Published in final edited form as:

Int J Biostat. 2014 November 1; 10(2): 231–249. doi:10.1515/ijb-2013-0037.

Semiparametric Odds Rate Model for Modeling Short-Term and Long-Term Effects with Application to a Breast Cancer Genetic Study

Mengdie Yuan and

Department of Statistics, George Mason University

Guoqing Diao

Department of Statistics, George Mason University

Mengdie Yuan: myuan2@gmu.edu; Guoqing Diao: gdiao@gmu.edu

Abstract

The proportional odds model is commonly used in the analysis of failure time data. The assumption of constant odds ratios over time in the proportional odds model, however, can be violated in some applications. Motivated by a genetic study with breast cancer patients, we propose a novel semiparametric odds rate model for the analysis of right-censored survival data. The proposed model incorporates the short-term and long-term covariate effects on the failure time data and includes the proportional odds model as a nested model. We develop efficient likelihood-based inference procedures and establish the large sample properties of the proposed non-parametric maximum likelihood estimators. Simulation studies demonstrate that the proposed methods perform well in practical settings. An application to the motivating example is provided.

Keywords

nonparametric likelihood; odds rate; proportional odds; semiparametric efficiency

1 Introduction

The proportional odds (PO) model (Bennett, 1983a,b, Pettitt, 1984) is commonly used for analyzing survival data. The PO model specifies that the survival function of the failure time T given the covariates $\mathbf{X}$ takes the form

$$S(t|\mathbf{X}) = \frac{e^{-\beta^T \mathbf{X}}}{H(t) + e^{-\beta^T \mathbf{X}}},$$

where $H(t)$ is a strictly increasing function with $H(0) = 0$, and β is a vector of regression parameters. Unlike the Cox proportional hazards model (Cox, 1972), the hazard ratio between two sets of covariates is time-varying under the PO model. Instead, the odds ratio

of survival is assumed to be constant over time under the PO model. Consequently, the hazard ratio converges to one as time goes to infinity.

Much attention has been paid to the development of inference procedures under the PO model and its generalizations. Murphy, Rossini, and van der Vaart (1997) provided statistical inference for the maximum likelihood estimation with right-censored data under the PO model and demonstrated that the estimators of the regression parameters provided by Bennett (1983a) are efficient. The sieve maximum likelihood estimation and the estimation based on weighted empirical odds functions were developed by Shen (1998) and Yang and Prentice (1999), respectively. Computational issues in the PO models are discussed in Hunter and Lange (2002). The PO model has also been extensively studied for current status data and interval censored data, see Rossini and Tsiatis (1996), Huang and Rossini (1997), Rabinowitz, Betensky, and Tsiatis (2000) and Sun, Sun, and Zhu (2007), among others. For correlated failure time data, the PO model with random effects has been investigated in literature (Lam, Lee, and Leung, 2002, Zeng, Lin, and Yin, 2005).

All the aforementioned methods are developed under the assumption of constant odds ratio over time. However, in some applications, this assumption is violated. Violation of this assumption may lead to biased and misleading results. A motivating example is from a copy number alteration (CNA) study with breast cancer patients (Zhang, Martens, Yu, Jiang, Sieuwerts, Smid, Klijn, Wang, and Foekens, 2009). The objective of the study was to determine whether the CNAs influence the time to develop distant metastasis among breast cancer patients. We expect the log odds curves of survival to be parallel for two groups if the proportional odds assumption is true. Figure 1 presents the Kaplan-Meier curves for two groups partitioned by the median of copy number variation data at two SNP loci. It appears that the proportional odds assumption is reasonable at SNP #2124; however, the two curves cross at SNP #39687 indicating a violation of the proportional odds assumption. The PO model failed to capture the phenomenon of the crossing survival at SNP #39687.

To test the proportional odds assumption, Dauxois and Kirmani (2003) developed a formal analytical test for two-sample problems. Resampling approaches using observed score processes for the regression coefficients have been developed by Scheike (2006) to check the goodness-of-fit of the PO model. Liu, Mukherjee, Suesse, Sparrow, and Park (2009) provided graphical diagnostics to check model misspecification for the PO model. To alleviate the proportional odds assumption, Scheike (2006) proposed a flexible transformation model which allows for time-varying coefficients for some covariates in the model. They developed estimation and inference procedures based on martingale estimating equations. Additionally, they proposed to use a robust sandwich-type covariance matrix for the estimates of the regression coefficients. More recently, Chen, Hu, Cheng, Musoke, and Zhao (2012) proposed an extended PO model by incorporating “external” time-varying covariates and developed appropriate inference procedures for the regression parameters. Note that both Liu et al. (2009) and Chen et al. (2012) do not incorporate time-varying coefficients.

In this paper, we propose an alternative approach for relating covariates of interest to the odds of survival for right-censored data. Specifically, we propose a novel semiparametric

odds rate model to incorporate the short-term and long-term covariate effects. This model includes the PO model as a special case, leading to a new diagnostic tool for checking the proportional odds assumption. We establish efficient likelihood-based estimation and inference procedures and derive the asymptotic properties of the proposed estimators. Results from numerical studies demonstrate that the proposed methods perform well in practical settings.

The remaining of the article is organized as follows. In Section 2, we propose the new model and derive the nonparametric maximum likelihood estimation. The model assumptions and the large sample properties are presented in Section 3. In Section 4, we conduct extensive simulation studies to investigate the finite sample properties of the proposed estimators. An application to the motivating CNA study is provided in Section 5 and we conclude with a brief discussion in Section 6. Theoretical proofs are provided in the Appendix.

2 Method

Suppose there are n independent subjects in a random sample. For the i th subject, let T_i be the failure time, C_i be the censoring time, and $\mathbf{X}_i$ be a $p \times 1$ vector of time-independent covariates. The observed data contain $\{(Y_i = \min(T_i, C_i), \delta_i = I(T_i \leq C_i), \mathbf{X}_i), i = 1, \dots, n\}$, where $I(\cdot)$ is the indicator function.

Let $F(t|\mathbf{X})$ and $S(t|\mathbf{X})$ be the conditional distribution function and survival function of the failure time T given the covariates $\mathbf{X}$, respectively. We propose the following generalized odds model

$$\frac{F(t|\mathbf{X})}{S(t|\mathbf{X})} = H(t) \left[e^{\beta^T \mathbf{X}} F(t) + e^{\gamma^T \mathbf{X}} S(t) \right], \quad (1)$$

where $H(t)$ is a non-decreasing function with $H(0) = 0$. In model (1), $H(t)$ is the baseline odds function of failure, $S(t)$ is the baseline survival function which is the survival function for a subject with $\mathbf{X} = \mathbf{0}$, and $F(t) = 1 - S(t)$ is the baseline cumulative distribution function.

Under the proposed model (1), the regression coefficients β and γ have attractive and intuitive interpretations. For two sets of covariates $\mathbf{X}$ and $\mathbf{X}^*$, we can show that

$$\lim_{t \rightarrow 0} \frac{F(t|\mathbf{X})/S(t|\mathbf{X})}{F(t|\mathbf{X}^*)/S(t|\mathbf{X}^*)} = e^{\gamma^T (\mathbf{X} - \mathbf{X}^*)}, \quad \lim_{t \rightarrow \infty} \frac{F(t|\mathbf{X})/S(t|\mathbf{X})}{F(t|\mathbf{X}^*)/S(t|\mathbf{X}^*)} = e^{\beta^T (\mathbf{X} - \mathbf{X}^*)}.$$

Therefore, β and γ can be interpreted as the long-term and short-term log-odds ratios, respectively. When $\beta = \gamma$, model (1) reduces to the PO model. Hence, this model provides a useful diagnostic tool for testing the PO model. To be more specific, we can check the PO model by testing the null hypothesis $H_0: \beta = \gamma$. The test can be performed by using Wald, score or likelihood ratio test statistics.

We now derive the likelihood function for the unknown parameters (β, γ, H) . It can be shown that

$$S(t) = \frac{1}{H(t)+1}, \quad F(t) = \frac{H(t)}{H(t)+1}.$$

It follows that

$$S(t|\mathbf{X}) = \frac{H(t)+1}{e^{\beta^T \mathbf{X}} H^2(t) + (e^{\gamma^T \mathbf{X}} + 1)H(t) + 1}.$$

Consequently, the hazard function is given by

$$\lambda(t|\mathbf{X}) = \frac{H'(t) \left\{ e^{\beta^T \mathbf{X}} H^2(t) + 2e^{\beta^T \mathbf{X}} H(t) + e^{\gamma^T \mathbf{X}} \right\}}{\{H(t)+1\} \left\{ e^{\beta^T \mathbf{X}} H^2(t) + (e^{\gamma^T \mathbf{X}} + 1)H(t) + 1 \right\}},$$

where $H'(t)$ is the first derivative of $H(t)$. Under the conditional independent censoring assumption, the likelihood function for (β, γ, H) has the form

$$\begin{aligned} L_n(\beta, \gamma, H) &= \prod_{i=1}^n \lambda^{\Delta_i}(Y_i|\mathbf{X}_i) S(Y_i|\mathbf{X}_i) \\ &= \prod_{i=1}^n \left[\frac{H'(Y_i) \left\{ e^{\beta^T \mathbf{X}_i} H^2(Y_i) + 2e^{\beta^T \mathbf{X}_i} H(Y_i) + e^{\gamma^T \mathbf{X}_i} \right\}}{\{H(Y_i)+1\} \left\{ e^{\beta^T \mathbf{X}_i} H^2(Y_i) + (e^{\gamma^T \mathbf{X}_i} + 1)H(Y_i) + 1 \right\}} \right]^{\Delta_i} \times \frac{H(Y_i)+1}{e^{\beta^T \mathbf{X}_i} H^2(Y_i) + (e^{\gamma^T \mathbf{X}_i} + 1)H(Y_i) + 1}. \end{aligned} \quad (2)$$

We would like to make inference based on the likelihood function. Note that the maximum of the likelihood function does not exist since we can always choose $H'(Y_i) = 0$ for some Y_i with $i = 1$. Therefore, following the techniques of non-parametric maximum likelihood estimation in the Cox proportional hazards model (Murphy, 1994, 1995), the proportional odds model (Murphy et al., 1997, Zeng et al., 2005), and the semiparametric transformation model (Zeng, Lin, and Lin, 2008), we allow H to be a right continuous function. Therefore, we allow the estimator of H to be a step function with positive jumps only at the observed failure time points. Let $H\{Y_i\}$ denote the jump size of $H(t)$ at Y_i . We replace $H'(Y_i)$ with $H\{Y_i\}$ in the above equation (2) to obtain the nonparametric likelihood function, still denoted by $L_n(\beta, \gamma, H)$ for simplicity,

$$L_n(\beta, \gamma, H) = \prod_{i=1}^n \left[\frac{H\{Y_i\} \left\{ e^{\beta^T \mathbf{X}_i} H^2(Y_i) + 2e^{\beta^T \mathbf{X}_i} H(Y_i) + e^{\gamma^T \mathbf{X}_i} \right\}}{\{H(Y_i)+1\} \left\{ e^{\beta^T \mathbf{X}_i} H^2(Y_i) + (e^{\gamma^T \mathbf{X}_i} + 1)H(Y_i) + 1 \right\}} \right]^{\Delta_i} \times \frac{H(Y_i)+1}{e^{\beta^T \mathbf{X}_i} H^2(Y_i) + (e^{\gamma^T \mathbf{X}_i} + 1)H(Y_i) + 1}.$$

Assume the data have m unique observed failure time points $Y_{(1)} < Y_{(2)} < \dots < Y_{(m)}$. Easily we have $H(t) = \sum_{k: Y_{(k)} \leq t} H\{Y_{(k)}\}$. Then

$$S(t) = \frac{1}{\sum_{k: Y_{(k)} \leq t} H\{Y_{(k)}\} + 1}, \quad F(t) = \frac{\sum_{k: Y_{(k)} \leq t} H\{Y_{(k)}\}}{\sum_{k: Y_{(k)} \leq t} H\{Y_{(k)}\} + 1}.$$

In this case, the odds function $H(t|\mathbf{X}) = F(t|\mathbf{X})/S(t|\mathbf{X})$ is given by

$$\sum_{k: Y_{(k)} \leq t} \left\{ e^{\beta^T \mathbf{X}} F(Y_{(k)}) + e^{\gamma^T \mathbf{X}} S(Y_{(k)}) \right\} H\{Y_{(k)}\}.$$

Hence, to estimate the unknown parameters, we maximize the nonparametric likelihood function with respect to the regression parameters β and γ , and jump sizes $H\{Y_{(1)}\}$, $H\{Y_{(2)}\}$, $\dots$, $H\{Y_{(m)}\}$. The resulting nonparametric maximum likelihood estimators (NPMLEs) are denoted by $(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)$.

The maximization procedure can be performed by using the quasi-Newton optimization algorithm in Press, Teukolsky, Vetterling, and Flannery (1992). Based on Fisher information theory, the covariance matrix for $(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)$ can be obtained by inverting the observed information matrix, which is the negative second derivatives of the nonparametric likelihood function with respect to the regression parameters β and γ , and the jump sizes $H\{Y_{(1)}\}$, $H\{Y_{(2)}\}$, $\dots$, $H\{Y_{(m)}\}$. The justification follows the arguments in Theorem 2 of Parner (1998).

3 Asymptotic Properties

Before we derive the asymptotic properties of the NPMLEs, we impose the following regularity conditions:

- C1** There exists some positive constant δ_0 such that $Pr(C_i \leq \tau | \mathbf{X}_i) \geq \delta_0$ almost everywhere, where τ is the end of the study. In addition, $Pr(T_i > \tau | \mathbf{X}_i) > 0$.
- C2** Conditional on the vector of covariates $\mathbf{X}_i$, the censoring time C_i is independent of T_i for $i = 1, \dots, n$.
- C3** $\mathbf{X}_i$ are bounded for all $i = 1, \dots, n$. Furthermore, $\mathbf{a}^T \mathbf{X}_i = 0$ almost everywhere if and only if $\mathbf{a} = \mathbf{0}$.
- C4** The true values β_0 and γ_0 for parameters β and γ belong to the interior of a known compact set $\mathcal{B}_0$ in $\mathbb{R}^{2p}$.
- C5** The true baseline odds function H_0 belongs to

$$\mathcal{H}_0 = \{H: H \text{ is strictly increasing in } [0, \tau] \text{ and is continuously differentiable with } H(0)=0 \text{ and } H(\tau) < \infty\}.$$

Remark—Condition (C3) implies that $\mathbf{X}_i$ is linearly independent, which ensures the identifiability of the model. The proof of identifiability of the model is provided in Appendix A. Condition (C1) implies that any subject who survives at the end of the study is considered censored at time τ , and every subject has a positive probability to relapse after the study. Conditions (C4) and (C5) guarantee the existence of the NPMLEs.

We first establish the consistency of the NPMLEs.

Theorem 1

Under conditions (C1) – (C5), the NPMLEs $(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)$ exist. Furthermore $\|\hat{\beta}_n - \beta_0\| \rightarrow 0$, $\|\hat{\gamma}_n - \gamma_0\| \rightarrow 0$ and $\sup_{t \in [0, \tau]} |\hat{H}_n(t) - H_0(t)| \rightarrow 0$ almost surely as $n \rightarrow \infty$, where $\|\cdot\|$ is the Euclidean norm.

The existence of (β, γ) can be proved from the compactness of the parameter space and the boundedness of covariates. Suppose for some Y_i with $i = 1$, we have $\hat{H}_n\{Y_i\} = \infty$. In that case, the likelihood function must be 0, which implies $\hat{H}_n\{Y_i\}$ must be finite. Therefore, the NPMLEs exist. The detailed proof for Theorem 1 is provided in Appendix B. The key part of the proof is to establish the boundedness of $\hat{H}_n(\tau)$. Then we can construct a limit, (β^*, γ^*, H^*) , for the sequences $(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)$ and prove that the limit is actually (β_0, γ_0, H_0) by using the Kullback-Leibler information.

We next establish the asymptotic normality of the NPMLEs. Let $l^\infty[0, \tau]$ be a space consisting of all uniformly bounded functions on $[0, \tau]$ and the norm on it is defined as the supremum absolute value on $[0, \tau]$.

Theorem 2

Under conditions (C1)–(C5), $\sqrt{n}(\hat{\beta}_n - \beta_0, \hat{\gamma}_n - \gamma_0, \hat{H}_n - H_0)^T$ converges weakly to a Gaussian process with mean zero in the metric space $l^\infty(\mathcal{A})$ where

$$\begin{aligned} \mathcal{A} = \{(\theta_1, \theta_2, g) : \theta_1 \in \mathcal{R}^p, \theta_2 \in \mathcal{R}^p, g \text{ is a function on } [0, \tau], \\ \text{and } \|\theta_1\| \leq 1, \|\theta_2\| \leq 1, |g|_V \leq 1, \\ \text{where } |g|_V \text{ is the total variation of } g \text{ on } [0, \tau]\}. \end{aligned}$$

Detailed proofs for Theorem 2 are provided in Appendix C. Moreover, we can show that $(\hat{\beta}_n, \hat{\gamma}_n)$ is an asymptotic linear estimator for (β_0, γ_0) and the corresponding influence functions are in the spanned space via score functions, which implies $(\hat{\beta}_n, \hat{\gamma}_n)$ is semiparametrically efficient (Bickel, Klaassen, Ritov, and Wellner, 1993).

4 Simulation Studies

We conducted simulation studies to examine the finite-sample performance of our proposed methods. We generate the failure time from the following model

$$F(t|X) = \frac{e^{\beta X} t^2 + e^{\gamma X} t}{e^{\beta X} t^2 + e^{\gamma X} t + 1},$$

where X was set to be uniform $(-0.5, 0.5)$. Under the above model, $H(t) = t$. Results were obtained under four scenarios of regression parameters (β, γ) : $(0.5, 0.5)$, $(0, 0.5)$, $(-0.5, 0.5)$, and $(-0.5, 0)$. The long-term and short-term effects are equal under the first scenario where the model reduces to the PO model; there is no long-term effect under scenario 2; the long-term and short-term effects are in opposite directions under scenario 3; and there is no short-term effect under scenario 4. We generated the censoring time from an exponential distribution with mean 6. The censoring time was truncated at 4. The censoring rates were about 30% under all scenarios. For each setting, we considered the sample sizes of 100 and 200 and generated 10000 replicates. The variance-covariance matrix of the NPMLEs was approximated by the inverse of the observed information matrix.

Table 1 presents the summary statistics of the NPMLEs of the unknown regression parameters and the baseline odds functions at the time points 0.3, 0.8, 2, and 4, the first three of which correspond to the approximate first quartile, median and third quartile of the failure time, and the last one is the end of the study. In the table, *Bias* is the sampling average of the biases of the estimates; *SEE* denotes the average of the standard error estimates; *SE* denotes the empirical standard deviations of the parameter estimates; *CP* is the coverage probability of 95% confidence interval estimates; and *PO* denotes the average of the parameter estimates under the PO model.

From Table 1, we observed that the biases of the NPMLEs were small under all simulation settings. The standard deviation estimate exhibited the true variation. The coverage probabilities of the 95% confidence interval estimates were close to the nominal level. As expected, as sample size increases, the coverage probabilities improve. The PO model yielded biased results when the proportional odds assumption is violated. Particularly, when the long-term and short-term effects are in opposite directions, the PO model had little power to detect covariate effects.

The next set of simulation studies were concerned with the performance of our proposed procedures for testing the covariate effects and the proportional odds assumption. We used the same simulation settings as the previous simulation study except that the censoring time was from a uniform distribution in $(0, 20)$, yielding approximate censoring rates of 20% under all scenarios. We considered five hypothesis testings here: (H1) $H_0: \beta = 0$; (H2) $H_0: \gamma = 0$; (H3) $H_0: \beta = \gamma = 0$; (H4) $H_0: \beta = \gamma$; (H5) $H_0: \beta_{PO} = 0$. The first four tests were conducted based on our proposed model. The last one was conducted based on the PO model. We used Wald test statistics for testing (H1), (H2), (H4), and (H5) and likelihood ratio test statistics for testing (H3). The data were generated under the same model as above with $n = 400$.

Table 2 presents the sizes/powers at the nominal significance level of 0.05 based on 10000 replicates. Our method provides accurate control of type I error rates. We also observed reasonable powers of the proposed tests for testing (H3) compared to the test using the PO

model when the proportional odds assumption is true; the powers for testing the covariate effect under both scenarios $\beta = \gamma = -1$ and $\beta = \gamma = 1$ are approximately 84.2% and 90.7% using the proposed model and the PO model, respectively. However, when the proportional odds assumption is violated, the proposed tests tend to be more powerful than the test based on the PO model, especially when the long-term and short-term effects are in opposite directions. For example, when $(\beta, \gamma) = (-1, 1)$, the powers for testing (H1), (H2), and (H3) were 0.4553, 0.5213, and 0.5804, respectively, compared to the power of 0.0569 from the PO model. The proposed test for the proportional odds assumption also yielded reasonable powers; for example, we obtained powers of 0.6755 and 0.6713 for $(\beta, \gamma) = (-1, 1)$ and $(\beta, \gamma) = (1, -1)$, respectively.

In the last set of simulation studies, we compared the performance of our method with the method by Scheike (2006). We considered the simulation settings similar to those in Scheike (2006). Specifically, we generated data from the following model

$$\frac{F(t|X, Z)}{S(t|X, Z)} = \{A_0(t) + X A_1(t)\} \exp(Z\beta),$$

where Z is a standard normal random variable, X is a log-normal random variable truncated at 30 with mean $0.3Z$ and standard deviation 0.4, $A_0(t) = 0.1t$, and $A_1(t) = 0.05t^2$. We varied the value of β from -0.2 to 0.2 . The failure times were censored at 12 yielding approximate censoring rates of 30%. For each setting, we considered the sample sizes of 200 and 400. We fit the proposed model with the constraint of equal short-term and long-term effects of Z and compared its performance with that of Scheike (2006).

Table 3 summarizes the results for estimating the effect of Z . The NPMLE of the effect of Z under the proposed model appears to have little bias suggesting that the proposed method is robust to moderate model mis-specifications. Compared to the method of Scheike (2006), the NPMLE has slightly larger biases but the empirical standard deviations are smaller. Additionally, the NPMLE appears to be more efficient in terms of the mean squared error since the method of Scheike (2006) involves the estimation of two unknown functions $A_0(t)$ and $A_1(t)$ instead of one unknown function $H(t)$ in the proposed model.

5 Example

In this Section, we apply the proposed methods to the motivating CNA study (Zhang et al., 2009) described in the Introduction Section. In this study, the copy number data for 313 lymph node-negative breast cancer patients who were selected from the tumor bank at the Erasmus Medical Center are available. Survival analysis on time to develop distant metastasis (DM) was conducted at 56859 SNP loci in the breast cancer genome. The follow up time ranged from 20 to 169 months with a median of 99 months. Among the 313 breast cancer patients, 114 patients developed distant metastasis yielding a censoring rate of 63.6%.

There are two types of estrogen receptor (ER) for tumors, ER positive or ER negative. Among 313 patients, 199 tumors were ER positive; and 161 patients experienced menopause. The ages (years) of patients at the time of surgery were recorded ranging from 26 to 83 with a median of 54. As suggested by Zhang et al. (2009), we conduct the analysis for ER positive group and ER negative group separately. For each group, we include surgery age and menopause status as confounders and evaluate the long-term and short-term effects of the CNV data at each of the 56859 SNP loci. As a common practice, we standardize all the covariates in the numerical implementation.

Using the proposed method, the proportional odds assumption holds at about 99.5% of SNP loci for age and about 66.4% for menopause status. Table 4 presents the results for the ER positive group. The first part of Table 4 presents the top ten SNPs with the smallest P -values for testing the proportional odds assumption. It evidently shows that the PO model is inappropriate when the long-term and short-term effects are in opposite directions. At these loci, the proposed methods are able to detect significant short-term and/or long-term genetic effects missed by the PO model. The second part of Table 4 presents the top ten SNPs with the most significant copy number effect under the PO model. The results demonstrate that our proposed model can also detect significant long-term and/or short-term effects at those SNP loci. In contrast to the first part, we present the first 10 SNPs with the largest P -values for testing the proportional odds assumption in the third part of Table 4. The proposed methods yield short-term and long-term effects similar to the estimates obtained from the PO model, suggesting that the proposed methods work well when the proportional odds assumption is true.

As described in the Introduction Section, the proportional odds assumption appears to be true at SNP #2124, but is violated at SNP #39687. In Table 4, at SNP #2124, the estimates for the long-term and short-term effects were both -0.203 with standard error estimates of 0.285 and 0.242 , respectively. The estimate for the effect under the PO model was -0.193 , which is close to the estimates under the proposed model. At SNP #39687, we obtained significant long-term and short-term effects in opposite directions with estimates 1.482 and -1.130 . However, the PO model yielded insignificant results with regression parameter estimate of 0.006 (P -value = 0.967). These results suggest that in this case the proposed model works as well as the PO model when the proportional odds assumption is true and outperforms the PO model when the assumption is violated.

We also calculated the adjusted odds ratio comparing patients with copy number value of 1 against 0 while fixing the two confounders at their sample means. Figure 2 displays the estimated curves of the log odds ratios with their pointwise 95% confidence bands at SNP #2124 and SNP #39687. It appears that the log odds ratio curve at SNP #2124 is close to a horizontal line whereas the log odds ratio curve at SNP #39687 is non-decreasing and cross 0 around 60 months suggesting violation of the proportional odds assumption.

6 Discussion

We have proposed a novel odds rate model incorporating short-term and long-term covariate effects and have established efficient likelihood-based nonparametric estimation and

inference procedures. The proposed model allows one to explicitly model the long-term and short-term covariate effects on the odds of survival without the use of smoothing. The simulation studies demonstrate that our model can detect the short-term and long-term effects when the proportional odds assumption is violated, while the PO model is inappropriate in this circumstance. We have implemented our new method in an efficient computer program. This program is freely available upon request.

One limitation of the proposed method is that we assume implicitly that the odds ratio function is monotonic over time. Consider two univariate covariates $X > X^*$. It can be shown that

$$\frac{F(t|X)/S(t|X)}{F(t|X^*)/S(t|X^*)} = e^{\beta(X-X^*)}F(t) + e^{\gamma(X-X^*)}S(t) \\ = e^{\beta(X-X^*)}F(t) + e^{\gamma(X-X^*)}\{1-F(t)\}. \quad (3)$$

When $\beta < \gamma$, the odds ratio in (3) is decreasing, whereas when $\beta > \gamma$, the odds ratio is increasing. On the other hand, the transformation model proposed by Scheike (2006) postulates that

$$\frac{F(t|\mathbf{X}, \mathbf{Z})}{S(t|\mathbf{X}, \mathbf{Z})} = \mathbf{X}^T \mathbf{A}(t) \exp(\mathbf{Z}^T \beta),$$

where $\mathbf{Z}$ is another set of covariates which satisfy the proportional odds assumption, $\mathbf{A}(\cdot)$ is a vector of functions, and $H(t|\mathbf{X}) \equiv \mathbf{X}^T \mathbf{A}(t)$ is assumed to be increasing for all $\mathbf{X}$. This model is more general than the proposed model in the sense that it does not impose any assumption on the odds ratio functions $\mathbf{A}(t)$. However, it is not clear whether the estimators of $\mathbf{A}(t)$ are constrained to be monotonic (see Equation (7) in Scheike (2006)). Alternatively, one may consider another model described by Scheike (2006)

$$\frac{F(t|\mathbf{X}, \mathbf{Z})}{S(t|\mathbf{X}, \mathbf{Z})} = H(t) \exp\{\mathbf{Z}^T \beta + \mathbf{X}^T \alpha(t)\}.$$

This model is not identifiable as $t \rightarrow 0$. Furthermore, it is not clear how one can ensure that the estimator of $S(t|\mathbf{X}, \mathbf{Z})$ is non-increasing.

We assume in this paper that the failure time and the censoring time are conditionally independent given the covariates. Violation of this assumption may lead to biased results. For example, in the analysis of time to develop distant metastasis in the CNA study, censoring is likely to be informative because some local recurrences might censor the observations while being driven by the same tumor. To account for such dependent censoring, one may assume certain dependence structures between the failure time and the censoring time (e.g., see Tanaka and Rao (2005) and Li, Tiwari, and Guha (2007)). Alternatively, we can adopt the technique proposed by Othus, Li, and Tiwari (2009) through

the use of the conditional crude hazard function of the censoring time given the covariates. Future research along this direction is warranted.

In this paper, we assume independent observations. In several applications, we may encounter clustered or family data. One interesting future work is to extend our proposed method by incorporating random effects to account for the correlations within a cluster/family. Suppose $\mathbf{X}_{ij}$ is the set of covariates for the j th subject in the i th family. We can incorporate random effects in the following model

$$\frac{F(t|\mathbf{X}_{ij}, \omega_i)}{S(t|\mathbf{X}_{ij}, \omega_i)} = H(t) \omega_i \left[e^{\beta^T \mathbf{X}_{ij}} F(t) + e^{\gamma^T \mathbf{X}_{ij}} S(t) \right],$$

where ω_i denotes the shared frailty for the i th family. We are currently investigating the theoretical and computational aspects of this model.

Yang and Prentice (2005) proposed a two-sample model that accommodates the short-term and long-term covariate effects on the hazard function. More recently, Diao, Zeng, and Yang (2013) extended the above two-sample model to a regression setting accommodating possibly external time-dependent covariates. While the models of Yang and Prentice (2005) and Diao et al. (2013) are extensions of the Cox proportional hazards model, the proposed model renders direct interpretations of short-term and long-term covariate effects on the survival function and includes the proportional odds model as a special case. It is well known that both the proportional hazards model and the proportional odds model belong to a general class of semiparametric transformation models (Zeng et al., 2008). One may be interested in investigating the short-term and long-term covariate effects under this class of transformation models. Future research is warranted.

The flat lines after about 80 months in Figure 1 suggest that a proportion of patients in the CNA study may never develop distant metastasis, i.e., there exists a cured sub-population. To account for the presence of a cured sub-population, we can combine the proposed model and the standard mixture cure model. Specifically, given the covariates $\mathbf{Z}_i$ and $\mathbf{X}_i$, the mixture cure model postulates that the population survival function takes the form

$$S_{pop}(t|\mathbf{Z}_i, \mathbf{X}_i) = 1 - \theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i) S(t|\mathbf{X}_i),$$

where

$$\theta(\mathbf{Z}_i) = \frac{\exp(\beta^T \mathbf{Z}_i)}{1 + \exp(\beta^T \mathbf{Z}_i)}$$

and $S(t|\mathbf{X}_i)$ is the conditional survival function in the un-cured sub-population. Note that $1 - \theta(\mathbf{Z}_i)$ is the cure fraction. We assume the proposed model for $S(t|\mathbf{X}_i)$. The likelihood function is then given by

$$\prod_{i=1}^n \left[\{\theta(\mathbf{Z}_i) f(Y_i | \mathbf{X}_i)\}^{\Delta_i} (1 - \theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i) S(Y_i | \mathbf{X}_i))^{1 - \Delta_i} \right]^{I(Y_i \leq \infty)} \{1 - \theta(\mathbf{Z}_i)\}^{I(Y_i = \infty)},$$

where $f(t|\mathbf{X}_i)$ is the conditional density function of the failure time in the uncured sub-population. We can perform nonparametric likelihood estimation and inference using techniques similar to those in Zeng, Yin, and Ibrahim (2006) and Diao and Yin (2012).

In this study, we performed hypothesis testing at each SNP locus to search for the significant location at which the CNV impacts the time to develop distant metastasis, leading to the so-called “large p small n” problem. Appropriate procedures are needed to control for the family-wise error rate introduced by the multiple comparisons. Many statistical methods have been proposed in literature to account for multiple testing. Particularly, Lin (2005) developed a Monte Carlo approach to approximate the joint distribution of the test statistics along the genome. It would be interesting to adopt this approach in our model setting to adjust for multiple testing.

An interesting area is focused on the selection of important variables that impact the failure time. Recently, Lu and Zhang (2007) proposed variable selection procedures using adaptive LASSO for the PO model. It would be desirable to extend the adaptive LASSO procedure to select variables with significant long-term and/or short-term effects.

7 Software

Program in the form of C code is available on request from the corresponding author.

Acknowledgments

This work was supported by National Institutes of Health [CA150698].

References

- Bennett S. Analysis of survival data by the proportional odds model. *Statistics in Medicine*. 1983a; 2:273–277. [PubMed: 6648142]
- Bennett S. Log-logistic regression-models for survival data. *Applied Statistics - Journal of the Royal Statistical Society Series C*. 1983b; 32:165–171.
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press; 1993.
- Chen YQ, Hu N, Cheng SC, Musoke P, Zhao LP. Estimating regression parameters in an extended proportional odds model. *Journal of the American Statistical Association*. 2012; 107:318–330. [PubMed: 22904583]
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972; 34:187–220.
- Dauxois JY, Kirmani S. Testing the proportional odds model under random censoring. *Biometrika*. 2003; 90:913–922.
- Diao G, Yin G. A general transformation class of semiparametric cure rate frailty models. *Annals of the Institute of Statistical Mathematics*. 2012; 64:959–989.
- Diao G, Zeng D, Yang S. Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics*. 2013 in press.

- Huang J, Rossini A. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*. 1997; 92:960–967.
- Hunter DR, Lange K. Computing estimates in the proportional odds model. *Annals of the Institute of Statistical Mathematics*. 2002; 54:155–168.
- Lam KF, Lee YW, Leung TL. Modeling multivariate survival data by a semiparametric random effects proportional odds model. *Biometrics*. 2002; 58:316–323. [PubMed: 12071404]
- Li Y, Tiwari RC, Guha S. Mixture cure survival models with dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69:285–306.
- Lin DY. An efficient monte carlo approach to assessing statistical significance in genomic studies. *Bioinformatics*. 2005; 21:781–787. [PubMed: 15454414]
- Liu I, Mukherjee B, Suesse T, Sparrow D, Park SK. Graphical diagnostics to check model misspecification for the proportional odds regression model. *Statistics in Medicine*. 2009; 28:412–429. [PubMed: 18693299]
- Lu WB, Zhang HH. Variable selection for proportional odds model. *Statistics in Medicine*. 2007; 26:3771–3781. [PubMed: 17266170]
- Murphy SA. Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics*. 1994; 22:712–731.
- Murphy SA. Asymptotic theory for the frailty model. *The Annals of Statistics*. 1995; 23:182–198.
- Murphy SA, Rossini AJ, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*. 1997; 92:968–976.
- Othuis M, Li Y, Tiwari RC. A class of semiparametric mixture cure survival models with dependent censoring. *Journal of the American Statistical Association*. 2009; 104:1241–1250. [PubMed: 20706564]
- Parner E. Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*. 1998; 26:183–214.
- Pettitt AN. Proportional odds models for survival data and estimates using ranks. *Applied Statistics - Journal of the Royal Statistical Society Series C*. 1984; 33:169–175.
- Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. *Numerical recipes in C: the art of scientific computing*. 2. New York: Cambridge University Press; 1992.
- Rabinowitz D, Betensky RA, Tsiatis AA. Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*. 2000; 56:511–518. [PubMed: 10877311]
- Rossini AJ, Tsiatis AA. A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*. 1996; 91:713–721.
- Scheike TH. A flexible semiparametric transformation model for survival data. *Lifetime Data Analysis*. 2006; 12:461–480. [PubMed: 17031497]
- Shen X. Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*. 1998; 85:165–177.
- Sun JG, Sun LQ, Zhu C. Testing the proportional odds model for interval-censored data. *Lifetime Data Analysis*. 2007; 13:37–50. [PubMed: 17160547]
- Tanaka Y, Rao PV. A proportional hazards model for informatively censored survival times. *Journal of statistical planning and inference*. 2005; 129:253–262.
- van der Vaart, AW.; Wellner, JA. *Weak convergence and empirical processes: with applications to statistics*. New York: Springer-Verlag; 1996.
- Yang S, Prentice R. Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*. 2005; 92:1–17.
- Yang S, Prentice RL. Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association*. 1999; 94:125–136.
- Zeng D, Lin DY. Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*. 2007; 102:167–180.
- Zeng D, Lin DY, Lin X. Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica*. 2008; 18:355–377. [PubMed: 19809573]
- Zeng D, Lin DY, Yin G. Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association*. 2005; 100:470–483.

Zeng D, Yin G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*. 2006; 101:670–684.

Zhang Y, Martens JWM, Yu JX, Jiang J, Sieuwerts AM, Smid M, Klijn JGM, Wang Y, Foekens J. Copy number alterations that predict metastatic capability of human breast cancer. *Cancer Research*. 2009; 69:3795–3801. [PubMed: 19336569]

A Proof of Identifiability

Suppose two sets of parameters (β, γ, H) and $(\tilde{\beta}, \tilde{\gamma}, \tilde{H})$ give the same likelihood function for any observed data. That is,

$$\begin{aligned} & \left[\frac{H'(Y) \left\{ e^{\beta^T \mathbf{X}} H^2(Y) + 2e^{\beta^T \mathbf{X}} H(Y) + e^{\gamma^T \mathbf{X}} \right\}}{\{H(Y)+1\} \left\{ e^{\beta^T \mathbf{X}} H^2(Y) + (e^{\gamma^T \mathbf{X}} + 1)H(Y) + 1 \right\}} \right]^\Delta \\ & \quad \times \frac{H(Y)+1}{e^{\beta^T \mathbf{X}} H^2(Y) + (e^{\gamma^T \mathbf{X}} + 1)H(Y) + 1} \\ & = \left[\frac{\tilde{H}'(Y) \left\{ e^{\tilde{\beta}^T \mathbf{X}} \tilde{H}^2(Y) + 2e^{\tilde{\beta}^T \mathbf{X}} \tilde{H}(Y) + e^{\tilde{\gamma}^T \mathbf{X}} \right\}}{\{\tilde{H}(Y)+1\} \left\{ e^{\tilde{\beta}^T \mathbf{X}} \tilde{H}^2(Y) + (e^{\tilde{\gamma}^T \mathbf{X}} + 1)\tilde{H}(Y) + 1 \right\}} \right]^\Delta \quad (4) \\ & \quad \times \frac{\tilde{H}(Y)+1}{e^{\tilde{\beta}^T \mathbf{X}} \tilde{H}^2(Y) + (e^{\tilde{\gamma}^T \mathbf{X}} + 1)\tilde{H}(Y) + 1}. \end{aligned}$$

Since we assume (4) holds for any Y and $\mathbf{X}$, setting $\gamma = 1$, $Y = 0$ and $\mathbf{X} = \mathbf{0}$, we have $H'(0) = \tilde{H}'(0)$. And for any $\mathbf{X} = \mathbf{0}$ with $\gamma = 1$, $Y = 0$, we have $e^{\gamma^T \mathbf{X}} = e^{\tilde{\gamma}^T \mathbf{X}}$, i.e.

$$(\gamma - \tilde{\gamma})^T \mathbf{X} = 0.$$

By (C3), we obtain $\gamma = \tilde{\gamma}$. With $\gamma = 0$ and $\mathbf{X} = \mathbf{0}$, for any $Y = y$, we have $H(y) = \tilde{H}(y)$. It follows that $e^{\beta^T \mathbf{X}} = e^{\tilde{\beta}^T \mathbf{X}}$, which implies $\beta = \tilde{\beta}$ with (C3).

B Proof of Consistency

By constructing a new step function $H_n(t) = \hat{H}_n(t)/\hat{H}_n(\tau)$, we show that its corresponding likelihood will be larger than the likelihood of the NPMLEs if $\hat{H}_n(\tau)$ goes to infinity.

In fact, let $l_n(\beta, \gamma, H)$ denote the log-likelihood with sample size n . We have

$$\begin{aligned} & \frac{1}{n} \left[l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) - l_n(\hat{\beta}_n, \hat{\gamma}_n, \tilde{H}_n) \right] = \frac{1}{n} \left[l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n(\tau)\tilde{H}_n) - l_n(\hat{\beta}_n, \hat{\gamma}_n, \tilde{H}_n) \right] \\ & = \frac{1}{n} \sum_{i=1}^n \left\{ \Delta_i \log \left[\frac{\hat{H}_n(\tau) \left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(\tau) \tilde{H}_n^2(Y_i) + 2e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n(\tau) \tilde{H}_n(Y_i) + e^{\hat{\gamma}_n^T \mathbf{X}_i} \right\}}{\{\hat{H}_n(\tau)\tilde{H}_n(Y_i)+1\} \left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(\tau) \tilde{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1)\hat{H}_n(\tau)\tilde{H}_n(Y_i) + 1 \right\}} \right. \right. \\ & \quad \left. \left. \frac{\{\tilde{H}_n(Y_i)+1\} \left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \tilde{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1)\tilde{H}_n(Y_i) + 1 \right\}}{\left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \tilde{H}_n^2(Y_i) + 2e^{\hat{\beta}_n^T \mathbf{X}_i} \tilde{H}_n(Y_i) + e^{\hat{\gamma}_n^T \mathbf{X}_i} \right\}} \right] \right. \\ & \quad \left. + \log \left[\frac{\hat{H}_n(\tau)\tilde{H}_n(Y_i)+1}{e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(\tau) \tilde{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1)\hat{H}_n(\tau)\tilde{H}_n(Y_i) + 1} \frac{e^{\hat{\beta}_n^T \mathbf{X}_i} \tilde{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1)\tilde{H}_n(Y_i) + 1}{\tilde{H}_n(Y_i)+1} \right] \right\}. \end{aligned}$$

Given the compactness of the parameter space and the boundedness of covariates and H_n , it is easy to show when $\hat{H}_n(\tau)$ goes to infinity, the first log term goes to a bounded limit, and the second log term goes to $-\infty$. That is,

$$0 \leq \frac{1}{n} \left[l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) - l_n(\hat{\beta}_n, \hat{\gamma}_n, \tilde{H}_n) \right] \rightarrow -\infty, \text{ as } n \rightarrow \infty.$$

By the contradiction, $\hat{H}_n(\tau)$ is bounded almost everywhere. In addition, $H_n(t)$ is bounded by 1. This holds for arbitrary n .

Hence, it follows from compactness theorems that there exist convergent sub-sequences of $(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)$ with limits (β^*, γ^*, H^*) .

We next show that $\beta^* = \beta_0$, $\gamma^* = \gamma_0$, and $H^* = H_0$. We begin with constructing a new sequence of step functions $H_n(t)$ with jumps at those Y_i 's for which $i = 1$. We define the jump size as

$$\frac{\Delta_i}{\bar{H}_n\{Y_i\}} = \sum_{k=1}^n I(Y_k \geq Y_i) \left[\frac{\Delta_k A_k(\beta_0, \gamma_0, H_0) + \left\{ e^{\beta_0^T \mathbf{X}_k} H_0^2(Y_k) + 2e^{\beta_0^T \mathbf{X}_k} H_0(Y_k) + e^{\gamma_0^T \mathbf{X}_k} \right\}^2}{B_k(\beta_0, \gamma_0, H_0)} \right],$$

where

$$\begin{aligned} A_k(\beta, \gamma, H) = & e^{2\beta^T \mathbf{X}_k} H^4(Y_k) + 4e^{2\beta^T \mathbf{X}_k} H^3(Y_k) + (2e^{2\beta^T \mathbf{X}_k} + 4e^{\beta^T \mathbf{X}_k} + e^{\gamma^T \mathbf{X}_k}) H^2(Y_k) \\ & + (2e^{\beta^T \mathbf{X}_k} + e^{\gamma^T \mathbf{X}_k} - 2e^{\beta^T \mathbf{X}_k} + 2e^{2\gamma^T \mathbf{X}_k} + 2e^{\gamma^T \mathbf{X}_k}) H(Y_k) \\ & - 2e^{\beta^T \mathbf{X}_k} + e^{2\gamma^T \mathbf{X}_k} + 2e^{\gamma^T \mathbf{X}_k} \end{aligned}$$

and

$$B_k(\beta, \gamma, H) = (H(Y_k) + 1) \left\{ e^{\beta^T \mathbf{X}_k} H^2(Y_k) + (e^{\gamma^T \mathbf{X}_k} + 1) H(Y_k) + 1 \right\} \times \left\{ e^{\beta^T \mathbf{X}_k} H^2(Y_k) + 2e^{\beta^T \mathbf{X}_k} H(Y_k) + e^{\gamma^T \mathbf{X}_k} \right\}.$$

The above equation is derived similarly from the score functions of $l_n(\beta, \gamma, H)$ with respect to $H\{Y_i\}$ as follows:

$$\frac{\Delta_i}{\hat{H}_n\{Y_i\}} = \sum_{k=1}^n I(Y_k \geq Y_i) \left[\frac{\Delta_k A_k(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) + \left\{ e^{\hat{\beta}_n^T \mathbf{X}_k} \hat{H}_n^2(Y_k) + 2e^{\hat{\beta}_n^T \mathbf{X}_k} \hat{H}_n(Y_k) + e^{\hat{\gamma}_n^T \mathbf{X}_k} \right\}^2}{B_k(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)} \right],$$

where $\bar{H}_n(t) = \sum_{i=1}^n \bar{H}_n\{Y_i\} I(Y_i \leq t)$ and $\hat{H}_n(t) = \sum_{i=1}^n \hat{H}_n\{Y_i\} I(Y_i \leq t)$.

Let

$$M_k(\beta, \gamma, H) = \frac{\Delta_k A_k(\beta, \gamma, H) + \left\{ e^{\beta^T \mathbf{x}_k} H^2(Y_k) + 2e^{\beta^T \mathbf{x}_k} H(Y_k) + e^{\gamma^T \mathbf{x}_k} \right\}^2}{B_k(\beta, \gamma, H)}.$$

We have

$$\hat{H}_n(t) = \int_0^t \frac{\sum_{k=1}^n I(Y_k \geq Y_i) M_k(\beta_0, \gamma_0, H_0)}{\sum_{k=1}^n I(Y_k \geq Y_i) M_k(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n)} d\bar{H}_n(t). \quad (5)$$

Therefore $\hat{H}_n(t)$ is absolutely continuous with respect to $H_n(t)$. If we could verify that $H_n(t)$ converges uniformly to $H_0(t)$ in $[0, \tau]$ with probability 1, then in (5), letting $n \rightarrow \infty$, we obtain

$$H_n^*(t) = \int_0^t \frac{\sum_{k=1}^n I(Y_k \geq Y_i) M_k(\beta_0, \gamma_0, H_0)}{\sum_{k=1}^n I(Y_k \geq Y_i) M_k(\beta^*, \gamma^*, H^*)} dH_0(t).$$

Therefore $H_n^*(t)$ is differentiable with respect to $H_0(t)$. It follows that $d\hat{H}_n(t)/dH_n(t)$ converges uniformly to $dH^*(t)/dH_0(t)$ in $[0, \tau]$. Here we introduce a lemma before the following proof.

Lemma 1

The class

$$\mathcal{F} = \{I(Y_1 \geq y)M_1(\beta, \gamma, H) : y \in [0, \tau], (\beta, \gamma) \in \mathcal{B}, H \in \mathcal{H}\}$$

is a bounded p-Donsker class, therefore a Glivenko-Cantelli class, where $\mathcal{B}$ is a compact set and

$$\mathcal{H} = \{h : h \text{ is increasing in } [0, \tau] \text{ with } h(0)=0, h(\tau) < B_0\}$$

where B_0 is a positive constant such that $\hat{H}_n(\tau) \leq B_0$ with probability one.

Detailed proof of Lemma 1 is provided in Appendix D. By the Glivenko-Cantelli theorem given by van der Vaart and Wellner (1996),

$$\bar{H}_n(t) \rightarrow E \left[\frac{I(Y_1 \leq t)\Delta_1}{\mu(Y_1)} \right] \text{ uniformly, where } \mu(y) = E[I(Y_1 \geq y)M_1(\beta_0, \gamma_0, H_0)].$$

uniformly, where $\mu(y) = E[I(Y_1 \geq y)M_1(\beta_0, \gamma_0, H_0)]$.

Let $S_c(\cdot|\mathbf{X})$ be the survival function for the censoring time. We can show that

$$\mu(y) = E \left[\frac{H'(y) \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(y) + 2e^{\beta_0^T \mathbf{X}} H_0(y) + e^{\gamma_0^T \mathbf{X}} \right\} S_c(y|\mathbf{X})}{\{H_0(y)+1\} \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(y) + (e^{\gamma_0^T \mathbf{X}} + 1) H_0(y) + 1 \right\}} \right].$$

Then we have

$$\begin{aligned} E \left[\frac{I(Y_1 \leq t) \Delta_1}{\mu(Y_1)} \right] &= E \left[\int_0^t \frac{H'(y) \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(y) + 2e^{\beta_0^T \mathbf{X}} H_0(y) + e^{\gamma_0^T \mathbf{X}} \right\} S_c(y|\mathbf{X})}{\mu(y) \{H_0(y)+1\} \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(y) + (e^{\gamma_0^T \mathbf{X}} + 1) H_0(y) + 1 \right\}} dH_0(y) \right] \\ &= \int_0^t dH_0(y) = H_0(t). \end{aligned}$$

Thus, $H_n(t)$ converges uniformly to $H_0(t)$ in $[0, \tau]$ with probability 1. By the argument above, $d\hat{H}_n(t)/dH_n(t)$ converges uniformly to $dH^*(t)/dH_0(t)$ in $[0, \tau]$.

On the other hand,

$$\begin{aligned} & \frac{1}{n} \left[l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) - l_n(\beta_0, \gamma_0, \bar{H}_n) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \Delta_i \log \left[\frac{\hat{H}_n(Y_i) \left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(Y_i) + 2e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n(Y_i) + e^{\hat{\gamma}_n^T \mathbf{X}_i} \right\}}{\{\hat{H}_n(Y_i)+1\} \left\{ e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1) \hat{H}_n(Y_i) + 1 \right\}} \right] \right. \\ & \quad \left. \frac{\{\bar{H}_n(Y_i)+1\} \left\{ e^{\beta_0^T \mathbf{X}_i} \bar{H}_n^2(Y_i) + (e^{\gamma_0^T \mathbf{X}_i} + 1) \bar{H}_n(Y_i) + 1 \right\}}{\bar{H}_n(Y_i) \left\{ e^{\beta_0^T \mathbf{X}_i} \bar{H}_n^2(Y_i) + 2e^{\beta_0^T \mathbf{X}_i} \bar{H}_n(Y_i) + e^{\gamma_0^T \mathbf{X}_i} \right\}} \right] \\ & \quad + \log \left[\frac{\hat{H}_n(Y_i)+1}{e^{\hat{\beta}_n^T \mathbf{X}_i} \hat{H}_n^2(Y_i) + (e^{\hat{\gamma}_n^T \mathbf{X}_i} + 1) \hat{H}_n(Y_i) + 1} \frac{e^{\beta_0^T \mathbf{X}_i} \bar{H}_n^2(Y_i) + (e^{\gamma_0^T \mathbf{X}_i} + 1) \bar{H}_n(Y_i) + 1}{\bar{H}_n(Y_i)+1} \right] \Big\} \\ & \geq 0. \end{aligned}$$

Letting $n \rightarrow \infty$, given $\mathbf{X}$, we have

$$\begin{aligned} & E \left[\frac{1}{n} \sum_{i=1}^n \lim_{n \rightarrow \infty} \left\{ l_{n,i}(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) - l_{n,i}(\beta_0, \gamma_0, \bar{H}_n) \right\} \right] \\ &= E \left\{ \log \left[\left\{ \frac{H^{*'}(Y_1) \left\{ e^{\beta^{*T} \mathbf{X}} H^{*2}(Y_1) + 2e^{\beta^{*T} \mathbf{X}} H^*(Y_1) + e^{\gamma^{*T} \mathbf{X}} \right\}}{(H^*(Y_1)+1) P_1(\beta^*, \gamma^*, H^*)} \right\}^{\Delta_1} \left\{ \frac{H^*(Y_1)+1}{P_1(\beta^*, \gamma^*, H^*)} \right\} \right. \right. \\ & \quad \left. \left. / \left\{ \frac{H'_0(Y_1) \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(Y_1) + 2e^{\beta_0^T \mathbf{X}} H_0(Y_1) + e^{\gamma_0^T \mathbf{X}} \right\}}{(H_0(Y_1)+1) P_1(\beta_0, \gamma_0, H_0)} \right\}^{\Delta_1} \left\{ \frac{H_0(Y_1)+1}{P_1(\beta_0, \gamma_0, H_0)} \right\} \right] \right\} \\ & \geq 0, \end{aligned}$$

where $l_{n,i}(\beta, \gamma, H)$ is the log-likelihood contributed from the i th subject and

$$P_1(\beta, \gamma, H) = e^{\beta^T \mathbf{X}} H^2(Y_1) + (e^{\gamma^T \mathbf{X}} + 1) H(Y_1) + 1.$$

This expectation is the negative Kullback-Leibler information, which is non-positive. As a result, it equals zero. Consequently, we have

$$\begin{aligned} & \left[\frac{H^{*'}(Y_1) \left\{ e^{\beta^{*T} \mathbf{X}} H^{*2}(Y_1) + 2e^{\beta^{*T} \mathbf{X}} H^*(Y_1) + e^{\gamma^{*T} \mathbf{X}} \right\}}{(H^*(Y_1) + 1) P_1(\beta^*, \gamma^*, H^*)} \right]^{\Delta_1} \left\{ \frac{H^*(Y_1) + 1}{P_1(\beta^*, \gamma^*, H^*)} \right\} \\ &= \left[\frac{H'_0(Y_1) \left\{ e^{\beta_0^T \mathbf{X}} H_0^2(Y_1) + 2e^{\beta_0^T \mathbf{X}} H_0(Y_1) + e^{\gamma_0^T \mathbf{X}} \right\}}{(H_0(Y_1) + 1) P_1(\beta_0, \gamma_0, H_0)} \right]^{\Delta_1} \left\{ \frac{H_0(Y_1) + 1}{P_1(\beta_0, \gamma_0, H_0)} \right\}. \end{aligned}$$

Using the same argument in the proof of identifiability, we have $\beta^* = \beta_0$, $\gamma^* = \gamma_0$, and $H^* = H_0$.

Therefore, $\|\hat{\beta}_n - \beta_0\| \rightarrow 0$, $\|\hat{\gamma}_n - \gamma_0\| \rightarrow 0$ and $|\hat{H}_n(t) - H_0(t)| \rightarrow 0$ point-wise, almost surely as $n \rightarrow \infty$. Since H_0 is a continuous function, we have further uniform convergence $\sup_{t \in [0, \tau]} |\hat{H}_n(t) - H_0(t)| \rightarrow 0$ almost surely as $n \rightarrow \infty$.

C Proof of Asymptotic Normality

First, we introduce some notation in the context of empirical process. Let $\mathbb{P}_n$ and $\mathbb{P}$ be the empirical measure and the population distribution of n i.i.d. observations $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_n$. Let $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ denotes the empirical process. Then for any measurable function h ,

$$\mathbb{P}_n h = \frac{1}{n} \sum_{i=1}^n h(\mathbf{O}_i), \quad \mathbb{P} h = \mathbb{E}_{\mathbb{P}}[h],$$

and

$$\mathbb{G}_n h = \sqrt{n}(\mathbb{P}_n h - \mathbb{P} h).$$

By the consistency theorem, $\|\hat{\beta}_n - \beta_0\| + \|\hat{\gamma}_n - \gamma_0\| + \sup_{t \in [0, \tau]} |\hat{H}_n(t) - H_0(t)| < \varepsilon$ with probability 1 when sample size n is large enough. Therefore we consider a neighborhood of (β_0, γ_0, H_0) ,

$$\mathcal{U} = \{(\beta, \gamma, H) : \|\beta - \beta_0\| + \|\gamma - \gamma_0\| + \sup_{t \in [0, \tau]} |H(t) - H_0(t)| < \varepsilon\}$$

for a small constant ε , and define a sequence of functions K_n mapping $\mathcal{U}$ to $l^\infty(\mathcal{A})$ as

$$K_n(\beta, \gamma, H) [\theta_1, \theta_2, g] = \frac{1}{n} \frac{d}{d\varepsilon} l_n \left(\beta + \varepsilon \theta_1, \gamma + \varepsilon \theta_2, H(t) + \varepsilon \int_0^t g dH(s) \right) \Big|_{\varepsilon=0}.$$

Particularly, consider a single observation $(Y, \mathbf{X})$, we have

$$\begin{aligned} & K(\beta, \gamma, H) [\theta_1, \theta_2, g] \\ = & \Delta \left[g(Y) + \frac{e^{\beta^T \mathbf{X}} \{ \theta_1^T \mathbf{X} H^2(Y) + 2\theta_1^T \mathbf{X} H(Y) + 2(H(Y)+1)E_H g(Y) \}}{e^{\beta^T \mathbf{X}} H(Y) (H(Y)+2) + e^{\gamma^T \mathbf{X}}} + \frac{e^{\gamma^T \mathbf{X}} \theta_2^T \mathbf{X}}{e^{\beta^T \mathbf{X}} H(Y) (H(Y)+2) + e^{\gamma^T \mathbf{X}}} \right] \\ & - (\Delta+1) \left[\frac{e^{\beta^T \mathbf{X}} (\theta_1^T \mathbf{X} H^2(Y) + 2H(Y)E_H g(Y))}{e^{\beta^T \mathbf{X}} H^2(Y) + (e^{\gamma^T \mathbf{X}} + 1)H(Y) + 1} + \frac{e^{\gamma^T \mathbf{X}} (\theta_2^T \mathbf{X} H(Y) + E_H g(Y)) + E_H g(Y)}{e^{\beta^T \mathbf{X}} H^2(Y) + (e^{\gamma^T \mathbf{X}} + 1)H(Y) + 1} \right] \\ & - (\Delta-1) \frac{E_H g(Y)}{H(Y)+1}, \end{aligned} \quad (6)$$

where $E_H g(Y) = \int_0^Y g(s) dH(s)$.

Essentially, the empirical process $\mathbb{G}_n K$ also maps from $\mathcal{U}$ to $l^\infty(\mathcal{A})$, and

$$\mathbb{P}_n K(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) [\theta_1, \theta_2, g] = 0, \mathbb{P} K(\beta_0, \gamma_0, H_0) [\theta_1, \theta_2, g] = 0.$$

This is implied from the definition of K_n and the consistency of NPMLEs.

To prove the asymptotic normality, we verify three Properties a) – c) below. Then it follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that

$\sqrt{n}(\hat{\beta}_n - \beta_0, \hat{\gamma}_n - \gamma_0, \hat{H}_n - H_0)^T$ converges in distribution to a Gaussian process in $l^\infty(\mathcal{A})$.

- a. $\mathbb{G}_n K(\hat{\beta}_n, \hat{\gamma}_n, \hat{H}_n) [\theta_1, \theta_2, g] - \mathbb{G}_n K(\beta_0, \gamma_0, H_0) [\theta_1, \theta_2, g] = o_P[1 + \sqrt{n}(\|\hat{\beta}_n - \beta_0\| + \|\hat{\gamma}_n - \gamma_0\| + \sup_{t \in [0, \tau]} |\hat{H}_t - H_t|)]$.
- b. $\mathbb{G}_n K(\beta_0, \gamma_0, H_0) [\theta_1, \theta_2, g]$ converges to a tight random element ξ .
- c. $\mathbb{P} K(\beta, \gamma, H) [\theta_1, \theta_2, g]$ is Frechet-differentiable at (β_0, γ_0, H_0) with a continuously invertible derivative $(\dot{\mathbb{P}} K)(\beta_0, \gamma_0, H_0)$.

Note that we assume that β and γ belong to the interior of a known compact set $\mathcal{B}_0$ in $\mathcal{A}^p$, $H(\tau) < \infty$, and $\mathbf{X}$ are bounded. Therefore it is easy to see from the explicit expression (6) where $(\theta_1, \theta_2, g) \in \mathcal{A}$ that $K(\beta, \gamma, H)$ is continuously differentiable with respect to (β, γ) and the norm of the derivative is bounded. Furthermore, $K(\beta, \gamma, H_1) - K(\beta, \gamma, H_2)$ is dominated by

$$\left\{ |H_1(Y) - H_2(Y)| + \int_0^Y d|H_1(s) - H_2(s)| \right\}.$$

Therefore, we have

$$\sup_{(\theta_1, \theta_2, g) \in \mathcal{A}} \mathbb{P}(K(\beta, \gamma, H)[\theta_1, \theta_2, g] - K(\beta_0, \gamma_0, H_0)[\theta_1, \theta_2, g])^2 \rightarrow 0,$$

as $\|\beta - \beta_0\| + \|\gamma - \gamma_0\| + \sup_{t \in [0, \tau]} |H(t) - H_0(t)| \rightarrow 0$.

Using the arguments in Appendix D, we prove that the class

$$\{K(\beta, \gamma, H)[\theta_1, \theta_2, g] - K(\beta_0, \gamma_0, H_0)[\theta_1, \theta_2, g] : (\beta, \gamma, H) \in \mathcal{U}, (\theta_1, \theta_2, g) \in \mathcal{A}\}$$

is p-Donsker. Therefore, by applying Lemma 3.3.5 of van der Vaart and Wellner (1996), Property a) holds.

Similarly, we can prove that the class $\{K(\beta_0, \gamma_0, H_0)[\theta_1, \theta_2, g] : (\theta_1, \theta_2, g) \in \mathcal{A}\}$ is p-Donsker. Then Property b) holds. In fact, ξ is a Gaussian process indexed by $(\theta_1, \theta_2, g) \in \mathcal{A}$. The covariance between $\xi(\theta_1, \theta_2, g)$ and $\xi(\theta_1^*, \theta_2^*, g^*)$ is denoted by

$$\mathbb{P}(K(\beta_0, \gamma_0, H_0)[\theta_1, \theta_2, g] \times K(\beta_0, \gamma_0, H_0)[\theta_1^*, \theta_2^*, g^*]).$$

The smoothness of $\mathbb{P}(\beta, \gamma, H)$ implies the Frechet differentiability in Property c).

According to the proof of Theorem 2 in Zeng and Lin (2007), to show the continuously invertibility of $(\dot{\mathbb{P}}_K)$ can be implemented by showing that $(\dot{\mathbb{P}}_K)$ is one-to-one; that is,

$$K(\beta_0, \gamma_0, H_0)[\theta_1, \theta_2, g] = 0$$

implies $\theta_1 = 0, \theta_2 = 0, g = 0$. In fact, to prove the continuously invertibility, it suffices to prove the Fisher information along the path $\beta_0 + \varepsilon\theta_1, \gamma_0 + \varepsilon\theta_2, H_0 + \varepsilon E_{H_0}g$ is nonsingular. This can be proved by using the idea of proving the identifiability in Appendix A.

Let $\varepsilon = 1, Y = 0$, we have $\theta_2^T \mathbf{X} + g(0) = 0$. By Condition (C3), we have $\theta_2 = 0$ and $g(0) = 0$.

Let $\varepsilon = 0, \mathbf{X} = \mathbf{0}$, we obtain $E_H g = 0$. While when $\varepsilon = 1, \mathbf{X} = \mathbf{0}, g = \frac{2E_H g}{H+1} = 0$. Thus, $g = 0$.

Plugging in $\theta_2 = 0, g = 0$, we have

$$K(\beta_0, \gamma_0, H_0)[\theta_1, \mathbf{0}, 0] = \Delta \frac{e^{\beta_0^T \mathbf{X}} (\theta_1^T \mathbf{X} H^2 + 2\theta_1^T \mathbf{X} H)}{e^{\beta_0^T \mathbf{X}} H(H+2) + e^{\gamma_0^T \mathbf{X}}} - (\Delta+1) \frac{e^{\beta_0^T \mathbf{X}} \theta_1^T \mathbf{X}}{e^{\beta_0^T \mathbf{X}} H^2 + (e^{\gamma_0^T \mathbf{X}} + 1)H + 1}.$$

By letting $\varepsilon = 0$, we immediately have $\theta_1 = \mathbf{0}$ by Condition (C3). Therefore, Property c) is proved.

It follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that

$\sqrt{n}(\hat{\beta}_n - \beta_0, \hat{\gamma}_n - \gamma_0, \hat{H}_n - H_0)^T$ converges in distribution to a tight Gaussian random element $(\mathbb{P}K)^{-1} \xi$ in $l^\infty(\mathcal{A})$.

D Proof of p-Donsker of

Recall that

$$\mathcal{F} = \{I(Y_1 \geq y)M_1(\beta, \gamma, H) : y \in [0, \tau], (\beta, \gamma) \in \mathcal{B}, H \in \mathcal{H}\}.$$

First we prove that

$$\mathcal{F}^* = \{M_1(\beta, \gamma, H) : y \in [0, \tau], (\beta, \gamma) \in \mathcal{B}, H \in \mathcal{H}\}$$

is a bounded p-Donsker class. And $\{I(Y_1 \geq y), y \in [0, \tau]\}$ is uniformly bounded and monotone on the real line, therefore is p-Donsker. By the preservation of Donsker property under product, we can prove Lemma B.1.

It remains to prove that $\mathcal{F}^*$ is a bounded p-Donsker class. Despite the complex expression for $M_1(\beta, \gamma, H)$, it is easy to know that it is bounded and continuously differentiable with respect to (β, γ) for any $(\beta, \gamma) \in \mathcal{B}$ and the norm of the derivative are also bounded. It is also dominated by H . Therefore, by the mean-value theorem, $|M_1(\beta, \gamma, H) - M_1(\tilde{\beta}, \tilde{\gamma}, \tilde{H})|$ is dominated by $\|\beta - \tilde{\beta}\| + \|\gamma - \tilde{\gamma}\| + \sup_{t \in (0, \tau)} |H(t) - \tilde{H}(t)|$. Then $\mathcal{F}^*$ is a p-Donsker class (van der Vaart and Wellner, 1996). It follows that $\mathcal{F}$ is a p-Donsker class.

Figure 1.
The Kaplan-Meier survival curves of time to develop distant metastasis for the CNA study at SNP #2124 and SNP #39687.

Figure 2.

Log odds ratios (solid curves) and the 95% pointwise confidence bands (dashed curves) at SNP #2124 and SNP #39687.

Table 1

Summary statistics of the proposed NPMLEs

Parameter	Bias	SEE	SE	CP	PO
Scenario 1: $(\beta, \gamma) = (0.5, 0.5)$					
n=100					
β	0.024	1.277	1.404	0.955	0.498
γ	-0.024	1.068	1.132	0.941	
$H(0.3)$	-0.013	0.071	0.073	0.908	
$H(0.8)$	-0.031	0.172	0.173	0.921	
$H(2.0)$	-0.054	0.486	0.501	0.913	
$H(4.0)$	-0.028	1.273	1.353	0.898	
n=200					
β	0.012	0.849	0.868	0.959	0.499
γ	-0.012	0.739	0.741	0.952	
$H(0.3)$	-0.006	0.051	0.051	0.937	
$H(0.8)$	-0.014	0.120	0.119	0.938	
$H(2.0)$	-0.020	0.335	0.337	0.934	
$H(4.0)$	-0.001	0.871	0.893	0.929	
Scenario 2: $(\beta, \gamma) = (0, 0.5)$					
n=100					
β	-0.010	1.275	1.411	0.956	0.268
γ	-0.015	1.064	1.130	0.941	
$H(0.3)$	-0.013	0.072	0.073	0.911	
$H(0.8)$	-0.031	0.172	0.174	0.919	
$H(2.0)$	-0.053	0.487	0.504	0.917	
$H(4.0)$	-0.022	1.278	1.378	0.901	
n=200					
β	0.000	0.849	0.869	0.959	0.270
γ	-0.009	0.737	0.739	0.951	
$H(0.3)$	-0.005	0.051	0.051	0.936	
$H(0.8)$	-0.013	0.120	0.120	0.938	
$H(2.0)$	-0.020	0.336	0.338	0.934	
$H(4.0)$	0.000	0.873	0.902	0.928	
Scenario 3: $(\beta, \gamma) = (-0.5, 0.5)$					
n=100					
β	-0.049	1.285	1.419	0.954	0.038

Parameter	Bias	SEE	SE	CP	PO
γ	-0.005	1.065	1.125	0.940	
$H(0.3)$	-0.013	0.072	0.073	0.909	
$H(0.8)$	-0.030	0.173	0.176	0.919	
$H(2.0)$	-0.053	0.491	0.509	0.916	
$H(4.0)$	-0.038	1.277	1.357	0.898	
β	-0.016	0.856	0.878	0.958	0.041
γ	-0.006	0.736	0.739	0.949	
$H(0.3)$	-0.006	0.051	0.051	0.935	
$H(0.8)$	-0.013	0.121	0.121	0.938	
$H(2.0)$	-0.019	0.340	0.344	0.934	
$H(4.0)$	-0.004	0.877	0.905	0.932	
Scenario 4: $(\beta, \gamma) = (-0.5, 0)$					
β	-0.016	1.277	1.416	0.955	-0.234
γ	-0.012	1.066	1.131	0.939	
$H(0.3)$	-0.013	0.071	0.072	0.908	
$H(0.8)$	-0.032	0.172	0.174	0.918	
$H(2.0)$	-0.056	0.486	0.504	0.914	
$H(4.0)$	-0.039	1.268	1.345	0.896	
β	-0.004	0.849	0.871	0.957	-0.231
γ	-0.007	0.738	0.741	0.948	
$H(0.3)$	-0.006	0.051	0.051	0.934	
$H(0.8)$	-0.014	0.120	0.119	0.938	
$H(2.0)$	-0.021	0.336	0.339	0.936	
$H(4.0)$	-0.004	0.871	0.897	0.930	

Table 2

Empirical sizes/powers at significance level of 0.05 based on 10000 replicates

β	γ	$H_0: \beta = 0$	$\gamma = 0$	$\beta = \gamma = 0$	$\beta = \gamma$	$\beta_{FO} = 0$
-1	-1	0.4789	0.5126	0.8457	0.0494	0.9065
-1	0	0.4812	0.0537	0.4212	0.2249	0.3548
-1	1	0.4553	0.5213	0.5804	0.6755	0.0569
0	-1	0.0511	0.5223	0.4534	0.2361	0.3932
0	0	0.0502	0.0521	0.0533	0.0488	0.0518
0	1	0.0488	0.5169	0.4532	0.2358	0.3952
1	-1	0.4497	0.5259	0.5787	0.6713	0.0568
1	0	0.4727	0.0536	0.4181	0.2261	0.3586
1	1	0.4767	0.5022	0.8422	0.0492	0.9065

Table 3

Comparison between the Proposed Method and the Method of Scheike (2006)[†]

	New			Scheike(2006)				
	<i>True</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>Bias</i>	<i>SE</i>	<i>MSE</i>	<i>RE</i>
n=200	-0.2	0.009	0.122	0.015	0.004	0.133	0.018	1.170
	-0.1	0.005	0.122	0.015	0.002	0.131	0.017	1.169
	0.0	0.001	0.121	0.015	0.001	0.131	0.017	1.172
	0.1	-0.003	0.122	0.015	0.000	0.132	0.017	1.172
n=400	0.2	-0.007	0.123	0.015	-0.002	0.133	0.018	1.170
	-0.2	0.012	0.086	0.007	0.003	0.094	0.009	1.196
	-0.1	0.007	0.085	0.007	0.003	0.094	0.009	1.211
	0.0	0.002	0.085	0.007	0.002	0.094	0.009	1.217
	0.1	-0.003	0.085	0.007	0.001	0.094	0.009	1.216
	0.2	-0.009	0.085	0.007	0.000	0.094	0.009	1.203

[†] SE is the empirical standard deviation; MSE is the mean squared error; and RE is the MSE relative efficiency of the proposed estimator compared to the estimator of Scheike (2006).

Table 4

Results for the CNA study¹

SNP loci with the smallest <i>P</i> -values for testing the PO assumption										
SNP	$\hat{\beta}_3$	$se(\hat{\beta}_3)$	$\hat{\gamma}_3$	$se(\hat{\gamma}_3)$	$\hat{\beta}_3 = 0$	$\gamma_3 = 0$	$\hat{\beta}_3 = \gamma_3 = 0$	$\hat{\beta}_{3PO}$	$\beta_{3PO} = 0$	
39687	1.482	0.249	-1.130	0.212	2.55E-09	1.01E-07	2.93E-10	1.82E-12	0.006	0.967
39734	1.481	0.249	-1.130	0.212	2.55E-09	1.01E-07	2.94E-10	1.82E-12	0.006	0.967
36091	1.463	0.249	-1.139	0.212	4.52E-09	8.01E-08	3.02E-10	2.48E-12	0.001	0.996
11222	1.404	0.246	-1.118	0.209	1.14E-08	8.88E-08	9.13E-10	4.75E-12	-0.022	0.885
27435	-2.410	0.341	0.708	0.195	1.68E-12	2.84E-04	4.21E-07	5.12E-12	-0.172	0.222
7879	1.411	0.251	-1.122	0.209	1.82E-08	7.88E-08	9.19E-10	6.54E-12	-0.028	0.851
7934	1.411	0.251	-1.122	0.209	1.82E-08	7.89E-08	9.19E-10	6.54E-12	-0.028	0.851
52493	1.356	0.235	-1.151	0.221	8.15E-09	2.00E-07	1.10E-09	7.42E-12	0.029	0.844
55032	1.355	0.235	-1.152	0.222	8.43E-09	2.00E-07	1.12E-09	7.63E-12	0.029	0.844
54335	1.355	0.235	-1.152	0.222	8.43E-09	2.00E-07	1.12E-09	7.63E-12	0.029	0.844

SNP loci with the most significant effects in the PO model

SNP	$\hat{\beta}_3$	$se(\hat{\beta}_3)$	$\hat{\gamma}_3$	$se(\hat{\gamma}_3)$	$\hat{\beta}_3 = 0$	$\gamma_3 = 0$	$\hat{\beta}_3 = \gamma_3 = 0$	$\hat{\beta}_{3PO}$	$\beta_{3PO} = 0$	
38136	0.856	0.201	0.426	0.188	2.07E-05	2.38E-02	1.39E-06	0.114	0.623	5.04E-06
9940	-0.974	0.531	0.902	0.168	6.63E-02	8.37E-08	8.48E-06	0.002	0.620	5.64E-06
9732	-0.978	0.530	0.902	0.168	6.49E-02	8.14E-08	8.35E-06	0.002	0.620	5.64E-06
9889	-0.975	0.531	0.902	0.168	6.60E-02	8.34E-08	8.47E-06	0.002	0.620	5.65E-06
24812	-0.373	0.364	-0.789	0.218	3.05E-01	3.04E-04	5.69E-06	0.399	-0.649	6.79E-06
8495	0.984	0.287	0.418	0.235	6.05E-04	7.51E-02	2.28E-06	0.194	0.611	6.93E-06
25940	0.968	0.319	0.439	0.245	2.43E-03	7.35E-02	3.01E-06	0.275	0.614	7.04E-06
31353	0.969	0.319	0.438	0.246	2.41E-03	7.48E-02	3.06E-06	0.274	0.613	7.16E-06
31412	0.969	0.319	0.438	0.246	2.42E-03	7.48E-02	3.06E-06	0.274	0.613	7.17E-06
31253	0.969	0.319	0.438	0.246	2.42E-03	7.48E-02	3.07E-06	0.274	0.613	7.18E-06

SNP loci with the largest P -values for testing the PO assumption

SNP	$\hat{\beta}_3$	$se(\hat{\beta}_3)$	$\hat{\gamma}_3$	$se(\hat{\gamma}_3)$	$\hat{\beta}_3 = 0$	$\gamma_3 = 0$	$\hat{\beta}_3 = \gamma_3 = 0$	$\hat{\beta}_3 = \gamma_3$	$\hat{\beta}_{3PO}$	$\beta_{3PO} = 0$
19200	0.108	0.354	0.108	0.243	0.760	0.656	0.437	1.000	0.113	0.420
33046	0.057	0.429	0.057	0.205	0.894	0.780	0.676	1.000	0.053	0.699
38039	-0.322	0.445	-0.322	0.230	0.470	0.160	0.027	1.000	-0.326	0.025
38133	-0.322	0.446	-0.322	0.230	0.471	0.161	0.027	1.000	-0.326	0.026
18446	0.219	0.251	0.219	0.210	0.382	0.297	0.095	1.000	0.219	0.096
8897	-0.128	0.304	-0.128	0.231	0.674	0.578	0.373	1.000	-0.127	0.378
2124	-0.203	0.285	-0.203	0.242	0.476	0.401	0.172	1.000	-0.193	0.196
26366	0.224	0.279	0.224	0.191	0.421	0.241	0.075	1.000	0.217	0.083

$\hat{\beta}_3$ and $\hat{\gamma}_3$ are the estimates of the long-term genetic effect and short-term genetic effect under our proposed model. $se(\hat{\beta}_3)$ and $se(\hat{\gamma}_3)$ are their estimated standard errors. $\hat{\beta}_{3PO}$ is the estimate of the genetic effect under the PO model. Columns 6, 7, 8, 9 and 11 are the P -values for testing the corresponding null hypotheses.