University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2004 *Paper* 152

Semiparametric Quantitative-Trait-Locus Mapping: II. on Censored Age-at-Onset

Ying Qing Chen* Chengcheng Hu†

Rongling Wu‡

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Dept. of Biostatistics, School of Public Health, Harvard University, Boston, MA, cchu@hsph.harvard.edu

‡Dept. of Statistics, Institute of Food & Agriculture Sciences, University of Florida, Gainesville, FL, rwu@hes.hmc.psu.edu

Semiparametric Quantitative-Trait-Locus Mapping: II. on Censored Age-at-Onset

Ying Qing Chen, Chengcheng Hu, and Rongling Wu

Abstract

In genetic studies, the variation in genotypes may not only affect different inheritance patterns in qualitative traits, but may also affect the age-at-onset as quantitative trait. In this article, we use standard cross designs, such as backcross or F2, to propose some hazard regression models, namely, the additive hazards model in quantitative trait loci mapping for age-at-onset, although the developed method can be extended to more complex designs. With additive invariance of the additive hazards models in mixture probabilities, we develop flexible semiparametric methodologies in interval regression mapping without heavy computing burden. A recently developed multiple comparison procedures is adapted to identify the QTL in dense maps. The proposed methodologies will be evaluated by simulation studies and demonstrated in an actual data analysis of forest tree growth.

1 INTRODUCTION

Genetic mapping has long been a major approach for geneticists to study and locate the chromosomal regions that may link to the patterns of inheritance of a trait. With new development in recombinant DNA, it is possible for genetic mapping to further isolate a gene by positional cloning (Bender, Spierer and Hogness, 1983). Several hundreds of rare human diseases of simple Mendelian inheritance have been mapped and dozens of them have been positionally cloned. More complex traits that do not follow Mendelian monogenic inheritance, such as diabetes, cancer and Parkinson's disease, have been considered for the genetic mapping as well. Age-at-onset, or time-to-event in general, as an inheritable quantitative trait, usually does not follow the classical Mendelian inheritance paradigm, due to various reasons, such as gene-gene interactions, gene-environment interactions or random chance. It has been nevertheless of important scientific interest. For example, a total of 3,796 individuals in 263 prostate cancer families were analyzed in Conlon, et al. (2003), and there was for two to three QTLs to contribute to the variation in the age-at-onset of hereditary prostate cancer.

To map complex trait such as the age-at-onset, there are usually four types of epidemiological or experimental ways: linkage analysis, allele-sharing methods, association studies and genetic analysis of experimental crosses, as summarized in Lander and Schork (1994). Among them, the method of experimental crosses is relatively more powerful in the QTL mapping of complex traits than three other methods, due to its ability to control the non-genetic noise. Although experimental crosses are often done in animal and plant studies, they are able to identify the key genes and help understand the possible biochemical pathway in a disease, when the biochemical pathologies are similar or the same. Another advantage of the experimental crosses is their easy adaptability of genomic information into the QTL mapping. With the interval mapping method and its variants (Lander and Botstein, 1989), the QTL mapping can be conducted in whole genome to search for possible QTLs with available phenotypic and genetic marker data.

The statistical methods in QTL mapping, even though they have been developed rapidly with recent development in DNA-based genetic linkage maps, has been largely based on parametric methods with normal theory, since the seminal work of Lander and Botstein (1989). Consider a prototype version of their framework. Let Y be the measurement of the phenotype of interest and G be the indicator of causal genotype, respectively. The following multiple linear regression model

is often considered,

$$Y = \beta_0 + \beta_G G + e, \tag{1}$$

where e are normal deviates with mean 0 and variance σ^2, β_0 and β_G are the unknown parameters. Here, $\beta_G \neq 0$ means potential existence of QTL. Then the so-called LOD-score profile can be obtained by way of the maximum likelihood and plotted against the genetic distance of makers to locate the potential QTLs. In practice, since the exact genotypes are unknown, the normal mixture models, along with the EM-algorithm, are used according to the specific design of experimental crosses (Lincoln, Daly and Lander, 1993). The normality assumption, however if violated, may lead to false QTL detection (Morton, 1984), although it greatly simplifies the EM-algorithm used.

Apparently the age-at-onset, which is mostly positively distributed, usually does not satisfy the normal assumptions. Although special transformations, such as the log-transformation or the Box-Cox transformation, can be applied to symmetrize the distributions, they often tend to be arbitrarily chosen and sometimes the interpretation of parameters is not clear. More importantly, the age-at-onset may be often subject to censoring, for which it is not fully observed due to reasons, such as limited observation period. Special methods are thus needed in the QTL mapping of censored age-at-onset, as those in conventional survival analysis.

One of the most widely used models for time-to-event is the semiparametric Cox proportional hazards model (Cox, 1972). If it is applied to the age-at-onset following the similar fashion of the aforementioned multiple linear regression model, that would be

$$\lambda(t \mid G) = \lambda_0(t) \exp(\beta_G G), \tag{2}$$

where $\lambda(\cdot)$ are the hazard functions and $\lambda_0(\cdot)$ is usually unknown. In this model, β_G is the parameter for the relative hazards between two genotypes at one locus. If $\beta_G = 0$, it means no difference in hazard functions due to variation in genotypes, whereas it implies a possible QTL if $\beta_G \neq 0$. This model has been fairly successful in the usual survival analysis. However, it may not have significant advantage when it becomes mixed due to unknown G. If the model is marginalized over all possible G's, then it is well known that the proportionality in the hazard functions would not be preserved. As a result, the simple form of estimation by way of maximum partial likelihood would not work well to eliminate the infinite-dimensional baseline hazard function as nuisance parameter. If the maximum likelihood method has to be implemented onto the full likelihood function as in the multiple linear regression models, then the baseline hazard functions have to be explicitly specified

in some parametric form. However, the EM-algorithm may not have the clean forms any longer as in the normal mixture except for some special distributions, such as the exponential distributions.

In this article, we will develop a new methodology in the semiparametric QTL mapping of censored age-at-onset by applying the additive hazards model instead. The new methodology will take the full advantage of the simple invariance property of the mixed additive hazards model in hazards additivity. They will greatly relieve computing burden and preserve the appealing interpretation of relative hazards in the parameters. In the rest of this article, we will first discuss the genetic designs in experimental crosses. Then the new methodology and relevant theory are studied. A recently developed multiple comparison procedures is adapted to identify the potential QTLs based on the parameter estimates. Numerical studies including simulations and an application to actual data are in §3. Some issues are discussed in §4. The technical proofs of asymptotic properties are collected in the Appendix.

2 METHODS AND THEORY

2.1 Genetic designs

In this article, we mainly focus on the genetic designs with inbreed experimental crosses. Unlike natural population such as human families, these designs exercise more control on the nongenetic noise and individual unobserved heterogeneity of the genetic materials, and hence the difference in the quantitative trait of age-at-onset, if there is, may be mostly likely caused by possible genetic factors. There are two important designs of inbreed experimental crosses for the QTL mapping, namely, the backcross and F_2 designs.

Both of the designs are initiated with two contrasting homozygous inbred lines, that is, their paternal (A) and the maternal (a) alleles are identical at any given locus of the genome. Thus, their F_1 generation are completely heterozygous. In a backcross design, the F_1 generation is backcrossed with one of their parents, for example, their paternal parents. In an F_2 design, the F_1 is selfed or two F_1's are crossed. A marker-based genetic linkage map of the crossed offsprings is constructed and aims to the QTL identification. Denote the marker positions as P_l, $l = 1, 2, \ldots, L$, with L to be the total number of markers on the genome. The possible genotypes at P_l are A_lA_l and A_la_l in the backcross design, and A_lA_l, A_la_l and a_la_l in the F_2 design, respectively.

Consider the interval mapping introduced by Lander and Botstein (1989). In this mapping scheme, a putative QTL is assumed to be bracketed by two flanking markers P_l and P_{l+1}, $l = 1, 2, \ldots, L-1$. Let M_l be the indicator for different combinations of genotypes at P_l and P_{l+1}. Specifically, it would be a value of $\{1, 2, 3, 4\}$ in the backcross design and $\{1, 2, \ldots, 9\}$ in the F_2 design, respectively. In addition, there are assumed two distinct genotypes, Q and q at a specific locus to affect the trait of age-at-onset. They segregate with two different genotypes of Qq and qq in the backcross population, and three different genotypes of QQ, Qq and qq in the F_2 population, respectively. Let G be the genotype indicator at a putative QTL. It would be a value of $\{0, 1\}$ in the backcross design and $\{0, 1, 2\}$ in the F_2 design, respectively. In total, there are $2^3 = 8$ and $3^3 = 27$ different combinations of genotypes for the putative QTL and its flanking markers in the backcross design and in the F_2 design, respectively.

In order to conduct the interval mapping, we also need to collect the phenotypic data in addition to the marker information to probe possible QTLs and estimate the genotype effect. Suppose there are n progeny subjects in the data set and they are indexed by i, $i = 1, 2, \ldots, n$. After an experiment is conducted, their phenotypic traits of age-at-onset and other variables are collected. Let T_i be the age-at-onset, which however may be censored at time C_i. Thus the smaller value of the underlying age-at-onset T_i and its censoring time C_i is often observed and denoted as $X_i = \min(T_i, C_i)$. Let the event indicator be $\Delta_i = I(T_i \leq C_i)$, indicating whether or not an event is censored. Let $\boldsymbol{M}_{l,i}$ be the p_1-vector of dummy indicators of marker information, and let $\boldsymbol{R}_i(t)$ be the p_2-vector of possible confounding covariates, such as temperature, that need to be adjusted. The $(p_1 + p_2 + 1)$-vector of $(\boldsymbol{M}_{l,i}^{\mathrm{T}}, \boldsymbol{R}_i(\cdot), G_i)^{\mathrm{T}}$ is denoted by $\boldsymbol{Z}_{l,i}(\cdot)$.

2.2 Statistical models

To analyze age-at-onset as time-to-event, the semiparametric Cox proportional hazards model is often used. Specifically, a Cox model would assume that the hazard function of $\boldsymbol{Z_i}$ follows

$$\lambda\{t \mid \boldsymbol{Z_i}(t)\} = \lambda_0(t) \exp\{\boldsymbol{\beta}_{\mathrm{R}}^{\mathrm{T}} \boldsymbol{R}_i(t) + \beta_{\mathrm{G}} G_i\}, \tag{3}$$

where $\boldsymbol{\beta}_{\mathrm{R}}$ and β_G are parameters. Here $\lambda_0(\cdot)$ is usually unspecified. The parameter β_G characterizes the proportionality on the hazard functions due to different genotypes. Thus the model may imply a potential QTL if $\beta_G \neq 0$. Implicitly, this model assumes that the genetic effect at any putative QTL is multiplicative, regardless of the flanking markers. If the genotypes are known, then the

usual maximum partial likelihood method can be implemented for inferences on β_{G}, considering the censoring.

In reality, the exact genotype of a progeny subject, G_i, is usually unknown, although its probability distribution can be calculated given the genotypes of the two-locus flanking markers and the QTL position within the marker interval. With known conditional distributions, the aforementioned Cox model is a mixture model and can be marginalized over the unknown G_i conditional on the marker information. An advantage of doing so is that the statistical analysis would be solely based on the observed marker information, instead of the unknown genotypes at the putative QTL. In Haley and Knott (1992), the similar regression mapping approach was applied in the multiple linear regression models such as (1), when the trait is normally distributed without censoring. In linear regression model, the advantage is greater since the marginalized model is still linear regression, and hence the usual computing routines can be applied, which greatly decrease the computing burden. This, however, does not apply in straightforward terms to the Cox model (3), since its marginalized version does not maintain the proportionality any longer. It actually leads to a complicated form in hazard functions. So the usual maximum partial likelihood does not apply, which may still need computer-intensive methods to estimate the unspecified baseline functions and the parameters of $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\mathrm{R}}^{\mathrm{T}}, \beta_{\mathrm{G}})^{\mathrm{T}}$ jointly.

One alternative model to the Cox proportional hazards model is the additive hazards model proposed by Lin and Ying (1994). In an additive hazards model, the hazard function of $\boldsymbol{Z}_i$ is assumed as

$$\lambda\{t \mid \boldsymbol{Z}_i(t)\} = \lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}G_i. \tag{4}$$

Another alternative would be the additive-multiplicative model

$$\lambda\{t \mid \boldsymbol{Z}_i(t)\} = \lambda_0(t)\exp\{\boldsymbol{\beta}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(t)\} + \beta_{\mathrm{G}}G_i. \tag{5}$$

. In both models, the genotypes at a putative QTL have additive effect on the hazard functions of age-at-onset as trait. That is, the parameter β_{G} characterizes the additional adjusted instantaneous hazard rate caused by one genotype versus the other. Therefore, β_{G} would mean a potential QTL to affect the trait of age-at-onset if $\beta_{\mathrm{G}} \neq 0$. Straightforward calculation shows that the marginal additive hazards model (4) over unknown G_i given its associated marker information becomes

$$\lambda\{t \mid \boldsymbol{R}_i(t), \boldsymbol{M}_{l,i}\} = \lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}^{\mathrm{T}}\boldsymbol{M}_{l,i}, \tag{6}$$

Table 1: *Conditional probabilities of genotype indicator G at a QTL bracketed by markers P_l and P_{l+1} in a backcross population. When r_{l1} or r_{l2} is relatively small, $r_{l1} + r_{l2}$ approximates $r_{l\cdot}$.*

Marker type	Marker genotypes		Conditional probabilities	
M_l	P_l	P_{l+1}	$\boldsymbol{p}_l^{\mathrm{G}} = \mathrm{pr}\{G=1 \mid M_l\}$	$1-\boldsymbol{p}_l^{\mathrm{G}} = \mathrm{pr}\{G=0 \mid M_l\}$
1	$A_l a_l$	$A_{l+1} a_{l+1}$	$\frac{(1-r_{l1})(1-r_{l2})}{1-r_{l\cdot}}$	$\frac{r_{l1} r_{l2}}{1-r_{l\cdot}}$
2	$A_l a_l$	$a_{l+1} a_{l+1}$	$\frac{(1-r_{l1}) r_{l2}}{r_{l\cdot}}$	$\frac{r_{l1}(1-r_{l2})}{r_{l\cdot}}$
3	$a_l a_l$	$A_{l+1} a_{l+1}$	$\frac{r_{l1}(1-r_{l2})}{r_{l\cdot}}$	$\frac{(1-r_{l1}) r_{l2}}{r_{l\cdot}}$
4	$a_l a_l$	$a_{l+1} a_{l+1}$	$\frac{r_{l1} r_{l2}}{1-r_{l\cdot}}$	$\frac{(1-r_{l1})(1-r_{l2})}{1-r_{l\cdot}}$

where $\boldsymbol{M}_{l,i}$ is the vector of dummy indicators of genotypes at markers l and $l+1$, and $\boldsymbol{p}_{l,i}$ is the associated conditional probabilities of G_i given $\boldsymbol{M}_{l,i}$, $l = 1, 2, \ldots, L-1$, and $i = 1, 2, \ldots, n$. The marginalized model (5) can derived similarly.

Assume that the recombination fractions between the marker P_l and the potential QTL, the potential QTL and the marker P_{l+1} and the markers P_l and P_{l+1} are r_{l1}, r_{l2} and r_l, respectively. The conditional probabilities of $\boldsymbol{p}_{l,i}$ in (6) can be further determined as function of $\boldsymbol{r} = (r_{l1}, r_{l2})^{\mathrm{T}}$, $\boldsymbol{p}_{l,i} = \boldsymbol{p}_{l,i}(\boldsymbol{r})$, say. Since the genetic distance between markers P_l and P_{l+1} is usually known and measured in *Morgans* or *centiMorgans*, the Haldane's mapping function can be used to determine

$$r_{l\cdot} = \frac{1}{2}\{1 - \exp(-2d_l)\},$$

where d_l is the genetic distance between P_l and P_{l+1}. This is not the only mapping function we can use. More comprehensive discussion on the genetic mapping functions can be found in Speed (1996). In Tables 1 and 2 for the backcross and F_2 designs, the conditional probabilities are listed for all the potential genotypes, respectively.

2.3 Estimation and inferences

To estimate the parameters in model (6), it is natural to consider the full likelihood function based on $\{X_i, \Delta_i, \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i}\}$, $i = 1, 2, \ldots, n$, when the baseline hazard function of $\lambda_0(\cdot)$ is known. It is in fact proportional to

$$\prod_{i=1}^{n} \lambda(X_i \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i}; \boldsymbol{\beta}, \boldsymbol{r})^{\Delta_i} S(X_i \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i}; \boldsymbol{\beta}, \boldsymbol{r}),$$

Table 2: *Conditional probabilities of genotype indicator G at a QTL bracketed by markers P_l and P_{l+1} in an F_2 population.*

Marker type	Marker genotypes		Conditional probabilities		
M_l	P_l	P_{l+1}	$\boldsymbol{p}_l^{\mathrm{G}} = \mathrm{pr}\{G=0\}$	$\boldsymbol{p}_l^{\mathrm{G}} = \mathrm{pr}\{G=1\}$	$\boldsymbol{p}_l^{\mathrm{G}} = \mathrm{pr}\{G=2\}$
1	A_lA_l	$A_{l+1}A_{l+1}$	$\frac{(1-r_{l1})^2(1-r_{l2})^2}{(1-r_{l\cdot})^2}$	$\frac{2r_{l1}r_{l2}(1-r_{l1})(1-r_{l2})}{(1-r_{l\cdot})^2}$	$\frac{r_{l1}^2r_{l2}^2}{(1-r_{l\cdot})^2}$
2	A_lA_l	$A_{l+1}a_{l+1}$	$\frac{(1-r_{l1})^2(1-r_{l2})r_{l2}}{(1-r_{l\cdot})r_{l\cdot}}$	$\frac{r_{l1}(1-r_{l1})\{r_{l2}^2+(1-r_{l2})\}^2}{(1-r_{l\cdot})r_{l\cdot}}$	$\frac{r_{l1}^2r_{l2}(1-r_{l2})}{(1-r_{l\cdot})r_{l\cdot}}$
3	A_lA_l	$a_{l+1}a_{l+1}$	$\frac{(1-r_{l1})^2r_{l2}^2}{r_{l\cdot}^2}$	$\frac{2r_{l1}r_{l2}(1-r_{l1})(1-r_{l2})}{r_{l\cdot}^2}$	$\frac{r_{l1}^2(1-r_{l2})^2}{r_{l\cdot}^2}$
4	A_la_l	$A_{l+1}A_{l+1}$	$\frac{r_{l1}(1-r_{l1})(1-r_{l2})^2}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{\{r_{l1}^2+(1-r_{l1})^2\}r_{l2}(1-r_{l2})}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{r_{l1}(1-r_{l1})r_{l2}^2}{r_{l\cdot}(1-r_{l\cdot})}$
5	A_la_l	$A_{l+1}a_{l+1}$	$\frac{2r_{l1}r_{l2}(1-r_{l1})(1-r_{l2})}{r_{l\cdot}^2+(1-r_{l\cdot})^2}$	$\frac{\{(1-r_{l1})^2+r_{l1}^2\}\{(1-r_{l2})^2+r_{l2}^2\}}{r_{l\cdot}^2+(1-r_{l\cdot})^2}$	$\frac{2r_{l1}r_{l2}(1-r_{l1})(1-r_{l2})}{r_{l\cdot}^2+(1-r_{l\cdot})^2}$
6	A_la_l	$a_{l+1}a_{l+1}$	$\frac{r_{l1}(1-r_{l1})r_{l2}^2}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{\{(1-r_{l1})^2+r_{l1}^2\}r_{l2}(1-r_{l2})}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{r_{l1}(1-r_{l1})(1-r_{l2})^2}{r_{l\cdot}(1-r_{l\cdot})}$
7	a_la_l	$A_{l+1}A_{l+1}$	$\frac{r_{l1}^2(1-r_{l2})^2}{r_{l\cdot}^2}$	$\frac{2r_{l1}(1-r_{l1})(1-r_{l2})}{r_{l\cdot}^2}$	$\frac{(1-r_{l1})^2r_{l2}^2}{r_{l\cdot}^2}$
8	a_la_l	$A_{l+1}a_{l+1}$	$\frac{r_{l1}^2r_{l2}(1-r_{l2})}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{r_{l1}(1-r_{l1})\{r_{l2}^2+(1-r_{l2})^2\}}{r_{l\cdot}(1-r_{l\cdot})}$	$\frac{(1-r_{l1})^2r_{l2}(1-r_{l2})}{r_{l\cdot}(1-r_{l\cdot})}$
9	a_la_l	$a_{l+1}a_{l+1}$	$\frac{r_{l1}^2r_{l2}^2}{(1-r_{l\cdot})^2}$	$\frac{2r_{l1}(1-r_{l1})r_{l2}(1-r_{l2})\}}{(1-r_{l\cdot})^2}$	$\frac{(1-r_{l1})^2(1-r_{l2})^2}{(1-r_{l\cdot})^2}$

where $S(\cdot) = \exp\{-\Lambda(\cdot)\}$ is survival function and $\Lambda(\cdot) = \int_0^{\cdot} \lambda(u)du$ is cumulative hazard function, respectively. Let $\{N_i(t) = I(X_i \le t, \Delta_i = 1), i = 1, 2, \ldots, n\}$ be the counting processes, and let a filtration be

$$\mathcal{F}_t = \sigma\{N_i(t), Y_i(t), \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i}(t); i = 1, 2, \ldots n\},$$

where $Y_i(t) = I(X_i \ge t)$. Then the likelihood function can be written alternatively as

$$\prod_{i=1}^{n}\left[\prod_{t\le\tau}\lambda(t \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i})^{dN_i(t)}\exp\left\{-\int_0^{\tau} Y_i(u)\lambda(u \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_{l,i})du\right\}\right]$$

where τ is some finite number such that $\lim_{n\to\infty}\sum_{i=1}^{n} Y_i(\tau) > 0$. Here the use of τ is to avoid technical discussion on tail behavior of asymptotic properties for the proposed estimation procedures. For interested readers, we refer to the work by Ying (1993), which can be adapted to extend τ to infinity. The associated score functions with respect to $\boldsymbol{\beta}$ and $\boldsymbol{r}$ are thus

$$\sum_{i=1}^{n}\int_0^{\tau}\frac{(\boldsymbol{R}_i(u)^{\mathrm{T}}, \boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i})^{\mathrm{T}}}{\lambda_0(u) + \beta_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}}dB_{0i}(u; \boldsymbol{\beta}, \boldsymbol{r}, \Lambda_0),$$

and

$$\sum_{i=1}^{n}\int_0^{\tau}\frac{\beta_{\mathrm{G}}\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{l,i}(\boldsymbol{r})^{\mathrm{T}}}{\lambda_0(u) + \beta_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}}dB_{0i}(u; \boldsymbol{\beta}, \boldsymbol{r}, \Lambda_0),$$

where $B_{0i}(\boldsymbol{\beta}, \boldsymbol{r}) = N_i(t) - \int_0^t Y_i(u)\{\lambda_0(u) + \beta_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}du$. Denote the true value of a parameter its same symbol but with subscript '$*$.' For example, the true value of $\boldsymbol{\beta}$ would be $\boldsymbol{\beta}_*$. Then $\{B_{0i}(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*, \Lambda_{0*})\}$ are local square integrable $\mathcal{F}_t$-martingales and hence the score functions

are martingale integrals at the true values of parameters. By further examining the score functions, it is not difficult to discover these two equations are only different in the integrands. Therefore, a more general version can be used to estimate the parameters as

$$\sum_{i=1}^{n} \int_0^{\tau} W(u) \boldsymbol{J}_i(u; \boldsymbol{\beta}, \boldsymbol{r}) dB_{0i}(u; \boldsymbol{\beta}, \boldsymbol{r}, \Lambda_0) = 0. \tag{7}$$

Here $W(\cdot)$ is $\mathcal{F}_t$-measurable weight function converging to a deterministic function of $w(\cdot)$, and $\boldsymbol{J}_i(\cdot)$ are $\mathcal{F}_t$-measurable smooth functions of dimension $p_2 + 3$ with known forms. However, since the baseline hazard function is usually unspecifed, an estimator needs to be developed before using these estimating functions to estimate $\boldsymbol{\beta}$ and $\boldsymbol{r}$ in (7).

Notice that $E\{B_{0i}(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*, \Lambda_{0*})\} = 0$. Thus the estimating equation, $\sum_{i=1}^{n} \int_0^{\tau} dB_{0i}(; \boldsymbol{\beta}, \boldsymbol{r}, \Lambda_0) = 0$, can be used to estimate Λ_0 as if $\boldsymbol{\beta}$ and $\boldsymbol{r}$ were known. This actually leads to a Breslow-type of estimator for Λ_0,

$$\widehat{\Lambda}_0(t; \boldsymbol{\beta}, \boldsymbol{r}) = \int_0^t \frac{\sum_{i=1}^{n} dB_i(u; \boldsymbol{\beta}, \boldsymbol{r})}{\sum_{i=1}^{n} Y_i(u)}.$$

Here $B_i(t; \boldsymbol{\beta}, \boldsymbol{r}) = N_i(t) - \int_0^t Y_i(u)\{\beta_{\mathrm{R}}^{\mathrm{T}} \boldsymbol{R}_i(u) + \beta_{\mathrm{G}} \boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}} \boldsymbol{M}_{l,i}\} du = B_{0i}(t; \boldsymbol{\beta}, \boldsymbol{r}, \Lambda_0) + \int_0^t Y_i(u) d\Lambda_0(u)$. Therefore, the following estimating equations can be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{r}$ by replacing Λ_0 with $\widehat{\Lambda}_0$,

$$\sum_{i=1}^{n} \int_0^{\tau} W(u) \boldsymbol{J}_i(u; \boldsymbol{\beta}, \boldsymbol{r}) d\widehat{B}_{0i}(u; \boldsymbol{\beta}, \boldsymbol{r}, \widehat{\Lambda}_0) = 0. \tag{8}$$

Straightforward calculation thus shows that the left-hand side of the above equation is

$$\mathcal{E}(t; \boldsymbol{\beta}, \boldsymbol{r}) = \sum_{i=1}^{n} \int_0^t W(u)\{\boldsymbol{J}_i(u; \boldsymbol{\beta}, \boldsymbol{r}) - \overline{\boldsymbol{J}}(u; \boldsymbol{\beta}, \boldsymbol{r})\} dB_i(u; \boldsymbol{\beta}, \boldsymbol{r}), \tag{9}$$

at $t = \tau$, where $\overline{\boldsymbol{J}}(u; \boldsymbol{\beta}, \boldsymbol{r}) = \{\sum_{i=1}^{n} Y_i(u) \boldsymbol{J}_i(u; \boldsymbol{\beta}, \boldsymbol{r})\}\{\sum_{i=1}^{n} Y_i(u)\}^{-1}$. Furthermore, since

$$\mathcal{E}(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*) = \sum_{i=1}^{n} \int_0^t W(u)\{\boldsymbol{J}_i(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*)\} dB_{0i}(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*, \Lambda_{0*}), \tag{10}$$

the process of $\mathcal{E}(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*)$ is also an $\mathcal{F}_t-$martingale. Denote $\mathcal{E}(\boldsymbol{\beta}, \boldsymbol{r}) = \mathcal{E}(\tau; \boldsymbol{\beta}, \boldsymbol{r})$ and $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})$ the solutions in $\mathcal{E}(\boldsymbol{\beta}, \boldsymbol{r}) = 0$.

Without loss of generality, we first establish the asymptotic properties for the estimating functions when $W(\cdot) \equiv 1$. In later this section, we will discuss more on the use of weight functions. By a Taylor expansion, we notice that

$$n^{-1/2}\{\mathcal{E}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}) - \mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{r}_*)\} = \left[n^{-1} \left(\left\{ \frac{\partial \mathcal{E}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{r}})}{\partial \boldsymbol{\beta}} \right\}^{\mathrm{T}}, \left\{ \frac{\partial \mathcal{E}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{r}})}{\partial \boldsymbol{r}} \right\}^{\mathrm{T}} \right) \right] \left\{ n^{1/2} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \\ \widehat{\boldsymbol{r}} - \boldsymbol{r}_* \end{pmatrix} \right\},$$

where $(\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}, \widetilde{\boldsymbol{r}}^{\mathrm{T}})^{\mathrm{T}}$ lies in the linear line segment between $(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{r}}^{\mathrm{T}})^{\mathrm{T}}$ and $(\boldsymbol{\beta}_*^{\mathrm{T}}, \boldsymbol{r}_*^{\mathrm{T}})^{\mathrm{T}}$. Therefore, we can establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{r}}$ accordingly.

Lemma 1. Assume that there exists an integrable function of $v(\cdot)$ such that

$$n^{-1}\sum_{i=1}^{n} Y_i(t)\lambda(t \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_i)\{\boldsymbol{J}_i(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*) - \overline{\boldsymbol{J}}(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)\}^{\otimes 2} \to v(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)$$

in probability; and for any $\epsilon > 0$

$$n^{-1}\sum_{i=1}^{n}\int_{n^{-1}\|\boldsymbol{J}_i-\overline{\boldsymbol{J}}\|^2>\epsilon} Y_i(u)\lambda(t \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_i)\|\boldsymbol{J}_i(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*)\|^2 du \to 0$$

in probability as well, as $n \to \infty$. Here $\|\cdot\|$ defines an appropriate Euclidean norm. Then $n^{-1/2}\mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{r}_*)$ converges weakly to a zero-mean Gaussian process in $\mathcal{D}[0,\tau]$ with independent increments and variance function of $V(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*) = \int_0^t v(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*)du$, which is the limit of

$$n^{-1}\sum_{i=1}^{n}\int_0^t Y_i(u)\lambda(t \mid \boldsymbol{M}_{l,i}, \boldsymbol{R}_i)\{\boldsymbol{J}_i(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*)\}^{\otimes 2} du.$$

Here $\mathcal{D}[0,\tau]$ denotes the space of cadlag functions on $[0,\tau]$ endowed with the Skorohod topology.

The stated conditions in the above lemma correspond to the ones of 2.5.1 and 2.5.3 in the Rebolledo's Theorem (Andersen, Borgan, Gill and Keiding, 1993, p. 83), which can be straightforwardly applied in proof and variance calculation. Furthermore in conjunction with *Lemma 1*, we establish the asymptotic properties of the proposed estimators in

Theorem 2. Assume that there exists a nonsingular matrix D such that

$$n^{-1}\int_0^\tau Y_i(u)\{\boldsymbol{J}_i(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u;\boldsymbol{\beta}_*,\boldsymbol{r}_*)\}\begin{pmatrix} \boldsymbol{R}_i(u) \\ \boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i} \\ \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{l,i}(\boldsymbol{r}) \end{pmatrix}^{\mathrm{T}} du \to D(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$$

in probability, in addition to the assumptions in *Lemma 1.* If the partial derivatives of $\boldsymbol{J}_i$ with respect to parameters $\boldsymbol{\beta}$ and $\boldsymbol{r}$ are uniformly continuous in $U(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$, a neighborhood of $(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$, then $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{r}}$ are uniquely defined and consistent in $U(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$. Furthermore,

$$n^{1/2}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \\ \widehat{\boldsymbol{r}} - \boldsymbol{r}_* \end{pmatrix} \xrightarrow{\mathcal{D}} N\{0, D^{-1}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)V(\boldsymbol{\beta}_*,\boldsymbol{r}_*)D^{-1}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)\}.$$

The proof of this theorem can be referred in the Appendix. It is worthwhile to point out that the uniqueness, consistency and asymptotic normality in this theorem are established for a

neighborhood of the true parameters. In practice, there are possibly multiple solutions to the proposed estimating equations over the entire parameter space. One solution to solve the multiple roots issue is to minimize the quadratic version of the estimating functions. The other solution is choose a second set of $\{\boldsymbol{J}_i\}$ to find the roots that make the estimating functions of both choices close to zero. However, as a special situation when the uniform continuity is satisfied on a compact set with the true parameters as interior point and $n^{-1}\mathcal{E}(\boldsymbol{\beta}, \boldsymbol{r})$ converges to a deterministic function at any pair of $\boldsymbol{\beta}$ and $\boldsymbol{r}$ in the compact set, the uniqueness can be extended to the entire compact set if the limiting function has unique solution as well.

To make inference in practice on the parameters, the variance-covariance matrix can be calculated by the empirical version of the asymptotic variance-covariance matrix. That is, it can be estimated by $\widehat{D}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})\widehat{V}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})\widehat{D}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})$. And similarly, the baseline cumulative hazard function can be estimated as

$$\widehat{\Lambda}_0(t; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}) = \int_0^t \frac{\sum_{i=1}^n d\widehat{B}_i(u; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})}{\sum_{i=1}^n Y_i(u)}.$$

We establish its asymptotic properties in the following corollary:

Corollary 3. Under the assumptions listed in *Lemma 1* and *Theorem 2*, $n^{1/2}\{\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}) - \Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process in $\mathcal{D}[0, \tau]$. The limiting covariance function is

$$\begin{aligned} V_\Lambda(s,t) = &\int_0^{\min(s,t)} \{y_0(u)\}^{-1}\{\lambda_0(u) + y_1(u)\}du + \left\{\int_0^t \boldsymbol{y}_2(u)du\right\}^{\mathrm{T}} D^{-1}VD^{-1}\int_0^s \boldsymbol{y}_2(u)du \\ &- \left\{\int_0^t \boldsymbol{y}_2(u)du\right\}^{\mathrm{T}} D^{-1}\int_0^s \{\boldsymbol{y}_{J1}(u) - y_1(u)\boldsymbol{y}_{J0}(u)\}\,du \\ &- \left\{\int_0^s \boldsymbol{y}_2(u)du\right\}^{\mathrm{T}} D^{-1}\int_0^t \{\boldsymbol{y}_{J1}(u) - y_1(u)\boldsymbol{y}_{J0}(u)\}\,du, \end{aligned}$$

where $y_0(u) = \lim_n n^{-1}\sum_{i=1}^n Y_i(u)$, $y_1(u) = \lim_n[\sum_i Y_i(u)\{\boldsymbol{\beta}_R^{\mathrm{T}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}^{\mathrm{T}}\boldsymbol{M}_{l,i}\}]\{\sum_i Y_i(u)\}^{-1}$, $\boldsymbol{y}_2(u) = \lim_n\{\sum_i Y_i(u)(\boldsymbol{R}_i(u)^{\mathrm{T}}, \boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}, \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{l,i})^{\mathrm{T}}\}\{\sum_i Y_i(u)\}^{-1}$, $\boldsymbol{y}_{J0}(u) = \lim_n \overline{\boldsymbol{J}}(u)$ and $\boldsymbol{y}_{J1}(u) = \lim_n[\sum_i Y_i(u)\{\boldsymbol{\beta}_R^{\mathrm{T}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{l,i}^{\mathrm{T}}\boldsymbol{M}_{l,i}\}\boldsymbol{J}_i(u)]\{\sum_i Y_i(u)\}^{-1}$.

To show this corollary, as in Fleming and Harrington (1991, p. 300), consider a decomposition of $\widehat{\Lambda}_0(t; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}) - \Lambda_0(t)$ into summation of three terms: $\widehat{\Lambda}_0(t; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}}) - \widehat{\Lambda}_0(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*)$, $\widehat{\Lambda}_0(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*) - \widetilde{\Lambda}_0(t)$ and $\widetilde{\Lambda}_0(t) - \Lambda_{0*}(t)$, where $\widetilde{\Lambda}_0(t) = \int_{\sum_i Y_i(u)>0} \lambda_0(u)du$. As shown in the Appendix, the first two terms can be written as summation of martingale integrals and the third term is negligible. The asymptotic normality follows and the variance calculation is straightforward. As a result of Corollary 3, the empirical variance estimator of $D^{-1}VD^{-1}$ is also consistent as stated in

Corollary 4. Assume that the total variations of $\{\boldsymbol{Z}_i, \boldsymbol{J}_i; i = 1, 2, \ldots, n\}$ are uniformly bounded. Then $\widehat{D}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})\widehat{V}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})\widehat{D}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{r}})$ converges to $D^{-1}VD^{-1}$ in probability.

The estimation procedure proposed in this section is originally motivated by the score functions in which the infinite-dimensional nuisance parameter of baseline hazard function λ_0 was treated as known. The choices of $W(\cdot)$ and $\boldsymbol{J}_i$'s yield flexibility, for instance, in designing estimating equations with unique solutions. However, the arbitrariness in choice may also lead to potential loss of efficiency, although the inference procedures are always valid. To address this issue, we consider the following parametric submodels:

$$\lambda_0(t \mid \boldsymbol{R}_i, \boldsymbol{M}_{l,i}; \boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\psi}) = \lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i} + \boldsymbol{\psi}^{\mathrm{T}}\overline{\boldsymbol{\lambda}}(t),$$

where $\boldsymbol{\beta}$, $\boldsymbol{r}$ and $\boldsymbol{\psi}$ are parameters, and $\lambda_0(\cdot)$ and $\overline{\boldsymbol{\lambda}}_0(\cdot)$ are known functions. Thus, the log-likelihood function $l(\boldsymbol{\beta}, \boldsymbol{r}, \boldsymbol{\psi})$ can be used to compute the Fisher information matrix at $\boldsymbol{\beta} = \boldsymbol{\beta}_*$, $\boldsymbol{r} = \boldsymbol{r}_*$ and $\boldsymbol{\psi} = \boldsymbol{0}$. By an application of the Cramér-Rao inequality, the lower bound of the covariance matrix for any semiparametric parameter estimators is

$$\left\{I_{(\boldsymbol{\beta},\boldsymbol{r}),(\boldsymbol{\beta},\boldsymbol{r})}(\overline{\boldsymbol{\lambda}}) - I_{(\boldsymbol{\beta},\boldsymbol{r}),\boldsymbol{\psi}}(\overline{\boldsymbol{\lambda}})I^{-1}_{\boldsymbol{\psi},\boldsymbol{\psi}}(\overline{\boldsymbol{\lambda}})I_{\boldsymbol{\psi},(\boldsymbol{\beta},\boldsymbol{r})}(\overline{\boldsymbol{\lambda}})\right\}^{-1},$$

where $I_{(\boldsymbol{\beta},\boldsymbol{r}),(\boldsymbol{\beta},\boldsymbol{r})} = E\{l''_{(\boldsymbol{\beta},\boldsymbol{r})}\}$, $I_{(\boldsymbol{\beta},\boldsymbol{r}),\boldsymbol{\psi}} = I_{\boldsymbol{\psi},(\boldsymbol{\beta},\boldsymbol{r})} = E\{l''_{(\boldsymbol{\beta},\boldsymbol{r}),\boldsymbol{\psi}}\}$ and $I_{\boldsymbol{\psi},\boldsymbol{\psi}} = E\{l''_{\boldsymbol{\psi},\boldsymbol{\psi}}\}$. In fact, the lower bound can be reached when $\overline{\boldsymbol{\lambda}}(t) = \overline{\boldsymbol{\lambda}}_0(t)$ is the limit of

$$\frac{E\left[\sum_i Y_i(t)(\boldsymbol{R}_i^{\mathrm{T}}(t), \boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}, \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{i,l}(\boldsymbol{r})^{\mathrm{T}})^{\mathrm{T}}/\{\lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}\right]}{E\left[\sum_i Y_i(t)/\{\lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}\right]}.$$

Thus a set of optimal estimating functions for $\boldsymbol{\beta}_*$ and $\boldsymbol{r}_*$ are

$$\mathcal{E}_{\mathrm{opt}}(\boldsymbol{\beta}, \boldsymbol{r}) = \sum_{i=1}^{n}\int_0^{\tau}\frac{\left\{\boldsymbol{J}_{\mathrm{opt},i}(u) - \overline{\boldsymbol{J}}_{\mathrm{opt}}(u)\right\}dB_i(u)}{\lambda_0(u) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(u) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}}, \tag{11}$$

where $\boldsymbol{J}_{\mathrm{opt},i}(t) = (\boldsymbol{R}_i^{\mathrm{T}}(t), \boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}, \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{i,l}(\boldsymbol{r})^{\mathrm{T}})^{\mathrm{T}}$, and

$$\overline{\boldsymbol{J}}_{\mathrm{opt}}(t) = \frac{\sum_i Y_i(t)/\{\lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}\boldsymbol{J}_{\mathrm{opt},i}(t)}{\sum_i Y_i(t)/\{\lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}}.$$

By comparing the optimal estimating functions with the general weighted estimating functions in (10), it is obvious that they are identical when $\boldsymbol{J}_i = \boldsymbol{J}_{\mathrm{opt},i}$ and $W(t) = \{\lambda_0(t) + \boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t) + \beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\}^{-1}$. To use the optimal estimating function derived above, a sample-splitting technique (Lin and Ying, 1994) can be used to yield most efficient estimators. However, the nonparametric estimation of $\lambda_0(\cdot)$ causes the fundamental difficulty. One simple way is use ad hoc estimate

of $\lambda_0(t)$ based on prior knowledge, for instance, piecewise exponential hazard function. Another simple way is ignore $\lambda_0(t)+\boldsymbol{\beta}_{\mathrm{R}}\boldsymbol{R}_i(t)+\beta_{\mathrm{G}}\boldsymbol{p}_{i,l}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}$ in $\mathcal{E}_{\mathrm{opt}}(\boldsymbol{\beta},\boldsymbol{r})$. That is, the following estimating functions

$$\widetilde{\mathcal{E}}(\boldsymbol{\beta},\boldsymbol{r})=\sum_{i=1}^{n}\int_0^{\tau}\left\{\boldsymbol{J}_{\mathrm{opt},i}(u;\boldsymbol{\beta},\boldsymbol{r})-\widetilde{\boldsymbol{J}}(u;\boldsymbol{\beta},\boldsymbol{r})\right\}dB_i(u;\boldsymbol{\beta},\boldsymbol{r}),$$

can be used, where $\widetilde{\boldsymbol{J}}(t;\boldsymbol{\beta},\boldsymbol{r})=\sum_i Y_i(t)\boldsymbol{J}_{\mathrm{opt},i}(t;\boldsymbol{\beta},\boldsymbol{r})/\sum_i Y_i(t)$. Apparently, the approximation of both ad hoc approaches to the optimal estimating functions depends on the choices of $\lambda_0(\cdot)$ and how close the ignored terms are to constant, respectively.

As stated previously, use of weight functions in estimating equations can help gain efficiency of estimators, or alleviate the problem of multiple roots. Another important application of weight functions in the estimating equations is for model adequacy checking, as in Lin (1991) for the goodness-of-fit in the Cox proportional hazards model. Specifically, suppose that there are two sets of weight functions for estimating equations,

$$\mathcal{E}_m(\boldsymbol{\beta},\boldsymbol{r})=\sum_{i=1}^{n}\int_0^{\tau}W_m(u)\{\boldsymbol{J}_i(u)-\overline{\boldsymbol{J}}(u)\}dB_i(u),$$

$m=1,2$, respectively. Then according to Theorem 2, their respective 'weighted' estimators, $(\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}},\widehat{\boldsymbol{r}}_1^{\mathrm{T}})^{\mathrm{T}}$ and $(\widehat{\boldsymbol{\beta}}_2^{\mathrm{T}},\widehat{\boldsymbol{r}}_2^{\mathrm{T}})^{\mathrm{T}}$, are both consistent and asymptotically normal, if the additive hazards model indeed holds. In fact, when the additive hazards model is true, the joint distribution of $n^{-1/2}\mathcal{E}_1(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$ and $n^{-1/2}\mathcal{E}_2(\boldsymbol{\beta}_*,\boldsymbol{r}_*)$ is asymptotically normal. By standard counting process techniques, $n^{1/2}\{(\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}},\widehat{\boldsymbol{r}}_1^{\mathrm{T}})-(\widehat{\boldsymbol{\beta}}_2^{\mathrm{T}},\widehat{\boldsymbol{r}}_2^{\mathrm{T}})\}^{\mathrm{T}}$ is shown to be zero-mean normal asymptotically. Let its asymptotic variance be Σ. Then the quadratic form of $n\{(\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}},\widehat{\boldsymbol{r}}_1^{\mathrm{T}})-(\widehat{\boldsymbol{\beta}}_2^{\mathrm{T}},\widehat{\boldsymbol{r}}_2^{\mathrm{T}})\}\Sigma^{-1}\{(\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}},\widehat{\boldsymbol{r}}_1^{\mathrm{T}})-(\widehat{\boldsymbol{\beta}}_2^{\mathrm{T}},\widehat{\boldsymbol{r}}_2^{\mathrm{T}})\}^{\mathrm{T}}$ can be used as an asymptotic central χ^2 test statistic with p_1+p_2+1 degrees of freedom.

2.4 Multiple comparison procedures in QTL detection

The regression models and their estimation are proposed mainly to evaluate the association between the genotypes and the functional quantitative trait at a putative locus bracketed by one specific pair of markers. To detect the QTLs, the following null hypotheses would be tested: $\mathrm{H}_{l,1}:\beta_{\mathrm{G}}=0$ versus $\mathrm{H}_{l,0}:\beta_{\mathrm{G}}\neq 0$, for $l=1,2,\ldots,L$. When the testing procedure is repeated at every pair of consecutive markers throughout the entire linkage map, L multiple-comparison procedures are thus conducted. There are many factors, like genetic map density to influence the distribution of

the test statistics and hence the determination of threshold of the test statistics under the null hypotheses.

2.4.1 Conventional approaches

In a sparse map where markers are spreaded broadly, the marker intervals are considered independent. The usual Bonferroni correction, although conservative, may be used for multiple comparisons. In a dense map when thousands of markers are tested, a common approach to identify the amount of support for QTL at a particular map position is often by graphically displaying the likelihood ratio test statistics as a function of the map position of a putative QTL (Lander and Botstein, 1989). Conventionally a LOD score exceeding 3 usually suggests a QTL for simple Mendelian disease. This threshold is calculated by a Bayesian argument of a prior probability which leads to false positive rate of 5%. In fact, this method may not work well for complex traits such as age-at-onset or highly dense map. Lander and Botstein (1989) showed that the LOD score approaches in large sample to an Ornstein-Uhlenbeck diffusion process in backcross design, while Dupuis and Siegmund (1999) reported similar result for F_2 design. These approximations can be used to determine the threshold of significance levels as given in Lander and Schork (1994). There is also a permutation approach available to permute the phenotypes while the marker information stays (Churchill and Doerge, 1994).

In fact, when a large number of hypothesis testing are performed, the rate of false QTL claims usually needs to be controlled. Conventional approaches, such as the ones discussed in Hochberg and Tamhane (1987), are mainly aimed to controlling the so-called family-wise error rate (FWER), i.e., the probability of at least one false QTL claim when there is no QTL bracketed by any pair of markers in the entire linkage map. When certain proportion of markers to be tested actually depart from their corresponding null hypotheses, these procedures are often conservative and less powerful, as discussed extensively in literature. An important alternative has been developed to focus on the control of the so-called false discovery rate (FDR), which is the expected false positive rate of the rejected hypotheses, since the work by Benjamini and Hochberg (1995). There are both Frequentist and Bayesian FDR-based approaches. Yet most of them rely on the assumptions of the independence among the test statistics, although certain specific form of dependence may be allowed.

Table 3: *Error types in QTL multiple comparisons*

	QTL not claimed	QTL claimed	Total hypotheses tested
No QTL existed	U	V	L_0
QTL existed	T	S	$L - L_0$
Total claims	$L - R$	R	L

2.4.2 A new approach

Nevertheless, most of the conventional approaches are mainly based on the critical assumptions of normality and pooling of independent meiosis. However, given the semiparametric framework of our models, the underlying distributional form of the errors are usually not assumed, and it is thus almost impossible to obtain the usual likelihood maps or profiles to construct a linkage map the based on likelihood ratios. In addition, the independence assumption does not always hold for QTL detection in dense map, given the same set of observations of age-on-set being repeatedly used in the semiparametric models. In this section, we adapt the framework recently constructed by Dudoit, van der Laan and Pollard (2003) and van der Laan, Dudoit and Pollard (2003) to the test statistics on the QTL parameter.

Two kinds of Type I error rate, θ_n, are considered: the generalized family-wise error rate (gFWER) and the proportion of false QTL claims of the rejected hypotheses (PFP). A $gFWER(k)$ is the probability of allowing at least k false claims for some $k+1 \geq 0$, while a $PFP(\kappa)$ is the probability of false claims larger than some κ in (0,1) among the total rejections. We use the notations in Benjamini and Hochberg (1995), as seen in Table 3. Then the $gFWER(k)$ and $PFP(\kappa)$ are actually $\text{pr}\{V \geq k+1\}$ and $\text{pr}\{V/R > \kappa\}$, respectively. Compared with the definitions of the FWER and FDR, it is not difficult to find that

$$FWER = gFWER(0), \text{ and } FDR = E(V/R) = \int_0^1 PFP(\kappa)d\kappa,$$

respectively. For a prespecified α-value, it is said to be of finite sample control if $\theta_n \leq \alpha$, whereas it is of asymptotic control if $\overline{\lim}_{n\to\infty} \theta_n \leq \alpha$. Usually α is chosen to be 0.05.

Consider two statistics that may be used for the lth pair of markers: one is the difference statistic of $\phi_{n,l} = n^{1/2}(\widehat{\beta}_{\text{G},l} - 0)$, and the other is its standardized version of $\varphi_{n,l} = n^{1/2}(\widehat{\beta}_{\text{G},l} - 0)/\sigma_{n,l}$, where $\sigma_{n,l}/\sqrt{n}$ is the estimated standard error of $\widehat{\beta}_{\text{G},l}$. Let $\boldsymbol{\phi}_n = (\phi_{n,1}, \phi_{n,2}, \ldots, \phi_{n,L})^{\text{T}}$ and $\boldsymbol{\varphi}_n = (\varphi_{n,1}, \varphi_{n,2}, \ldots, \varphi_{n,L})^{\text{T}}$, respectively. Assume that F is the underlying data generating distribution.

Denote $\Omega_{n,\phi}(F)$ and $\Omega_{n,\varphi}(F)$ the joint distributions of $\boldsymbol{\phi}_n$ and $\boldsymbol{\varphi}_n$ with limiting distributions of $\Omega_\phi(F)$ and $\Omega_\varphi(F)$, respectively. Then the distributions of V is determined by the corresponding $\Omega_{n,\phi}(F)$ and $\Omega_{n,\varphi}(F)$. Since P is usually unknown, it needs to be estimated to ensure appropriate control of $gFWER(k)$ and $PFP(\kappa)$ in the QTL detection under the null distributions of $\Omega_{0,\phi}(F)$ and $\Omega_{0,\varphi}(F)$, respectively. Since

$$\begin{aligned}\phi_{n,l} &= n^{1/2}(\widehat{\beta}_{\mathrm{G},l} - \beta_{\mathrm{G}*,l}) + n^{1/2}\beta_{\mathrm{G},l} = \phi^*_{n,l} + n^{1/2}\beta_{\mathrm{G}*,l}, \text{ and} \\ \varphi_{n,l} &= \frac{n^{1/2}(\widehat{\beta}_{\mathrm{G},l} - \beta_{\mathrm{G}*,l})}{\sigma_{l,n}} + \frac{n^{1/2}\beta_{\mathrm{G}*,l}}{\sigma_l} \cdot \frac{\sigma_l}{\sigma_{n,l}} = \varphi^*_{n,l} + \frac{n^{1/2}\beta_{\mathrm{G}*,l}}{\sigma_l} \cdot \frac{\sigma_l}{\sigma_{n,l}},\end{aligned}$$

it is therefore true that

$$\phi^*_{n,l} \xrightarrow{\mathcal{L}} N(0, V_\phi(F)) \text{ and } \varphi^*_{n,l} \xrightarrow{\mathcal{L}} N(0, \rho_\varphi(F)),$$

where $V_\phi(F)$ is the covariance matrix and $\rho_\varphi(F)$ is the correlation matrix. Thus according to the Theorem 2 in Dudoit, van der Laan and Pollard (2003), the bootstrapping algorithm such as the following can be used to estimate the null distribution:

Algorithm 1.

1. Obtain a bootstrapping set of samples as $\{(\boldsymbol{X}^b_i, \Delta^b_i, \boldsymbol{Z}^b_i), i = 1, 2, \ldots, n\}$;
2. Compute $\boldsymbol{\phi}^b_n$ and $\boldsymbol{\varphi}^b_n$, respectively;
3. Repeat Step 1 and 2 for a total of B times;
4. Compute the sample mean and the sample variance for each element in $\boldsymbol{\phi}^b_n$ and $\boldsymbol{\varphi}^b_n$;
5. Compute

$$\begin{aligned}\phi^{*,b}_{n,l} &= \sqrt{\min\{1, 1/\widehat{\mathrm{var}}(\phi^b_{n,l})\}}\{\phi^b_{n,l} - \widehat{E}(\phi^b_{n,l})\}, \text{ and} \\ \varphi^{*,b}_{n,l} &= \sqrt{\min\{1, 1/\widehat{\mathrm{var}}(\varphi^b_{n,l})\}}\{\varphi^b_{n,l} - \widehat{E}(\varphi^b_{n,l})\},\end{aligned}$$

   respectively.

6. Compute the empirical distributions of $\phi^{*,b}_{n,l}$ and $\varphi^{*,b}_{n,l}$ for $b = 1, 2, \ldots, B$.

After the null distribution Ω_0 is estimated, there are two procedures to choose actual cutoffs, $\boldsymbol{\beta}^c_{\mathrm{G}} = (\beta^c_{\mathrm{G},1}, \beta^c_{\mathrm{G},2}, \ldots, \beta^c_{\mathrm{G},L})^{\mathrm{T}}$, say, to decide the rejection regions for $\phi_{n,l}$ and $\varphi_{n,l}$, $l = 1, 2, \ldots, L$,

namely, single-step common-quantile and single-step common-cutoff, to control the FWER. For the single-step common-quantile procedure, the cutoffs can be selected as the common quantile of the marginal distributions of the estimated Ω_0. For the single-step common-cutoff, the common cutoff can be selected as $\inf\{c : \theta_n(R \mid \Omega_0) \leq \alpha\}$. Furthermore, their adjusted p-values can be computed as $\widetilde{p}_{n,l} = \inf\{\alpha : l \in S_n(\alpha)\}$, $l = 1, 2, \ldots, L$, where $S_n = \{l : \varphi_{n,l} > c_l(\alpha)\}$.

Based on the aforementioned control of FWER, there are augmentation procedures to select additional rejections to control the gFWER and PFP (van der Laan, Dudoit and Pollard, 2003). Specifically, the augmentations are done in the following algorithm:

Algorithm 2:

1. Sort the adjusted FWER p-values as

$$\widetilde{p}_{n,(1)} \leq \widetilde{p}_{n,(2)} \leq \ldots \leq \widetilde{p}_{n,(L)},$$

   where $(\cdot)$ defines a permutation of $\{1, 2, \ldots, L\}$. Then the rejected null hypotheses of S_n consist of $\{l : \widetilde{p}_{n,l} \leq \alpha\}$ or $\{(l) : l = 1, 2, \ldots, R\}$;

2. Additional rejections are selected as $\{(l) : l = R+1, \ldots, R+k\}$, for $k = k_0$ of a given $0 \leq k_0 \leq L - R$ in the *gFWER*-control, and for $k = \max\{0 \leq l \leq L - R : l/(l+R) \leq \kappa\}$ of a given κ in *FPF*-control, respectively.

Thus the adjusted p-value for controlling the $gFWER(k)$ is calculated as $\widetilde{p}_{n,(l-k)}I(l > k)$, and the adjusted p-value for controlling the $PFP(\kappa)$ is calculated as $\inf\{\alpha : \{l - R(\alpha)\}/l \leq \kappa\}$.

3 NUMERICAL STUDIES

Moderate simulation studies are conducted to evaluate the performance of the proposed models and inference procedures in simulated backcross experiments. For demonstration purpose, one chromosomal segment flanked by two markers are set at 0 and 20 cM. A QTL is assumed at the mid-point of 10cM. The following model is used in simulation to generate age-at-onset

$$\lambda(t \mid R_i, G_i) = \lambda_0(t) + \beta_{\mathrm{R}} R_i + \beta_{\mathrm{G}} G_i.$$

Here, the baseline hazard function is of a Weibull distribution, R_i is continuous and simulated following a uniform distribution on $[0, 1]$, and G_i are simulated as 0/1 following the probabilities

calculated according to Table 1. The parameters of $(\beta_R, \beta_G)^T$ are set to be $(0,0)^T$ and $(1,1)^T$, respectively. The total sample size is chosen to be 100 and 200, respectively. In addition, censoring times are simulated following an exponential distribution to yield about 15% and 30% of censoring, respectively. Estimating functions in (11) with and without weight are used in estimation of parameters. The simulation results are listed in Table 4. One thousand data sets are simulated for each entry in the table to calculate the bias and empirical coverage probabilities. The bias is defined as the difference between the sample mean of estimates over 1000 simulations and the true parameter value; and 95% empirical coverage probability is the percentage of Wald-type 95% confidence intervals that include the true parameter value. As shown in the table, the estimators are virtually unbiased and the nominal confidence intervals have sound coverage probabilities. The weighted estimators tend to have smaller variances, although the reduction is not dramatic under current simulation setting. More extensive simulations need to be conducted to evaluate the efficiency of different sets of $\{\boldsymbol{J}_i\}$.

A study of forest tree growth was conducted at a forest farm in Xuzhou City of Jiangsu Province in China since the Spring of 1988. The study materials used in the study were derived from the triple hybridization of Populus (poplar). As described in Wu, Wang and Huang (1992), a *Populus deltoides* clone (designated I-69) was used as a female parent to mate with an interspecific *P. deltoides* × *P. nigra* clone (designated I-45) as a male parent to produce the hybrids Euramerica poplar, *P. euramericana*. A total of 450 one-year-old rooted three-way hybrid seedlings were planted at a spacing four by five meters in the forest farm. In this study, the age-at-onsets were recorded at the times when the diameters reaches 20cm within 11 growing seasons.

The genetic linkage maps based on the pseudo-test backcross design were constructed using 90 randomly selected genotypes of the 450 hybrids with random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and intersimple sequence repeats (ISSRs), see Yin, Zhang, Huang, et al. (2002). These parent-specific maps consist of the 19 largest linkage groups for each parent parental map. They amount to 19 pairs of chromosomes. For demonstration purpose, we choose the linkage group 10 of the *P. deltoides* parental map to detect statistically meaningful QTLs that potentially affect the age-at-onset of tree growth.

By applying the additive hazards regression model to the time-to-events, one QTL is detected on the linkage group 10 between the markers CA/CCC-640R and CG/CCC-825 in interspecific

Table 4: Summary of simulation studies

			$(\beta_{R*}, \beta_{G*})^T = (0,0)^T$					
			$\widehat{\beta}_R$			$\widehat{\beta}_G$		
n	Censoring Percentage	Weight	Bias $\lvert\widehat{\beta}_R - \beta_{R*}\rvert$	Coverage Probability	Mean $SE(\widehat{\beta}_R)$	Bias $\lvert\widehat{\beta}_R - \beta_{R*}\rvert$	Coverage Probability	Mean $SE(\widehat{\beta}_G)$
100	15%	N	0.034	0.954	0.411	0.044	0.943	0.340
100	15%	Y	0.038	0.955	0.394	0.016	0.962	0.331
100	30%	N	0.045	0.954	0.430	0.064	0.946	0.323
100	30%	Y	0.049	0.942	0.409	0.047	0.953	0.355
200	15%	N	0.033	0.958	0.291	0.041	0.961	0.257
200	15%	Y	0.024	0.974	0.283	0.044	0.948	0.251
200	30%	N	0.047	0.956	0.301	0.021	0.962	0.262
200	30%	Y	0.041	0.958	0.297	0.026	0.938	0.259

			$(\beta_{R*}, \beta_{G*})^T = (1,1)^T$					
			$\widehat{\beta}_R$			$\widehat{\beta}_G$		
n	Censoring Percentage	Weight	Bias $\lvert\widehat{\beta}_R - \beta_{R*}\rvert$	Coverage Probability	Mean $SE(\widehat{\beta}_R)$	Bias $\lvert\widehat{\beta}_R - \beta_{R*}\rvert$	Coverage Probability	Mean $SE(\widehat{\beta}_G)$
100	15%	N	0.016	0.960	0.452	0.043	0.945	0.357
100	15%	Y	0.013	0.956	0.510	0.016	0.977	0.299
100	30%	N	0.018	0.967	0.467	0.012	0.935	0.378
100	30%	Y	0.047	0.974	0.488	0.064	0.925	0.349
200	15%	N	0.022	0.940	0.396	0.039	0.952	0.287
200	15%	Y	0.051	0.958	0.393	0.036	0.947	0.277
200	30%	N	0.031	0.957	0.411	0.037	0.949	0.375
200	30%	Y	0.018	0.971	0.388	0.058	0.942	0.313

hybrids of poplar at adjusted $p-$value of 0.05 based on 1000 bootstrapped samples. The 95% confidence interval for the QTL is (10.51cM, 20.79cM) from the marker CA/CCC-640R. And the estimated genotype effect is 0.517 with standard error of 0.112. In fact, when the overall growth curves of diameters were studied, a QTL was also reported between these same set of markers, at about 13cM from CA/CCC-640R (Ma, Casella and Wu, 2002). These discoveries consistently indicate that there may be existence of QTL between these two markers, whose alleles control not only the growth profiles but also the rate of growth measure by time-to-growths.

4 DISCUSSION

Statistical methods in mapping quantitative trait loci for age-at-onset in presence of censoring has been developed in literature. For example, an additive gamma frailty model with inheritance vectors was proposed in Li (2000) to develop likelihood ratio-based LOD score and test the disease gene locus. Nevertheless, in most of these methods, the popular Cox proportional hazards models have been used. The advantages of the Cox model are its appealing interpretation in hazard functions and readily available softwares. Its nonlinear form, however, often causes challenges in preserving proportionality in mixture models. As pointed in Lin and Ying (1997), the mixed Cox model often leads to "numerical and theoretical difficulties" in inference procedures and "awkward interpretation" in parameter interpretation.

The additive hazards model used in this article has same appealing interpretation in hazard function as the Cox model. More attractively, its additive invariance for mixture distributions enable itself with much more convenience and power in modeling and inferences. One caveat in using the additive hazards model is that it usually requires restricted parameter space to warrant positive hazard functions. A possible solution is to replace the linear combination terms with their respective exponentiated terms. Then it may lose the attractive feature of simplicity in additivity and also leads to cumbersome parameter interpretation.

Although the methodologies are demonstrated in this article with experimental crosses of standard backcross and F_2, they can be easily extended to other occasions. For examples, one such occasion is the so-called Composite Interval Mapping (Zeng, 1994), when the QTLs are suspected to link to multiple markers or intervals between markers. To extend the proposed methods to the

Composite Interval Mapping for age-at-onset, consider

$$\lambda\{t \mid G, \boldsymbol{M}\} = \lambda_0(t) + \beta_{\mathrm{G}} G + \boldsymbol{\beta}_{\mathrm{M}}^{\mathrm{T}} \boldsymbol{M},$$

where $\boldsymbol{M}$ are the selected markers for genetic background control. This would adjust for the effect of other potential QTLs outside the interval containing the putative QTL of interest. For more the type of family pedigree data structure, the additive hazards model with frailties can be also used (Lin and Ying, 1997).

APPENDIX: Proof of theorems

A.1 Proof of Lemma 1

Consider the predictable variation process of $n^{-1/2}\mathcal{E}(t; \boldsymbol{\beta}_*, \boldsymbol{r}_*)$,

$$n^{-1} \sum_{i=1}^{n} \int_0^t Y_i(u) \left\{ \boldsymbol{J}_i(u) - \overline{\boldsymbol{J}}(u) \right\}^{\otimes 2} \lambda(u \mid \boldsymbol{R}_i, \boldsymbol{M}_{l,i}) du,$$

which converges to V, due to the convergence of integrands on a finite closed interval $[0, \tau]$. Thus the first condition of 2.5.1 in the Rebolledo's theorem is satisfied. The third condition of 2.5.3 is also satisfied by the assumptions in the Lemma. Therefore, the martingale integral form of $n^{-1/2}\mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{r}_*)$ leads to its weak convergence according to the Theorem II.5.1 in Andersen, Borgan, Gill and Keiding (1992, p. 83). In fact, the listed conditions are also special forms of the stability and negligibility conditions in Lin and Ying (1995).

A.2 Proof of Theorem 2

Consider

$$n^{-1} \begin{pmatrix} \partial \mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{r}_*)/\partial \boldsymbol{\beta} \\ \partial \mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{r}_*)/\partial \boldsymbol{r} \end{pmatrix} = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \begin{pmatrix} \partial \{ \boldsymbol{J}_i(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) \}/\partial \boldsymbol{\beta} \\ \partial \{ \boldsymbol{J}_i(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) \}/\partial \boldsymbol{r} \end{pmatrix} dB_{0i}(u)$$
$$- n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(u) \left\{ \boldsymbol{J}_i(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) - \overline{\boldsymbol{J}}(u; \boldsymbol{\beta}_*, \boldsymbol{r}_*) \right\} \begin{pmatrix} \partial \lambda(u \mid \boldsymbol{R}_i, \boldsymbol{M}_{l,i})/\partial \boldsymbol{\beta} \\ \partial \lambda(u \mid \boldsymbol{R}_i, \boldsymbol{M}_{l,i})/\partial \boldsymbol{r} \end{pmatrix} du.$$

The first term on the right-hand side of the above equation is sum of martingale integrals. Their predictable variation is of $o_p(1)$, due to the factor of n^{-1}, and hence negligible asymptotically. The

second term converges to $-D$ according to the condition in theorem. When D is nonsingular, consider the Taylor expansion

$$n^{1/2}\begin{pmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_*\\ \widehat{\boldsymbol{r}}-\boldsymbol{r}_*\end{pmatrix}=\left[-n^{-1}\left(\left\{\frac{\partial\mathcal{E}(\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{r}})}{\partial\boldsymbol{\beta}}\right\}^{\mathrm{T}},\left\{\frac{\partial\mathcal{E}(\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{r}})}{\partial\boldsymbol{r}}\right\}^{\mathrm{T}}\right)^{-1}\right]\left\{n^{-1/2}\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)\right\}.$$

By the uniform continuity assumed in Theorem 2, for any $\epsilon>0$, there exists $\delta>0$ independent of n, such that

$$n^{-1}\left\|\begin{pmatrix}\partial\mathcal{E}(\boldsymbol{\beta},\boldsymbol{r})/\partial\boldsymbol{\beta}\\ \partial\mathcal{E}(\boldsymbol{\beta},\boldsymbol{r})/\partial\boldsymbol{r}\end{pmatrix}-\begin{pmatrix}\partial\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{\beta}\\ \partial\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{r}\end{pmatrix}\right\|<\frac{\epsilon}{2},$$

when $\|(\boldsymbol{\beta}^{\mathrm{T}}-\boldsymbol{\beta}_*^{\mathrm{T}},\boldsymbol{r}^{\mathrm{T}}-\boldsymbol{r}_*^{\mathrm{T}})^{\mathrm{T}}\|<\delta$. Therefore, as $n\to\infty$,

$$\mathrm{pr}\left\{\sup\left\|n^{-1}\begin{pmatrix}\partial\mathcal{E}(\boldsymbol{\beta},\boldsymbol{r})/\partial\boldsymbol{\beta}\\ \partial\mathcal{E}(\boldsymbol{\beta},\boldsymbol{r})/\partial\boldsymbol{r}\end{pmatrix}+D\right\|>\epsilon:\|(\boldsymbol{\beta}^{\mathrm{T}}-\boldsymbol{\beta}_*^{\mathrm{T}},\boldsymbol{r}^{\mathrm{T}}-\boldsymbol{r}_*^{\mathrm{T}})^{\mathrm{T}}\|<\delta\right\}$$

goes to 0. By the nonsingularity of D, there exist a neighborhood of $(\boldsymbol{\beta}_*^{\mathrm{T}},\boldsymbol{r}_*^{\mathrm{T}})^{\mathrm{T}}$ such that the uniqueness and consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{r}}$ are warranted. As a result, the asymptotic normality is straightforward.

A.3 Proof of Corollary 3

The first term in the decomposition of $n^{1/2}\{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})-\Lambda_0(t)\}$ is $n^{1/2}\{\widehat{\Lambda}_0(t;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})-\widehat{\Lambda}_0(t;\widehat{\boldsymbol{\beta}}_*,\widehat{\boldsymbol{r}}_*)\}$, which is equivalent to

$$\begin{pmatrix}\partial\widehat{\Lambda}_0(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{\beta}\\ \partial\widehat{\Lambda}_0(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{r}\end{pmatrix}^{\mathrm{T}}n^{1/2}\begin{pmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_*\\ \widehat{\boldsymbol{r}}-\boldsymbol{r}_*\end{pmatrix}+o_p(1)=\left\{\int_0^t\boldsymbol{y}_2(u)du\right\}^{\mathrm{T}}D^{-1}\{n^{-1/2}\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)\}+o_p(1),$$

due to

$$\sup_{t\in[0,\tau]}\left\|\begin{pmatrix}\partial\widehat{\Lambda}_0(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{\beta}\\ \partial\widehat{\Lambda}_0(t;\boldsymbol{\beta}_*,\boldsymbol{r}_*)/\partial\boldsymbol{r}\end{pmatrix}-\int_0^t\boldsymbol{y}_2(u)du\right\|\le\int_0^\tau\left\|\begin{pmatrix}\sum_iY_i(u)\boldsymbol{R}_i(u)/\sum_iY_i(u)\\ \sum_iY_i(u)\boldsymbol{p}_{l,i}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{M}_{l,i}\\ \sum_iY_i(u)\beta_{\mathrm{G}}\boldsymbol{p}_{l,i}^{\mathrm{T}}\boldsymbol{M}_{l,i}\boldsymbol{p}'_{l,i}/\sum_iY_i(u)\end{pmatrix}-\boldsymbol{y}_2(u)\right\|du$$

and the definition of $\boldsymbol{y}_2(\cdot)$. The second term of $n^{1/2}\{\widehat{\Lambda}_0(t;\boldsymbol{\beta},\boldsymbol{r})-\widetilde{\Lambda}_0(t)\}$ in the decomposition is sum of martingale integrals as

$$n^{1/2}\int_{\sum_iY_i(u)>0}\frac{\sum_{i=1}^n dB_{0i}(u)}{\sum_{i=1}^nY_i(u)}.$$

The last term goes to zero almost surely. Thus, $n^{1/2}\{\widehat{\Lambda}_0(t,\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})-\Lambda_0(t)\}$ is equivalent to

$$\left\{\int_0^t \boldsymbol{y}_2(u)du\right\}^{\mathrm{T}} D^{-1}\left[n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\boldsymbol{J}_i(u)-\overline{\boldsymbol{J}}(u)\right\}dB_{0i}(u)\right]+n^{1/2}\int_{\sum_i Y_i(u)>0}\frac{\sum_{i=1}^n dB_{0i}(u)}{\sum_{i=1}^n Y_i(u)}+o_p(1).$$

By a multivariate central limit theorem of martingales, the conclusion in Corollary 3 is proven and the variance calculation follows.

A.4 Proof of Corollary 4

As shown in the proof of Corollary 3, $\widehat{\Lambda}_0(\cdot;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})$ converges to $\Lambda_0(\cdot)$ uniformly on $[0,\tau]$. Furthermore, since $\widehat{D}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})$ is equivalent to $-n^{-1}(\partial\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)^{\mathrm{T}}/\partial\boldsymbol{\beta},\partial\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{r}_*)^{\mathrm{T}}/\partial\boldsymbol{r})^{\mathrm{T}}+o_p(1)$ and the consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{r}}$, $\widehat{D}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})$ is thus consistent. Similarly, $\widehat{V}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})$ is also consistent. Thus, $\widehat{D}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})^{-1}\widehat{V}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})\widehat{D}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{r}})^{-1}$ is consistent.

REFERENCES

Andersen, P. K., Borgan, Ø, Gill, R. D. and Keiding, N. (1992). *Statistical Models Based on Counting Processes*. New York: Springer.

Bender, W., Spierer, P., and Hogness, D. S. (1983), "Chromosomal walking and jumping to isolate DNA from the ACE and ROSY loci and the Bithorax complex in Drosophila-Melanogaster," *Journal of Molecular Biology*, 168, 17-33.

Churchill, G. A. and Doerge, R. W. (1994). "Empirical threshold values for quantitative trait mapping," *Genetics*, 138, 963-971.

Conlon E. M, Goode, E. L., Gibbs, M., Stanford, J. L., Badzioch, M., Janer, M., Kolb, S., Hood, L., Ostrander, E. A., Jarvik, G. P., Wijsman, E. M. (2003), "Oligogenic segregation analysis of hereditary prostate cancer pedigrees: evidence for multiple loci affecting age at onset," *International Journal of Cancer*, 105, 630-635.

Cox, D. R. (1972). "Regression models and life tables (with Discussion)," *Journal of Royal Statistical Society, Series B*, 34, 187-220.

Dupuis, J. and Siegmund, D. (1999). "Statistical methods for mapping quantitative trait loci from a dense set of markers," *Genetics*, 151, 373-386.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Dudoit, S., van der Laan and Pollard, K. S. (2003), "Multiple testing. Part I: Single-step procedures for control of general type I error rates," *UC Berkeley Biostatistics Technical Report #138*.

Haley, C. S. and Knott, S. A. (1992), "A simple regression method for mapping quantitative traits in line crosses using flanking markers," *Heredity*, 69, 315-324.

Lander, E. S. and Botstein, D. (1989), "Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps," *Genetics*, 121, 185-199.

Lander, E. S. and Schork, N. J. (1994), "Genetic dissection of complex traits," *Science*, 265, 2037-2048.

Li, H. (2000), "The additive genetic gamma frailty model for linkage analysis of age-of-onset variation," *Annals of Human Genetics*, 62, 1-14.

Lin, D. Y. (1991), "Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators," *Journal of American Statistical Association*, 86, 725-728.

Lin, D. Y. and Ying, Z. (1994), "Semiparametric analysis of the additive risk model," *Biometrika*, 81, 61-71.

Lin, D. Y. and Ying, Z. (1995), "Semiparametric analysis of general additive-multiplicative hazard models for counting processes," *Annals of Statistics*, 5, 1712-1734.

Lin, D. Y. and Ying, Z. (1997), "Additive hazards regression models for survival data," in *Lecture Notes in Statistics 123*, Eds. D. Y. Lin and T. R. Fleming, pp. 185-198. New York: Springer-Verlag.

Lincoln, S. E., Daly, M. J. and Lander, E. S. (1993), *Constructing Genetic Linkage Maps with MAPMAKER/EXP Version 3.0: A Tutorial and Reference Manual for MAPMAKER/QTL*. Whitehead Institute for Biomedical Research Technical Report. Boston: Whitehead Institute.

Ma, C.-X., Casella, G. and Wu, R. L. (2002), "Functional mapping of quantitative trait loci underlying the character process: a theoretical framework," *Genetics*, 161, 1751-1762.

Morton, N. E. (1984), "Trials of segregation analysis by deterministic and macro simulations," in *Human Population Genetics: the Pittsburgh Symposium*, Ed. A. Chakaravarti, pp. 83-107. New York: Van Nostrand Reinhold.

Speed, T. P. (1996). "What is a genetic mapping function?" in *Genetic Mapping and DNA Sequencing*, Eds. T. P. Speed and M. S. Waterman, pp. 65-88. New York: Springer.

van der Laan, M. J., Dudoit, S. and Pollard, K. S. (2003). "Multiple testing. Part III: Procedures of control of the generalized family wise error rate and proportion of false positives," *UC Berkeley Biostatistics Technical Report #141*.

Wu, R. L., Ma, C. X., Chang, M., Little, R. C., Wu, S. S., Yin, T. M., Huang, M. R., Wang, M. X. and Casella, G. (2002). A logistic mixture model for characterizing genetic determinants causing differentiation in growth trajectories. *Genetic Research* **19**, 235-245.

Wu, R. L., Wang, M. X. and Huang, M. R. (1992). "Quantitative genetics of yield breeding for Populous short rotation culture: I. Dynamics of genetic control and selection models of yield traits," *Canadian Journal of Forest Research*, 22, 175-182.

Yin, T. M., Zhang, X. Y., Huang, M. R., Wang, M. X., Zhuge, Q., Zhu, L. H., Zeng, Z.-B. and Wu, R. L. (2002), "Molecular linkage maps of the Populus genome," *Genome*, 45, 541-555.

Ying, Z. (1993). "A large sample study of rank estimation for censored regression data, " *Annals of Statistics* **21**, 76-99.

Zeng, Z. B. (1994), "Precision mapping of quantitative traits loci," *Genetics*, 136, 1457-1468.