z_i,V_i)Y_i(t)\Big|\frac{n^{-1}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big|\lambda(t,0,V_i)dt\right\} \\ \leq\ & E\left\{\sum_v \int_K^\infty \frac{\sum_{i:V_i=v}\varphi(t,Z_i,V_i)Y_i(t)}{\sum_j \varphi(t,Z_j,V_j)Y_j(t)}\Big|\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t)\Big|\lambda(t,0,v)dt\right\}. \end{aligned} \tag{A.15}$$

Let the non random constant C_0 be the number of levels of V. Since $\sum_j (Z_j - \pi)\varphi(t,Z_j,V_j)Y_j(t)$ is an iid sum, upon applying the Cauchy-Schwartz inequality to

(A.15) we have

$$
\begin{aligned}
& E\{|S_1^{(K,\infty)}|\} \\
\leq\ & C_0 \cdot \max_v \int_K^\infty \left[E\{(Z_1-\pi)\varphi(t, Z_1, V_1)Y_1(t)\}^2\right]^{\frac{1}{2}} \cdot \lambda(t,0,v)dt \\
\leq\ & C_0 \cdot \max_v \int_K^\infty \left[E\{\varphi(t, Z_1, V_1)Y_1(t)\}\right]^{\frac{1}{2}} \lambda(t,0,v)dt \\
\leq\ & C_0 \cdot \max_v \int_K^\infty E\{g(t,V)^{1/2} e^{-\frac{1}{2}\Lambda(t,0,V)}\} d\Lambda(t,0,v) \\
\leq\ & C_0 \cdot \max_{v,w} \int_K^\infty g(t,w)^{1/2} e^{-\frac{1}{2}\Lambda(t,w)} d\Lambda(t,0,v) \\
\leq\ & C_0 \cdot \max_{v,w} \int_K^\infty e^{-\frac{1}{2}\Lambda(t,0,w)} d\Lambda(t,0,v) \\
\rightarrow\ & \ 0 \text{ as } K \uparrow 0.
\end{aligned}
\tag{A.16}
$$

Furthermore, it is easy to check that

$$
S_2^{(K,\infty)} \equiv n^{-\frac{1}{2}} \sum_{i=1}^n \int_K^\infty \varphi(t, Z_i, V_i)(Z_i - \pi) dN_i(t)
$$

is a sum of iid random variables with mean 0, so we have

$$
\begin{aligned}
E\{|S_2^{(K,\infty)}|^2\} \ &\leq\ E\{\int_K^\infty (Z_1-\pi)^2 \varphi^2(t, Z_1, V_1) Y_1(t) \lambda(t,0,V_1)dt\} \\
&\leq\ E\{\int_K^\infty g(t,V_1) e^{-\Lambda(t,V_1)} d\Lambda(t,V_1)\} \\
&\leq\ \max_v \int_K^\infty g(t,v) e^{-\Lambda(t,v)} d\Lambda(t,v) \\
&\rightarrow\ \ 0 \text{ as } K \uparrow \infty.
\end{aligned}
\tag{A.17}
$$

Hence from (A.16) and (A.17) we obtain

$$
E\{|n^{-\frac{1}{2}} \hat{U}_\varphi^{(K,\infty)}|\} \leq E\{|S_1^{(K,\infty)}|\} + E\{|S_2^{(K,\infty)}|\} \ \rightarrow\ 0 \text{ as } K \uparrow \infty
$$

uniformly over all n. So (i) is proved.

Note that the second sum on the right hand side of

$$
n^{-\frac{1}{2}} \tilde{U}_\varphi^{(K,\infty)} \equiv S_2^{(K,\infty)} - n^{-\frac{1}{2}} \sum_{i=1}^n \int_K^\infty [Z_i - \pi] \varphi(t, Z_i, V_i) Y_i(t) \frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]} dt
$$

is also an iid sum with mean 0. By Cauchy-Schwartz inequality, we also have

$$
\begin{aligned}
\Big[E\{Y(t)\varphi(t,Z,V)\lambda(t,0,V)\}\Big]^2 &= \Big[E\{Y(t)\varphi(t,Z,V)Y(t)\lambda(t,0,V)\}\Big]^2 \\
&\le E\{Y^2(t)\varphi^2(t,Z,V)\}\cdot E\{Y^2(t)\lambda^2(t,0,V)\}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& E\Big\{\Big[n^{-\frac{1}{2}}\int_K^\infty \sum_{i=1}^n [Z_i-\pi]\varphi(t,Z_i,V_i)Y_i(t)\frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]}dt\Big]^2\Big\} \\
\le\ & E\Big\{\int_K^\infty [Z_1-\pi]^2\varphi^2(t,Z_1,V_1)Y_1^2(t)\Big(\frac{E[Y(t)\varphi(t,Z,V)\lambda(t,0,V)]}{E[Y(t)\varphi(t,Z,V)]}\Big)^2 dt\Big\} \\
\le\ & E\Big\{\int_K^\infty \varphi^2(t,Z_1,V_1)Y_1^2(t)\frac{E\{Y^2(t)\varphi^2(t,Z,V)\}\cdot E\{Y^2(t)\lambda^2(t,0,V)\}}{[E\{Y(t)\varphi(t,Z,V)\}]^2}dt\Big\} \\
\le\ & \int_K^\infty E\{\varphi(t,Z,V)Y(t)\}\frac{E\{Y(t)\varphi(t,Z,V)\}\cdot E\{Y^2(t)\lambda^2(t,0,V)\}}{[E\{Y(t)\varphi(t,Z,V)\}]^2}dt \\
=\ & \int_K^\infty E\{Y^2(t)\lambda^2(t,0,V)\}dt \\
\le\ & \int_K^\infty E\{S_C(t,Z,V)e^{-\Lambda(t,0,V)}\lambda^2(t,0,V)\}dt \\
\le\ & \max_v \int_K^\infty e^{-\Lambda(t,0,v)}\lambda^2(t,0,v)dt \\
\to\ & 0 \text{ as } K\uparrow 0 \qquad\qquad \text{(A.18)}
\end{aligned}
$$

Hence from (A.17) and (A.18) we have

$$
E\{|n^{-\frac{1}{2}}\tilde{U}_\varphi^{(K,\infty)}|\} \ \to\ 0 \text{ as } K\uparrow\infty.
$$

Thus (ii) is also proved. □

Lemma A.10 *Let iid random variables* (T_i, C_i), $i = 1, \ldots, n$ *be pairs of independent latent death- and censoring-time random variables.*

(i) Let $S(t)$ *and* $S_C(t)$ *be survival functions of* T_i *and* C_i *and let* $\hat{S}(t)$ *be the Kaplan-Meier estimator of* $S(t)$ *based on the right-censored survival data of all subjects,*

then

$$nE\{[\hat{S}(t)-S(t)]^2 I_{[\bar{Y}(t)\geq 1]}\} \;\leq\; \frac{1}{S(t)S_C(t)} \tag{A.19}$$

(ii)Let $\hat{S}_{(-j)}(t)$ be the Kaplan-Meier estimator of $S(t)$ based on the data of all except for the jth subject, then

$$I_{[\bar{Y}(t)\geq 2]}\left|\hat{S}(t)-\hat{S}_{(-j)}(t)\right|\Big/\hat{S}_{(-j)}(t) \;\leq\; 2/\bar{Y}_{(-j)}(t) \tag{A.20}$$

with probability one.

Proof. (i) Define a stopping time $\tau \equiv \inf\{t : \bar{Y}(t) = 1\}$. Then from Section IV.3.1 of Andersen et. al. (1992), we know the process $[\hat{S}(t) - S(t\wedge\tau)]/S(t\wedge\tau)$ is a martingale with probability 1, then

$$\begin{aligned}
nE\{[\hat{S}(t)-S(t)]^2 I_{[\tau\geq t]}\} &\leq E\{[\hat{S}(t\wedge\tau)-S(t\wedge\tau)]^2\} \\
&\leq S^2(t)E\int_0^{t\wedge\tau}\frac{\hat{S}^2(x-)}{S^2(x)}\frac{n\lambda(x)}{\bar{Y}(x)}dx \\
&\leq S^2(t)\int_0^t \frac{1}{S^2(x)}\frac{n\lambda(x)}{\bar{Y}(x)}I_{[\bar{Y}(x)\geq 1]}dx
\end{aligned} \tag{A.21}$$

Next we estimate a bound on $E\{I_{[\bar{Y}(t)\geq 1]}/\bar{Y}(t)\}$. Letting $M \equiv \bar{Y}(t) \sim \text{Binom}(n,p)$ with $p = S(t)S_C(t)$, we have

$$\begin{aligned}
E\{\frac{I_{[M\geq 1]}}{M}\} &= \sum_{k=1}^{n}\binom{n}{k}\frac{1}{k}p^k(1-p)^{n-k} \\
&\leq \frac{2}{(n+1)p}\sum_{k=1}^{n}\binom{n+1}{k+1}p^{k+1}(1-p)^{n-k} \\
&\leq \frac{2}{np}
\end{aligned} \tag{A.22}$$

Applying (A.22) into (A.21) we have,

$$\begin{aligned}
nE\{[\hat{S}(t)-S(t)]^2 I_{[\bar{Y}(t)\geq 1]}\} &\leq nS^2(t)\int_0^t E\{I_{[\bar{Y}(x)\geq 1]}/\bar{Y}(x)\}\frac{\lambda(x)}{S^2(x)}dx \\
&\leq n\frac{2S^2(t)}{n}\int_0^t \{S^3(x)S_C(x)\}^{-1}\lambda(x)dx \\
&\leq 2\frac{S^2(t)}{S_C(t)}\int_0^t \frac{\lambda(x)}{S^3(x)}dx \\
&\leq \frac{2}{3S(t)S_C(t)} \\
&\leq \frac{1}{S(t)S_C(t)}.
\end{aligned}$$

Hence (i) is proved.

(ii) Let $\tau_0 = \inf\{t : \Delta\bar{N}(t) = 1\}$ and $\tau_t = \sup\{s \leq t : \Delta\bar{N}(s) = 1\}$ and $X_j = T_j \wedge C_j$. Under the assumption that $\bar{Y}(t) \geq 2$, we can assert $0 < \tau_0 < \tau_t < t$. By definition,

$$\hat{S}(t) = \prod_{s\leq t}\{1-\frac{\Delta\bar{N}(s)}{\bar{Y}(s)}\},$$

$$\hat{S}_{(-j)}(t) = \prod_{s\leq t}\{1-\frac{\Delta\bar{N}_{(-j)}(s)}{\bar{Y}_{(-j)}(s)}\}.$$

Since $\bar{Y}(s) = \bar{Y}_{(-j)}(s)+1$ when $X_j \geq s$ and $\bar{Y}(s) = \bar{Y}_{(-j)}(s)$ when $X_j < s$, then

$$\begin{aligned}
I_{[X_j\leq t]}\hat{S}(t) &= I_{[X_j\leq t]}(1-\frac{1}{\bar{Y}(X_j)})\prod_{s<X_j}\frac{1-\Delta\bar{N}(s)/\bar{Y}(s)}{1-\Delta\bar{N}(s)/[\bar{Y}(s)-1]}\hat{S}_{(-j)}(t) \\
&\geq I_{[X_j\leq t]}(1-\frac{1}{\bar{Y}(X_j)})\hat{S}_{(-j)}(t) \\
&\geq I_{[X_j\leq t]}(1-\frac{1}{\bar{Y}(t)})\hat{S}_{(-j)}(t) \\
&\geq I_{[X_j\leq t]}(1-\frac{1}{\bar{Y}_{(-j)}(t)})\hat{S}_{(-j)}(t) \qquad \text{(A.23)}
\end{aligned}$$

and

$$
\begin{aligned}
I_{[X_j>t]}\hat{S}_{(-j)}(t) &= I_{[X_j>t]}\prod_{s\le t}\frac{1-\Delta\bar{N}(s)/[\bar{Y}(s)-1]}{1-\Delta\bar{N}(s)/\bar{Y}(s)}\hat{S}(t) \\
&= \frac{I_{[X_j>t]}}{1-1/\bar{Y}(\tau_0)} \times \frac{\prod_{s\le t}\left\{1-\Delta\bar{N}(s)/[\bar{Y}(s)-1]\right\}}{\prod_{\tau_0<s\le t}\left\{1-\Delta\bar{N}(s)/\bar{Y}(s)\right\} \times \{1-1/[\bar{Y}(\tau_t)-1]\}} \\
&\qquad \times \{1-1/[\bar{Y}(\tau_t)-1]\}\hat{S}(t) \\
&\ge I_{[X_j>t]}\frac{1}{1-1/\bar{Y}(\tau_0)}\{1-1/[\bar{Y}(t)-1]\}\hat{S}(t) \\
&\ge I_{[X_j>t]}\{1-1/[\bar{Y}(t)-1]\}\hat{S}(t). \qquad \text{(A.24)}
\end{aligned}
$$

Furthermore, by reasoning as in (A.24), with jump-times in $\hat{S}_{(-j)}(t)$ occurring immediately previous to those in $\hat{S}(t)$, and by inspection of the first line of (A.24), we also derive when $\bar{Y}(t) \ge 2$,

$$I_{[X_j\le t]}\hat{S}(t) \le I_{[X_j\le t]}\hat{S}_{(-j)}(t) \text{ and } I_{[X_j>t]}\hat{S}_{(-j)}(t) \le I_{[X_j>t]}\hat{S}(t).$$

Hence

$$
\begin{aligned}
& I_{[X_j\le t]}\left|\hat{S}(t)-\hat{S}_{(-j)}(t)\right| \\
&= I_{[X_j\le t]}\left\{\hat{S}_{(-j)}(t)-\hat{S}(t)\right\} \\
&\le I_{[X_j\le t]}\left\{1-1+\frac{1}{\bar{Y}_{(-j)}(t)}\right\}\hat{S}_{(-j)}(t) \\
&= I_{[X_j\le t]}\hat{S}_{(-j)}(t)/\bar{Y}_{(-j)}(t) \qquad \text{(A.25)}
\end{aligned}
$$

and

$$
\begin{aligned}
& I_{[X_j>t]}\Big|\hat{S}(t)-\hat{S}_{(-j)}(t)\Big| \\
= \;& I_{[X_j>t]}\Big\{\hat{S}(t)-\hat{S}_{(-j)}(t)\Big\} \\
\leq \;& I_{[X_j>t]}\left\{\frac{\bar{Y}(t)-1}{\bar{Y}(t)-2}-1\right\}\hat{S}_{(-j)}(t) \\
= \;& I_{[X_j>t]}\frac{1}{\bar{Y}_{(-j)}-1}\hat{S}_{(-j)}(t) \\
\leq \;& I_{[X_j>t]}2\hat{S}_{(-j)}(t)/\bar{Y}_{-j}(t). \qquad \text{(A.26)}
\end{aligned}
$$

Add (A.25) and (A.26)to get

$$
I_{[\bar{Y}(t)\geq 2]}\big|\hat{S}(t)-\hat{S}_{(-j)}(t)\big|\Big/\hat{S}_{(-j)}(t) \;\leq\; 2/\bar{Y}_{(-j)}(t).
$$

Hence (ii) is proved.

Lemma A.11 *When V is discrete with finite values and $\hat{\varphi}(t,z,v)$ is the Kaplan-Meier estimator for $\varphi(t,z,v)$ which is defined at Section 5.1. Further assume*

$$
\max_{z,v,w}\int_0^\infty \{S(t,0,w)/S_C(t,z,w)\}^{1/2}\lambda(t,0,v)dt<\infty, \qquad \text{(A.27)}
$$

then under Assumptions 1.1 and 1.3 , we have

$$
\text{(i) } n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{(K,\infty)} \overset{p,L_1}{\longrightarrow} 0 \text{ as } K\uparrow\infty
$$

$$
\text{(ii) } n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{(K,\infty)} \overset{p,L_1}{\longrightarrow} 0 \text{ as } K\uparrow\infty
$$

uniformly over all n.

proof.

From (5.6) and (5.7) we have

$$\begin{aligned} S_1^{*(K,\infty)} &\equiv n^{-\frac{1}{2}}\hat{U}_{\hat{\varphi}}^{(K,\infty)} - n^{-\frac{1}{2}}\int_K^\infty \sum_{i=1}^n \hat{\varphi}(t,Z_i,V_i)(Z_i-\pi)dN_i(t) \\ &= -n^{-\frac{1}{2}}\int_K^\infty \sum_{i=1}^n \hat{\varphi}(t,Z_i,V_i)\Big(\frac{n^{-1}\sum_j (Z_j-\pi)\hat{\varphi}(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \hat{\varphi}(t,Z_j,V_j)Y_j(t)}\Big)dN_i(t). \end{aligned}$$

Hence

$$\begin{aligned} & E\{|S_1^{*(K,\infty)}|\} \\ =\ & E\left\{n^{-\frac{1}{2}}\int_K^\infty \hat{\varphi}(t,Z_i,V_i)Y_i(t)\Big(\frac{n^{-1}\sum_j (Z_j-\pi)\hat{\varphi}(t,Z_j,V_j)Y_j(t))}{n^{-1}\sum_j \hat{\varphi}(t,Z_j,V_j)Y_j(t)}\Big)\lambda(t,0,V_i)dt\right\} \\ \le\ & E\left\{\sum_v \int_K^\infty \frac{\sum_{i:V_i=v}\hat{\varphi}(t,Z_i,V_i)Y_i(t)}{\sum_j \hat{\varphi}(t,Z_j,V_j)Y_j(t)}\Big|n^{-\frac{1}{2}}\sum_j (Z_j-\pi)\hat{\varphi}(t,Z_j,V_j)Y_j(t)\Big|\lambda(t,0,v)dt\right\} \\ \le\ & E\left\{\sum_v \int_K^\infty \Big|n^{-\frac{1}{2}}\sum_j (Z_j-\pi)\hat{\varphi}(t,Z_j,V_j)Y_j(t)\Big|\lambda(t,0,v)dt\right\}. \\ \le\ & E\left\{\sum_v \int_K^\infty n^{-\frac{1}{2}}\sum_j \Big|\hat{\varphi}(t,Z_j,V_j)-\varphi(t,Z_j,V_j)\Big|Y_j(t)\lambda(t,0,v)dt\right\} \\ & \qquad + E\left\{\sum_v \int_K^\infty \Big|n^{-\frac{1}{2}}\sum_j (Z_j-\pi)\varphi(t,Z_j,V_j)Y_j(t)\Big|\lambda(t,0,v)dt\right\} \qquad \text{(A.28)} \end{aligned}$$

The second term of (A.28) goes to 0 as $K\to\infty$ as shown in the proof of Lemma A.9. Hence it is sufficient to prove $E\{|S_1^{*(K,\infty)}|\}\to 0$ by showing

$$E\left\{\sum_v \int_K^\infty n^{-\frac{1}{2}}\sum_j \Big|\hat{\varphi}(t,Z_j,V_j)-\varphi(t,Z_j,V_j)\Big|Y_j(t)\lambda(t,0,v)dt\right\} \ \to\ 0 \quad \text{(A.29)}$$

as $K \uparrow \infty$. Furthermore, by Assumption 5.1,

$$
\begin{aligned}
& |\hat{\varphi}(t, Z_j, V_j) - \varphi(t, Z_j, V_j)| \\
= \;& \left| \frac{1}{S_C(t, Z_j, V_j)} \{\hat{g}(t, V_j) - g(t, V_j)\} \; - \; \frac{\hat{g}(t, Z_j, V_j)}{\hat{S}_C(t, Z_j, V_j)} \{ \frac{\hat{S}_C(t, Z_j, V_j)}{S_C(t, Z_j, V_j)} - 1\} \right| \\
\leq \;& \frac{1}{S_C(t, Z_j, V_j)} \, |\hat{g}(t, V_j) - g(t, V_j)| \; + \; \left| \frac{\hat{S}_C(t, Z_j, V_j)}{S_C(t, Z_j, V_j)} - 1 \right| \\
\leq \;& \frac{1}{S_C(t, Z_j, V_j)} \left\{ c_1 \left| \hat{S}_C(t, 1, V_j) - S_C(t, 1, V_j) \right| \; + \; c_0 \left| \hat{S}_C(t, 0, V_j) - S_C(t, 0, V_j) \right| \right\} \\
& \qquad + \frac{1}{S_C(t, Z_j, V_j)} \left| \hat{S}_C(t, Z_j, V_j) - S_C(t, Z_j, V_j) \right| \qquad \text{(A.30)}
\end{aligned}
$$

for some constant c_1 and c_0. By Lemma A.10 (i), for $n \geq 2$

$$
\begin{aligned}
& \sqrt{n} E \left| \hat{S}_{C,(-1)}(t, z, v) - S_C(t, z, v) \right| I_{[\bar{Y}_{(-1)}(t) \geq 1]} \\
= \;& \frac{\sqrt{n}}{\sqrt{n-1}} \sqrt{n-1} \; E \left| \hat{S}_{C,(-1)}(t, z, v) - S_C(t, z, v) \right| I_{[\bar{Y}_{(-1)}(t) \geq 1]} \\
\leq \;& \sqrt{2} \, \left\{ (n-1) E \{ \hat{S}_{C,(-1)}(t, z, v) - S_C(t, z, v) \}^2 \right\}^{1/2} I_{[\bar{Y}_{(-1)}(t) \geq 1]} \\
\leq \;& \sqrt{2} / \sqrt{S_C(t, z, v) S(t, 0, v)}. \qquad \text{(A.31)}
\end{aligned}
$$

By Lemma A.10 (ii),

$$
\begin{aligned}
& \sqrt{n} \left| \hat{S}_{C,(-1)}(t, z, v) - \hat{S}_C(t, z, v) \right| I_{[\bar{Y}(t) \geq 2]} \\
\leq \;& 2 \sqrt{n} I_{[\bar{Y}(t) \geq 2]} \hat{S}_{C,(-1)}(t, z, v) / \bar{Y}_{(-1)}(t) \\
\leq \;& \frac{2\sqrt{n}}{\bar{Y}_{(-1)}(t)} \left| \hat{S}_{C,(-1)}(t, z, v) - S_C(t, z, v) \right| I_{[\bar{Y}(t) \geq 2]} \; + \; 2 \frac{\sqrt{n} S_C(t, z, v)}{\bar{Y}_{(-1)}(t)} I_{[\bar{Y}(t) \geq 2]} \\
\leq \;& 2\sqrt{n} \left| \hat{S}_{C,(-1)}(t, z, v) - S_C(t, z, v) \right| I_{[\bar{Y}(t) \geq 2]} \; + \; 2 \frac{\sqrt{n} S_C(t, z, v)}{\bar{Y}_{(-1)}(t)} I_{[\bar{Y}(t) \geq 2]}.
\end{aligned}
$$

(A.32)

Furthermore, using (A.31) and (A.32) we derive the following,

$$
\begin{aligned}
& E\left\{\sum_v \int_K^\infty n^{-\frac{1}{2}} \sum_j \frac{1}{S_C(t,Z_j,V_j)} \left|\hat{S}_C(t,Z_j,V_j) - S_C(t,Z_j,V_j)\right| Y_j(t)\lambda(t,0,v)dt\right\} \\
= \;& E\left\{\sum_v \int_K^\infty \frac{\sqrt{n}}{S_C(t,Z_1,V_1)} \left|\hat{S}_C(t,Z_1,V_1) - S_C(t,Z_1,V_1)\right| Y_1(t)\lambda(t,0,v)dt\right\} \\
\leq \;& E\left\{\sum_v \int_K^\infty \frac{\sqrt{n}}{S_C(t,Z_1,V_1)} \left|\hat{S}_{C,(-1)}(t,Z_1,V_1) - S_C(t,Z_1,V_1)\right| Y_1(t)\lambda(t,0,v)dt\right\} \\
& + E\left\{\sum_v \int_K^\infty \frac{\sqrt{n}}{S_C(t,Z_1,V_1)} \left|\hat{S}_{C,(-1)}(t,Z_1,V_1) - \hat{S}_C(t,Z_1,V_1)\right| Y_1(t)\lambda(t,0,v)dt\right\} \\
\leq \;& E\left\{\sum_v \int_K^\infty \frac{3\sqrt{n}}{S_C(t,Z_1,V_1)} \left|\hat{S}_{C,(-1)}(t,Z_1,V_1) - S_C(t,Z_1,V_1)\right| Y_1(t)\lambda(t,0,v)dt\right\} \\
& + E\left\{\sum_v \int_K^\infty I_{[\bar{Y}_{(-1)}(t)\geq 2]} \frac{\sqrt{n}}{S_C(t,Z_1,V_1)} \frac{2S_C(t,Z_1,V_1)}{\bar{Y}_{(-1)}(t)} Y_1(t)\lambda(t,0,v)dt\right\} \\
\leq \;& E\left\{\sum_v \int_K^\infty 3\sqrt{n} \left|\hat{S}_{C,(-1)}(t,Z_1,V_1) - S_C(t,Z_1,V_1)\right| S(t,0,V_1)\lambda(t,0,v)dt\right\} \\
& + E\left\{(4 \wedge \sqrt{n}) \sum_v \int_K^\infty S(t,0,V_1)\lambda(t,0,v)dt\right\} \\
\leq \;& E\left\{3\sqrt{2} \sum_v \int_K^\infty \{1/S(t,0,V_1)S_C(t,Z_1,V_1)\}^{1/2} S(t,0,V_1)\lambda(t,0,v)dt\right\} \\
& + E\left\{(4 \wedge \sqrt{n}) \sum_v \int_K^\infty S(t,0,V_1)\lambda(t,0,v)dt\right\} \\
\leq \;& 6C_0 \max_{z,v,w} \left\{\int_K^\infty \{S(t,0,w)/S_C(t,z,w)\}^{1/2}\lambda(t,0,v)dt + \int_K^\infty S(t,0,w)\lambda(t,0,v)dt\right\}
\end{aligned}
\tag{A.33}
$$

Thus from (A.28) to (A.33) we can derive

$$
S_1^{*(K,\infty)} \xrightarrow{L_1} 0 \text{ as } K \uparrow \infty. \tag{A.34}
$$

Next consider $S_2^{*(K,\infty)}$. It is clear that

$$
\begin{aligned}
& E|S_2^{*(K,\infty)}| \\
\equiv \;\; & n^{-\frac{1}{2}} E \left| \sum_{i=1}^{n} \int_K^\infty \hat{\varphi}(t, Z_i, V_i)(Z_i - \pi) dN_i(t) \right| \\
= \;\; & n^{-\frac{1}{2}} E \left| \sum_{i=1}^{n} \int_K^\infty \{\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)\}(Z_i - \pi) dN_i(t) \; + \; S_2^{(K,\infty)} \right| \\
\leq \;\; & n^{-\frac{1}{2}} E \sum_{i=1}^{n} \int_K^\infty |\hat{\varphi}(t, Z_i, V_i) - \varphi(t, Z_i, V_i)| \, Y_i(t) \lambda(t, 0, V_i) dt \; + \; E|S_2^{(K,\infty)}| \\
\leq \;\; & E \left\{ \int_K^\infty \sqrt{n} \, |\hat{\varphi}(t, Z_1, V_1) - \varphi(t, Z_1, V_1)| \, Y_1(t) \lambda(t, 0, V_1) dt \right\} \; + \; E|S_2^{(K,\infty)}|
\end{aligned}
\tag{A.35}
$$

From the proof of Lemma A.9 we know $E|S_2^{(K,\infty)}| \to 0$ as $K \uparrow \infty$. By applying the similar technique of obtaining (A.34) on the first term of (A.35), it is sufficient to show that

$$S_2^{*(K,\infty)} \stackrel{L_1}{\to} 0 \text{ as } K \uparrow \infty. \tag{A.36}$$

Finally

$$n^{-\frac{1}{2}} \hat{U}_{\hat{\varphi}}^{(K,\infty)} \equiv S_1^{*(K,\infty)} + S_2^{*(K,\infty)} \stackrel{L_1}{\to} 0$$

as $K \uparrow \infty$. So (i) is proved.

Next consider $\tilde{U}_{\hat{\varphi}}^{(K,\infty)}$. By Definition,

$$n^{-\frac{1}{2}} \tilde{U}_{\hat{\varphi}}^{(K,\infty)} \equiv S_2^{*(K,\infty)} - n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_K^\infty [Z_i - \pi] \hat{\varphi}(t, Z_i, V_i) Y_i(t) \frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]} dt.$$

Since V is discrete with finite values,

$$\frac{E[Y(t)\varphi(t, Z, V)\lambda(t, 0, V)]}{E[Y(t)\varphi(t, Z, V)]} \leq \max_v \lambda(t, 0, v) \leq \sum_v \lambda(t, 0, v).$$

Then

$$\begin{aligned}
& E\left|n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{(K,\infty)} - n^{-\frac{1}{2}}\tilde{U}_{\varphi}^{(K,\infty)}\right| \\
\leq\ & E\left|S_2^{*(K,\infty)} - S_2^{(K,\infty)}\right| \\
& + n^{-\frac{1}{2}} E\left\{\sum_{i=1}^{n}\int_K^\infty |\hat{\varphi}(t,Z_i,V_i) - \varphi(t,Z_i,V_i)|\, Y_i(t)\sum_v \lambda(t,0,v)dt\right\} \\
\leq\ & E\left|S_2^{*(K,\infty)} - S_2^{(K,\infty)}\right| \\
& + C_0 \max_v E\left\{\sqrt{n}\int_K^\infty |\hat{\varphi}(t,Z_1,V_1) - \varphi(t,Z_1,V_1)|\, Y_1(t)\lambda(t,0,v)dt\right\} \quad \text{(A.37)}
\end{aligned}$$

When $K \uparrow \infty$, the first term of (A.37) goes to 0. By the similar reasoning of using (A.29),(A.30) and (A.33) to derive (A.34), the second term of (A.37) also go to 0. Finally by Lemma A.9,

$$n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{(K,\infty)} = \{n^{-\frac{1}{2}}\tilde{U}_{\hat{\varphi}}^{(K,\infty)} - n^{-\frac{1}{2}}\tilde{U}_{\varphi}^{(K,\infty)}\} \ + \ n^{-\frac{1}{2}}\tilde{U}_{\varphi}^{(K,\infty)} \stackrel{L_1}{\to} 0$$

as $K \uparrow 0$. so (ii) is also proved. $\square$

Lemma A.12 *Let (X_i, V_i), $i = 1, \ldots, n$, be iid vectors of random variables. Assumes that for each i, there exist*

$$0 \leq u_i = \sup_{0 \leq t < \infty} |g(t, V_i)| < \infty, \ \textit{with}\ E(u_i) < \infty$$

and $\{g(\cdot, v)\}_v$ are uniformly equicontinuous. Then

$$\sup_{0\leq t<\infty} |\frac{1}{n}\sum_{i=1}^{n}\{I_{[X_i \geq t]}g(t,V_i)\} - E\{I_{[X_1\geq t]}g(t,V_1)\}| \stackrel{P,L(\Omega)}{\longrightarrow} 0$$

Proof:

Since $E(u_i) < \infty$ and $u_i \cdot I_{[X_i \geq t]} \stackrel{a.s.}{\to} 0$ as $t \uparrow \infty$,

by the Dominated convergence theorem, as $t^* \to \infty$

$$E[\sup_{t>t^*} |I_{[X_i\geq t]}g(t,V_i)|] \to 0. \quad \text{(A.38)}$$

For $j = 1, 2, \ldots, m$ and $0 < t^* < \infty$, define $t_j \equiv \frac{j}{m} t^*$ and $g^*_{m,t^*}(s, V) \equiv g(t_j, V)$ for $t_j \le s < t_{j+1}$. Thus, for any $s \in [0, t^*]$, we have

$$|g(s, V) - g^*_{m,t^*}(s, V)| \le \sup_{x,y:|x-y|\le \frac{t^*}{m}} |g(x, V) - g(y, V)| \tag{A.39}$$

For any $\epsilon > 0$, there exist $0 < t^* < \infty$ and m, such that

$$E[\sup_{t>t*} |I_{[X_i \ge t]} g(t, V)|] < \frac{\epsilon}{2}$$

and

$$\sup_{x,y\in[0,t^*],|x-y|\le\frac{t^*}{m}} |g(x, V) - g(y, V)| < \frac{\epsilon}{6}.$$

Then from (A.39), we have

$$E[\sup_{t\in[0,t^*]} |g(t, V) - g^*_{m,t^*}(t, V)|] \le \frac{\epsilon}{6}.$$

Then From Lemma A.3 we know that there exist a large integer $N > 0$, such that for any $n \ge N$,

$$\begin{aligned}
& E\Big[\sup_{t\in[0,t^*]} \Big|\frac{1}{n}\sum_{i=1}^{n} g^*_{m,t^*}(t, V_i) \cdot I_{[X_i \ge t]} \;-\; E[g^*_{m,t^*}(t, V) \cdot I_{[X>t]}]\Big|\Big] \\
= \quad & \max_{0\le j\le m<\infty} E\Big[\sup_{t_j\le t<t_{j+1}} \Big|\frac{1}{n}\sum_{i=1}^{n} g(t_j, V_i) I_{[X_i\ge t]} - E[g(t_j, V) I_{[X\ge t]}]\Big|\Big] \\
< \quad & \frac{\epsilon}{6}.
\end{aligned}$$

Thus,

$$\begin{aligned}
& E\Big[\sup_{t\in[0,t^*]}\Big|\frac{1}{n}\sum_{i=1}^{n} g(t,V_i)I_{[X_i\geq t]} - E[g(t,V)I_{[X\geq t]}]\Big|\Big] \\
\leq\ & E\Big[\sup_{t\in[0,t^*]}\Big|\frac{1}{n}\sum_{i=1}^{n} g(t,V_i)I_{[X_i\geq t]} - \frac{1}{n}\sum_{i=1}^{n} g^*_{m,t^*}(t,V_i)I_{[X_i\geq t]}\Big|\Big] \\
& + E\Big[\sup_{t\in[0,t^*]}\Big|\frac{1}{n}\sum_{i=1}^{n} g^*_{m,t^*}(t,V_i)I_{[X_i\geq t]} - E[g^*_{m,t^*}(t,V)I_{[X\geq t]}]\Big| \\
& + \Big|E[g(t,V)I_{[X\geq t]}]\ -\ E[g^*m,t*(t,V)I)[X\geq t]]\Big| \\
& <\ \frac{\epsilon}{6}+\frac{\epsilon}{6}+\frac{\epsilon}{6} \\
& =\ \frac{\epsilon}{2}.
\end{aligned} \tag{A.40}$$

Thus, from (A.40) and (A.40), we have

$$\sup_{0\leq t<\infty}|\frac{1}{n}\sum_{i=1}^{n}\{I_{[X_i\leq t]}g(t,V_i)\} - E\{I_{[X_1\leq t]}g(t,V_1)\}| \stackrel{P,L(\Omega)}{\longrightarrow} 0$$

This lemma is proved. $\square$

Corollary A.1 *If $E(u_i^2)\leq\infty$ and in addition to the hypotheses of Lemma A.12, then*

$$\sup_{0\leq t<\infty}|\frac{1}{n}\sum_{i=1}^{n}\{I_{[X_i\geq t]}g(t,V_i)\} - E\{I_{[X_1\geq t]}g(t,V_1)\}| \stackrel{P,L^2(\Omega)}{\longrightarrow} 0$$

Proof.

The proof of this corollary is analogous to the proof of Lemma A.12. $\square$

Lemma A.13 *Let N be a counting process and $M=N-A$ be the corresponding local square integrable martingale. If H_n is a predicable and locally bounded process* *with* $\sup_{0\le t<\infty}|H_n(t)| \xrightarrow{p} 0$, *then for any bounded stopping time* τ,

$$\textit{(i)}\ \sup_{0\le t<\tau}\Big|\int_0^t H_n(s)dM(s)\Big| \xrightarrow{p}, \textit{as } n\to\infty;$$

$$\textit{(ii)}\ \sup_{0\le t<\tau}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t H_n(s)dM_i^{(n)}(s)\Big| \xrightarrow{p} 0, \textit{as } n\to\infty,$$

where $M_i^{(n)} = N_i^{(n)} - A_i^{(n)}$, $i=1,2,\dots,n$ *are iid realizations of* M.

proof.

(i) By Lenglart's inequality and Corollary 3.4.1 of Fleming and Harrington (1991), for any $\epsilon,\eta>0$,

$$P\Big\{\sup_{0\le t<\tau}\Big|\int_0^t H_n(s)dM(s)\Big|\ge\epsilon\Big\} \le \frac{\eta}{\epsilon^2} + P\Big\{\int_0^\tau H_n^2(s)d\langle M,M\rangle(s)\ge\eta\Big\}, \quad \text{(A.41)}$$

where $\langle M,M\rangle = A$ is nonegative and monotone increasing with $E(A)=E(N)<\infty$ for all $t\in[0,\tau]$. Since $\sup_{0\le t<\infty}|H_n(t)| \xrightarrow{p} 0$, then for any $\epsilon>0$ and $\eta<\epsilon^3/2$, there exists an integer $N>0$, such that, for any $n\ge N$,

$$0\le P\Big\{\int_0^\tau H_n^2(s)d\langle M,M\rangle(s)\ge\eta\Big\} < \frac{\epsilon^3-\eta}{\epsilon^2}$$

By (A.41)

$$P\Big\{\sup_{0\le t\le\tau}\Big|\int_0^t H_n(s)dM(s)\Big|\ge\epsilon\Big\} < \frac{\eta}{\epsilon^2} + \frac{\epsilon^3-\eta}{\epsilon^2} \;=\; \epsilon$$

Hence (i) is proved.

(ii) For each n,

$$M^{(n)}(t) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t H_n(s)dM_i^{(n)}(s)$$

is an $\mathcal{F}_t$-martingale with predictable-variance process

$$\langle M^{(n)}, M^{(n)}\rangle(t) = \sum_{i=1}^n\int_0^t \frac{1}{n}H_n^2(s)dA_i^{(n)}(s).$$

As in (A.41),

$$P\{\sup_{0\le t\le\tau}|M^{(n)}(t)|\ge\epsilon\}\le\frac{\eta}{\epsilon^2}+P\{\int_0^\tau H_n^2(s)\frac{1}{n}\sum_{i=1}^n dA_i^{(n)}(s)>\eta\} \tag{A.42}$$

Since $\sup_{0\le t<\infty}|H_n(t)|\xrightarrow{p}0$ and

$$\frac{1}{n}\sum_{i=1}^n A_i^{(n)}(t)\le\frac{1}{n}\sum_{i=1}^n A_i^{(n)}(\tau)\xrightarrow{p}E[A_1^{(n)}(\tau)]$$

then

$$0<[\sup_{0\le t\le\tau}H_n^2(t)]\cdot\frac{1}{n}\sum_{i=1}^n A_i(\tau)\xrightarrow{p}0,\text{as } n\to\infty,$$

hence

$$\int_0^\tau H_n^2(s)\frac{1}{n}\sum_{i=1}^n dA_i^{(n)}(s)\xrightarrow{p}0,\text{as }\to\infty.$$

Thus for any $\epsilon,\eta>0$ and $\eta<<\epsilon$, there exist an integer $N>0$, such that for any $n\ge N$, for the bounded stopping-time τ,

$$P[\int_0^\tau H_n^2(s)\frac{1}{n}\sum_{i=1}^n dA_i^{(n)}(s)>\eta]<\frac{\epsilon^3-\eta}{\epsilon^2},$$

hence

$$P[\sup_{0\le t\le\tau}|M^{(n)}(t)|\ge\epsilon]<\epsilon.$$

and (ii) is proved, since

$$\sup_{0\le t<\tau}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^t H_n(s)dM_i^{(n)}(s)\right|\xrightarrow{p}0,\text{as } n\to\infty,$$

□

Lemma A.14 *Write $\theta=b/\sqrt{n}$, and assume for all t,V that $\int_0^t\lambda(s,0,V)ds<\infty$, and also that, on each interval $[0,t]$, the function $\lambda'(s,\theta,V)$ are uniformly integrable,* *with respect to ds, over θ in a sufficiently small neighborhood of 0. Then for any integrable random variable of the form $K(V,Z)$,*

$$E_{\theta_n}\{Y(t)K(V,Z)\} = E_0\{Y(t)k(V,Z)\} - \frac{b}{\sqrt{n}}E_0\Big\{Y(t)k(V,Z)\int_0^t \lambda'(s,0,V)ds\Big\} + o(n^{-\frac{1}{2}}) \tag{A.43}$$

where the expectation is taken with respect to the true model. As a consequence, for integrable $k(V)$,

$$n^{\frac{1}{n}}E_{\theta_n}\{Y(t)[Z-\mu(t)]k(V)\} = -b\sigma_Z^2 E_0\Big\{Y(t)k(V)\int_0^t \lambda'(s,0,V)ds\Big\} + o(1) \tag{A.44}$$

This is Lemma A·1 of Kong and Slud (1997). □

Bibliography

[1] Aalen, O. (1978). Nonparametric inference for a family of counting process, *Ann. Statis.* **6** 701-726

[2] Aalen, O. O. (1980). A model for nonparametric regression analysis of counting process, *Lecture Notes in Statistics***2**, 1-25. New York: Springer-Verlag

[3] Anderson, P. K., Borgan, Ø., Gill, R. D. and Keiding, N., (1992), *Statistical Models Based on Counting Processes*, Springer, New York

[4] Casella, G. and Berger, R. L., (2002), *Statistical Inference*, Duxbury

[5] Cox, D. R.(1972). Regression models and life tables (with discussion), *J. R. Statist. Soc.* B **34**, 187-219

[6] DiRienzo, A. G. and Lagakos, S. W. (2001a), Bias correctioin for score tests arising from misspecified proportional hazards regression models, *Biometrika* **88**, 421-434

[7] DiRienzo, A. G. and Lagakos, S. W. (2001b), Effects of model misspecification on tests of no randomized treatment effect arising from Cox's Proportional hazards model, *J. R. Statist. Soc.* B **63**, 745-757

[8] Fleming, T.R. and Harrington, D. P., (1991), *Counting processes and survival analysis*, New York: Wiley

[9] Gill, R. D. (1980), Censoring and stochastic integrals. *Mathematical centre tracts*,**124**, Amsterdam: Mathematisch Centrum

[10] Green, S. B. and Byar, D. P. (1978), The effect of stratified randomization on size and power of statistical tests in clinical trials, *J. Chronic. Dis.*, **31**, 445-454

[11] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457-481

[12] Kalbfleisch, J. D. and Prentice, R.L. (1980) *Statistical Analysis of Failure Time Data.* New York: Wiley

[13] Kong, F. H. and Slud, E. V. (1997). Robust covariate-adjusted log-rank tests *Biometrika*, **84**, 847-862

[14] Lagakos, S. W. and Schoenfeld, D. A. (1984). Properties of Proportional-Hazards Score Tests under Misspecified Regression Models *Biometrics*, **40**, 1037-1048

[15] Lawless J. F.(1982), *Statistical Models and Methods for Lifetime Data*. New York: Wiley

[16] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemonther*, Rep. **50**, 163-170

[17] Morgan, T. (1986) Omitting covariates from the proportional hazards model. *Biometrics*, **42**, 993-995

[18] Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64**, 441-448

[19] McKeague, I. W. (1988), Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *Statistical inference from stochastic processes* **80**,139-152

[20] Peto, R. and Peto, J. (1972) Asymptotic efficient rank distribution theory for Cox-type regreesion models with general relative risk form. *Ann. Statist.* , **11**, 804-813

[21] Schoenfeld, D. (1981), The asymptotic properties of rank tests for the censored two-sample problem. *Biometrika*, **68**, 316-319

[22] Schoenfeld, D. and Tsiatis, A. A., (1987), A modified log rank test for highly stratified data, *Biometrika*, **74**, 167-175

[23] Slud, E. V. and Rubinstein, L.(1983), Dependent competing risks and summary survival curves, *Biometrika*,**70**, 643-649

[24] Slud, E. V. (1991), Relative efficiency of the log rank test within multiplcative intensity model, *Biometrika*, **78**, 621-630

[25] Segall, A. and Kailath, T. (1975). The modeling of randomly modulated jump processes. *IEEE Trans. Info. Theory*, **IT-21**, 135-143

[26] Tsiatis, A. (1975) A nonidentifiability aspect of the problem of Competing Risks, *Proc. Nat. Acad. Sci. USA*, **72** 20-22

[27] van der Varrt, A. W. (1998) *Asymptotic Statistics*, Cambridge University Press