https://doi.org/10.1016/j.csda.2020.107007

Competing risk modeling and testing for X-chromosome genetic association

Meiling Hao[a], Xingqiu Zhao[b], and Wei Xu[c1]

[a] *School of Statistics, University of International Business and Economics, Beijing 100029, P. R. of China*

[b] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, P. R. of China*

[c] *Department of Biostatistics, Princess Margaret Cancer Centre, Toronto; Dalla Lana School of Public Health, University of Toronto, Canada*

Abstract

The complexity of X-chromosome inactivation arouses the X-linked genetic association being overlooked in most of the genetic studies, especially for genetic association analysis on time to event outcomes. To fill this gap, we propose novel methods to analyze the X-linked genetic association for competing risk failure time data based on a subdistribution hazard function. Specifically, we consider two mechanisms for a single genetic variant on X-chromosome: (1) all the subjects in a population undergo the same inactivation process; (2) the subjects randomly undergo different inactivation processes. According to the assumptions, one of the proposed methods can be used to infer the unknown biological process under scenario (1), while another method can be used to estimate the proportion of a certain biological process in the population under scenario (2). Both of the two methods can infer the direction of skewness for skewed X-chromosome inactivation and derive asymptotically unbiased estimates of the model parameters. The asymptotic distributions for the parameter estimates and constructed score tests with nuisance parameters only presented under the alternative hypothesis are illustrated under both assumptions. Finite sample performance of these novel methods is examined via extensive simulation studies. An ap-

[1]Correspondence to: Department of Biostatistics, Princess Margaret Cancer Centre, Toronto; Dalla Lana School of Public Health, University of Toronto, Canada (E-mail: Wei.Xu@uhnresearch.ca)

plication is illustrated with implementation on a cancer genetic study with competing risk outcomes.

Keywords: Genetic association test, Subdistribution hazard function, X-chromosome association, X-chromosome inactivation.

1. Introduction

The genome-wide association study (GWA study, or GWAS), also known as the whole genome association study (WGA study, or WGAS), has provided a mass of successful approaches to identifying candidate germline genetic variants for diagnosis. It plays an important role in pharmacogenetics to identify prognostic genetic variants that affect overall survival, tumor response or treatment toxicity on many other complex diseases [1]. However, the X-chromosome, which contains almost 5% of the human genome (UCSC Genome Browser) [2], has been generally excluded from the majority of GWAS analyses [3]-[5]. That is mainly because the gene expression on X-chromosome is very complicated. [6] states that to address the imbalance of X-linked genes between males and females, one of the two copies of the X-chromosome genes present in each cell in females may be inactivated during the early embryonic development, which is the so called X-chromosome inactivation (XCI). Studies have suggested that besides XCI, skewed or non-random XCI is also biological plausible. The skewed XCI (denoted as XCI-S) is more common in affected women (namely women with disease) than that in unaffected women; see [7]-[11]. Therefore, it is important to consider XCI-S when assessing the X-chromosome genetic association on disease risk. Another complexity of X-linked genetic effects is the escape from XCI (denoted as XCI-E) outside the pseudo-autosomal regions on the X-chromosome; see [12]-[16]. This kind of gene has no-dosage compensation between males and females and accounts for around 25% X-linked genes.

As the complex inactivation biological process is totally unobserved at most of times, there are only a few statistical approaches, which are all focused on case-control studies, proposed to reveal the relationships between the X-linked genetic variants and various disease risks. [17] and [18] studied the X-chromosome association analysis deeming them only undergoing XCI-E or XCI, respectively. Both of them consider samples of unrelated individuals from a single population. [19] extended the association tests with both single marker and haplotype to the X-chromosome on related individu-

als. [20] proposed different test statistics to study the relationships between X-linked genetic markers and some complex disease traits with related individuals. [21] compared various test statistics regarding genetic markers on X-chromosome. Besides these joint tests for X-linked genetic association, another common approach is to conduct genetic analysis on male and female subjects separately (sex-stratified), see [22]. The limitation of this approach is losing in power caused by partitioning of the samples [23]. [24] proposed a unified likelihood ratio test for X-linked genes over all the potential biological processes, which is first considering XCI-S and can deal with the model misspecification. [25] considered capitalizing on variance heterogeneity due to various factors and predominately the process of X-inactivation. [26] systematically introduced a new software XWAS, which is designed for the X-linked association study and included the recent developed methods. Lately, [27] provided a Bayesian model averaging framework to account for the inherent model uncertainty.

Unfortunately, after an exhaustive review of the literature, methodology developed specifically for X-linked genetic association on time-to-event outcomes is very scarce. [28] first developed a novel statistical model on single X-linked genetic association with right censored data. To analyze the X-linked genetic association for time to event outcomes with the actual process unknown, they proposed a unified approach of maximizing the partial likelihood over all of the potential biological processes. Their proposed method can be used to infer the true biological process and derive asymptotically unbiased estimates of the genetic association parameters. [29] addressed a very important issue in [28] about chi-square distribution assumption. [30] proposed a random effect model to tackle the right censored data. However, in complex human genetic disease research, one may need to deal with more complicated time-to-event outcomes, such as the recurrence-free survival and the cause-specific survival. These involve competing risk events such as death without tumor progression or index cancer. A competing risk event can preclude the event of interest from occurring. It may lose information or provide misleading inference if we ignore this type of event.

To fill this gap, we propose a novel statistical methodology framework to explore the X-linked genetic prognostic and predictive associations on competing risk failure time data. For competing risk data, researchers usually pay attention to the cumulative incidence functions of a specific cause of failure [31]-[35] or the subdistribution hazard function [36]-[43], which is developed directly to model the cumulative incidence function. Thereby, our

goal is to handle the effects of X-linked genetic data on the proportional subdistribution hazard model. In this study, we develop both the uniform XCI model (assuming that for one specific gene, all the subjects in the population undergo the same X-chromosome inactivation biological process) and the random XCI model (assuming that for one specific gene, the subjects randomly go through different biological processes). For uniform XCI model, the proposed method is a unified pseudo-partial likelihood model over all the possible biological processes. Not only can it infer the true biological process, it can also determine the skewed direction and estimate the magnitude of the skewness for XCI-S. For the random XCI model, the proposed method can estimate the proportion of XCI-E and the direction of skewness for XCI-S. Moreover, both methods can provide consistent estimates of the X-linked genetic effects on the time to event outcomes, which is proved to enjoy the asymptotic normality. Further, to test whether the X-linked gene is related to the interest event or not, novel score tests with nuisance parameters that only present under the alternative are constructed with the true value outside the inter point region under the null hypothesis. The asymptotic distribution for the novel score test is derived and a resampling approach is illustrated to derive the reject critical value.

The outline of the remainder of the paper is as follows. We first describe the methodology framework and asymptotic properties of the model in Section 2. Section 3 shows the results from extensive simulations and in Section 4, our proposed methodologies are applied to analyze an X-chromosome wide genetic association study on colorectal cancer patients with competing risk outcomes. Conclusions and discussions are given in Section 5. All the detailed theoretical proofs are provided in the Appendix A of the Supplementary Material, while some further simulation studies are given in the Appendices B, C, and D of the Supplementary Material.

2. Statistical Methodology

Let T be the failure time, Z be the p-vector covariates. Assume that the failure may be aroused by $K \geq 2$ distinct causes. Let C be the censoring time, $Y = \min(T, C)$ be the observation time. $\Delta = 1$ denotes the event of interest occurring, $\Delta = 2$ indicates all the other events happening, and $\Delta = 0$ denotes the censoring happening. We use the Cox proportional hazards model to specify the subdistribution hazard function of the event of interest:

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta_0^\top Z),$$

where β_0 is the true parameter. From [36] and [42], the estimating equation for parametric coefficient β is:

$$U(\beta) = \sum_{i=1}^{n} \mathbf{1}(\Delta_i = 1) \left\{ Z_i - \frac{\sum_{j=1}^{n} w_j(Y_i) Z_j \exp(Z_j^\top \beta)}{\sum_{j=1}^{n} w_j(Y_i) \exp(Z_j^\top \beta)} \right\} = 0, \qquad (1)$$

where $w_j(t) = \mathbf{1}(Y_j \geq t, \Delta_j \neq 2) + \mathbf{1}(\Delta_j = 2)\hat{G}(t)/\hat{G}\{\min(Y_j, t)\}$, $\hat{G}(t)$ is the Kaplan-Meier estimator of $G(t) = Pr(C \geq t)$.

2.1. X-Chromosome Inactivation

For an X-linked genetic variant, such as a Single Nucleotide Polymorphism (SNP), with two alleles: normal allele a and risk allele A, it can be generalized and coded as Table 1 under different biological processes. In the table, γ is an unknown parameter representing the extent of skewness and ranges from 0 to 1. γ less than 0.5 represents XCI-S towards to the normal allele, while γ larger than 0.5 represents XCI-S towards deleterious allele. As suggested by one anonymous reviewer, the coding schemes are added for a better understanding. Table 1 means that no matter which kind of biological process the genes go through, aa on female and a on male have the same effect. Under the biological process of XCI and XCI-S, AA on female and A on male have the same effect. This is reasonable since under XCI and XCI-S, the genes on female X-chromosome will be inactivated to make the dosage compensation between males and females. While under the scenario of XCI-E, gene has no-dosage compensation between males and females, then the effect of AA is double than that of A. So we code AA as 2 and A as 1. This assumption is in accordance with that of Wang et al. (2014). The only difference of genotype coding between ours and theirs is that they assumed that AA on female should have the same effect while we assume that A on male has the same effect under different biological processes. In our model, we include the gender (X_{sex}) of the subjects as one factor, and the effect of gender can not simply be interpreted as the effect of the SNP, since it includes many aspects, such as the immunity difference aroused by gender, which in the majority situation is not determined by the SNP we analyzed. Let X_{snp} belong to a coding value set $\mathcal{X}$ that is in accordance to Table 1 and defined as:

$$\begin{aligned} \mathcal{X} = \Big\{ & X^F_{\text{XCI-S}} = \{0, \gamma, 1\}, \gamma \in [0,1]; X^F_{\text{XCI}} = \{0, 0.5, 1\}; \\ & X^F_{\text{XCI-E}} = \{0, 1, 2\}; X^M = \{0, 1\} \Big\}, \end{aligned}$$

where X^F denotes the coding for female genotypes with different subscript referring the corresponding biological processes and X^M denotes the male genotypes.

To study the relationships between the event of interest and the X-linked genetic marker, we first extent Clayton's methods [18] that assume all the SNPs undergoing XCI and PLINK methods [17] that assume all the SNPs undergoing XCI-E directly. Specifically, denote the log pseudo-partial likelihood as $l(\beta)$:

$$l(\beta) = \sum_{i=1}^{n} \mathbf{1}(\Delta_i = 1)\{Z_i^\top \beta - \log \sum_{j=1}^{n} w_j(Y_i) \exp(Z_j^\top \beta)\},$$

where $Z = (X_{\text{snp}}, X_{\text{sex}})^\top$ with X_{sex} being the gender indicator, $\beta = (\beta_{\text{snp}}, \beta_{\text{sex}})^\top$ with β_{snp} and β_{sex} being the SNP effect and the gender effect, respectively. Denote $\beta_{0,\text{snp}}$ as the elements according to the SNP part. Given X_{snp} as the XCI or XCI-E coding value, respectively, the traditional pseudo-partial likelihood method can be used directly. Maximizing $l(\beta)$ is equivalent to solving equation (1).

Motivated by $l(\beta)$, for XCI-S, to estimate the parameters, we introduce a new function $f(\beta, \gamma)$:

$$\begin{aligned} f(\beta, \gamma) = \sum_{i=1}^{n} \mathbf{1}(\Delta_i = 1)\big(&\{R_i X_{i\text{snp}} + (1 - R_i)\gamma\}\beta_{\text{snp}} + X_{i\text{sex}}\beta_{\text{sex}} \\ &- \log \sum_{j=1}^{n} w_j(Y_i) \exp[\{R_j X_{j\text{snp}} + (1 - R_j)\gamma\}\beta_{\text{snp}} + X_{j\text{sex}}\beta_{\text{sex}}]\big), \end{aligned}$$

where $X_{i\text{snp}}$ is coded under the XCI for subject i, R_i is the genotype indicator for subject i: if genotype is Aa, $R_i = 0$; otherwise, $R_i = 1$. Thus, we have $f(\beta, \gamma_0) = l(\beta)$ for XCI-S with true value γ_0. Denote $(\hat{\beta}^\top, \hat{\gamma})$ as a local maximizer of $f(\beta, \gamma)$. As the biological process is totally unobserved at most of the time, the Akaike information criteria (AIC) is selected to choose the true biological process. Specifically, for XCI and XCI-E, we have $\text{AIC} = -2l(\hat{\beta}) + 2\dim(\hat{\beta})$ with $\dim(z)$ being the dimension for any vector z. For XCI-S, we have: $\text{AIC} = -2f(\hat{\beta}, \hat{\gamma}) + 2\dim(\hat{\beta}^\top, \hat{\gamma})$. Such kind of method is called the unified approach based on the pseudo-partial likelihood.

To study the properties about $(\hat{\beta}^\top, \hat{\gamma})$ under XCI-S, we denote the first and second derivatives of $f(\beta, \gamma)$ as $U(\beta, \gamma)$ and $D(\beta, \gamma)$, respectively. Then

we have that $U(\hat{\beta}, \hat{\gamma}) = 0$. Denote $\tilde{Z}(\beta_{\text{snp}}, \gamma) = (RX_{\text{snp}} + (1-R)\gamma, X_{\text{sex}}, (1-R)\beta_{\text{snp}})^\top$, $Z_\gamma = (RX_{\text{snp}} + (1-R)\gamma, X_{\text{sex}})^\top$. For any vector z, define $z^{\otimes 0} = \mathbf{1}, z^{\otimes 1} = z, z^{\otimes 2} = zz^\top$, where $\mathbf{1}$ is the unit vector. Denote

$$S^{(k)}(t, \gamma, \beta) = \frac{1}{n}\sum_{j=1}^{n} w_j(t) \exp\{Z_{j\gamma}(\beta_{\text{snp}}, \gamma)^\top \beta\} \tilde{Z}_j(\beta_{\text{snp}}, \gamma)^{\otimes k}, k = 0, 1, 2,$$

$$\tilde{S}^{(k)}(t, \gamma, \beta) = \frac{1}{n}\sum_{j=1}^{n} \tilde{w}_j(t) \exp\{Z_{j\gamma}(\beta_{\text{snp}}, \gamma)^\top \beta\} \tilde{Z}_j(\beta_{\text{snp}}, \gamma)^{\otimes k}, k = 0, 1, 2,$$

with $\tilde{w}_j(t) = \mathbf{1}(Y_j \geq t, \Delta_j \neq 2) + \mathbf{1}(\Delta_j = 2)G(t)/[G\{\min(Y_j, t)\}]$. To state the asymptotic properties, define $N(t) = \mathbf{1}(Y \leq t, \Delta = 1)$, $M(t, \beta_0) = \mathbf{1}(Y \leq t, \Delta = 1) - \int_0^t \tilde{w}(s) \exp(\beta_0^\top Z)\lambda_0(s)\, ds$, $M^c(t, \beta_0) = \mathbf{1}(Y \leq t, \Delta = 0) - \int_0^t \mathbf{1}(Y \geq s) d\Lambda^c(s)$ with $\Lambda^c(s)$ being the general cumulative hazard of the censoring time C, $s^{(k)}(t, \gamma, \beta) = E\{\tilde{S}^{(k)}(t, \gamma, \beta)\}, k = 0, 1, 2$, $e(t, \gamma, \beta) = s^{(1)}(t, \gamma, \beta)/s^{(0)}(t, \gamma, \beta)$. Set

$$A = E\left[\int_0^\tau \left\{\frac{s^{(2)}(t, \gamma_0, \beta_0)}{s^{(0)}(t, \gamma_0, \beta_0)} - e(t, \gamma_0, \beta_0)^{\otimes 2}\right\} dN(t)\right], \Sigma = E\{(\eta - \phi)^{\otimes 2}\},$$

$$\eta = \int_0^\tau \left\{\tilde{Z}(\beta_{0,\text{snp}}, \gamma) - e(t, \beta_0, \gamma_0)\right\} dM(t, \beta_0), \phi = \int_0^\tau \frac{q(t)}{E\{\mathbf{1}(Y \geq t)\}} dM^c(t, \beta_0),$$

$$q(s) = \int_0^\tau \left\{\tilde{Z}(\beta_{0,\text{snp}}, \gamma) - e(t, \beta_0, \gamma_0)\right\} \mathbf{1}(t \geq s \geq Y)\, dM(t, \beta_0).$$

Theorem 1. *(Asymptotic Normality) Under the regularity conditions (C1)-(C3) in the Appendix A of the Supplementary Material, for $\gamma_0 \in (0, 1)$, we have that $\sqrt{n}\{\hat{\beta}^\top - \beta_0^\top, \hat{\gamma} - \gamma_0\}^\top$ converges to a normal distribution with mean zero and variance $A^{-1}\Sigma A^{-1}$.*

The proof of Theorem 1 is quite similar to those of Lemma 1 and Theorem 1 of reference [28], thus omitted.

Remark 1: Theorem 1 implies that the convergence rate for both $\hat{\beta}$ and $\hat{\gamma}$ is $n^{-1/2}$. Denote $\asymp$ as the asymptotic equal distribution. This theorem shows that when n is large enough, direct calculations can yield $\hat{\gamma} \in (0, 1)$, which is very important to clinicians. In finite sample performance, to constrain $\hat{\gamma}$

ranging from 0 to 1, the function *nlminb* in software R can be used to get asymptotically unbiased estimates of β and γ with $\hat{\gamma}$ in $(0,1)$. Actually, this is an optimization problem:

$$\begin{cases} \max f(\beta,\gamma) \\ \text{with} \quad \gamma(\gamma-1)\le 0. \end{cases}$$

Through the Karush-Kuhn-Tucker (KKT) conditions, the dual problem of the optimization problem is:

$$\begin{cases} U(\beta,\gamma)=(0,0,\mu(2\gamma-1))^\top, \\ \mu\gamma(\gamma-1)=0, \\ \mu\ge 0. \end{cases} \tag{2}$$

Thereby, for finite sample performance, we have

$$\sqrt{n}\{\hat{\beta}^\top-\beta_0^\top,\hat{\gamma}-\gamma_0\}^\top-\hat{A}^{-1}\{n^{-1/2}(0,0,\hat{\mu}(2\hat{\gamma}-1))^\top\}\asymp N(0,\hat{A}^{-1}\hat{\Sigma}\hat{A}),$$

with $\hat{A}$ being the estimate of A and $\hat{\Sigma}$ being the estimate of Σ. This result can be used to make a correction of $\hat{\beta}$.

Denote

$$\begin{aligned} U_i(\beta) &= \mathbf{1}(\Delta_i=1)\left\{Z_i-\frac{\sum_{j=1}^n w_j(Y_i)Z_j\exp(Z_j^\top\beta)}{\sum_{j=1}^n w_j(Y_i)\exp(Z_j^\top\beta)}\right\}, \\ U_{i,\beta_{\text{snp}}}(\beta) &= \mathbf{1}(\Delta_i=1)\left\{X_{i\text{snp}}-\frac{\sum_{j=1}^n w_j(Y_i)X_{j\text{snp}}\exp(Z_j^\top\beta)}{\sum_{j=1}^n w_j(Y_i)\exp(Z_j^\top\beta)}\right\}, \\ U_{i,\beta_{\text{sex}}}(\beta) &= \mathbf{1}(\Delta_i=1)\left\{X_{i\text{sex}}-\frac{\sum_{j=1}^n w_j(Y_i)X_{j\text{sex}}\exp(Z_j^\top\beta)}{\sum_{j=1}^n w_j(Y_i)\exp(Z_j^\top\beta)}\right\}, \end{aligned}$$

namely $U_{i,\beta_{\text{snp}}}(\beta)$ and $U_{i,\beta_{\text{sex}}}(\beta)$ are the elements of $U_i(\beta)$ according to the β_{snp} and β_{sex} parts. Let

$$\tilde{\beta}_{\text{sex}}=\max_{\beta_{\text{sex}}} l(0,\beta_{\text{sex}}),$$

$\Sigma_{\beta_{\text{snp}}}(0,\tilde{\beta}_{\text{sex}})$ and $\Sigma_{\beta_{\text{sex}}}(0,\tilde{\beta}_{\text{sex}})$ be the asymptotic variances of $U_{i,\beta_{\text{snp}}}(0,\tilde{\beta}_{\text{sex}})$ and $U_{i,\beta_{\text{sex}}}(0,\tilde{\beta}_{\text{sex}})$ respectively, $\Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0,\tilde{\beta}_{\text{sex}})$ be the asymptotic covariance of $U_{i,\beta_{\text{snp}}}(0,\tilde{\beta}_{\text{sex}})$ and $U_{i,\beta_{\text{sex}}}(0,\tilde{\beta}_{\text{sex}})$. To test whether $\beta_{\text{snp}}=0$ or not, we can

use the U-score test [44]. For the extension of Clayton's method ([18]) and the PLINK method ([17]), the U-score test statistic is defined as:

$$\text{Score} = \frac{\{\sum_{i=1}^{n} U_{i0}^{*}(0, \tilde{\beta}_{\text{sex}})\}^2}{n\Sigma_0^{*}(0, \tilde{\beta}_{\text{sex}})},$$

where $U_{i0}^{*}(0, \tilde{\beta}_{\text{sex}}) = U_{i,\beta_{\text{snp}}}(0, \tilde{\beta}_{\text{sex}}) - \Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})\Sigma^{-1}_{\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})U_{i,\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})$,

$$\Sigma_0^{*}(0, \tilde{\beta}_{\text{sex}}) = \Sigma_{\beta_{\text{snp}}}(0, \tilde{\beta}_{\text{sex}}) - \Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})\Sigma^{-1}_{\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})\Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}}),$$

and X_{snp} is the coding value under the XCI or the XCI-E mechanism. It follows from the definition of the test that the "Score" is asymptotic to the central χ_1^2 distribution.

For the XCI-S model, γ is not identifiable under the null hypothesis, and the null model is not an interior point in the alternative space. Thus, the U-score test is defined as:

$$\text{Score}_{\text{XCI-S}} = \max_{\gamma\in[0,1]} \frac{\{\sum_{i=1}^{n} U_{i0,\gamma}^{*}(0, \tilde{\beta}_{\text{sex}})\}^2}{n\Sigma_{\gamma}^{*}(0, \tilde{\beta}_{\text{sex}})} = \frac{\{\sum_{i=1}^{n} U_{i0,\hat{\gamma}}^{*}(0, \tilde{\beta}_{\text{sex}})\}^2}{n\Sigma_{\hat{\gamma}}^{*}(0, \tilde{\beta}_{\text{sex}})},$$

where $U_{i0,\gamma}^{*}(0, \tilde{\beta}_{\text{sex}})$ and $\Sigma_{\gamma}^{*}(0, \tilde{\beta}_{\text{sex}})$ have the same formula as $U_{i0}^{*}(0, \tilde{\beta}_{\text{sex}})$ and $\Sigma_0^{*}(0, \tilde{\beta}_{\text{sex}})$ replacing Z with Z_γ.

Theorem 2. *Under the regularity conditions (C1)-(C3) in the Appendix A of the Supplementary Material, Score$_{XCI\text{-}S}$ converges in distribution to* $\sup_{\gamma\in[0,1]} \mathcal{G}^2(\gamma)$ *under* $\beta_{snp} = 0$ *as* n *goes to infinity, where* $\{\mathcal{G}(\gamma), \gamma \in [0,1]\}$ *is a zero mean Gaussian process with the covariance function*

$$\Sigma(\gamma_1, \gamma_2) = E\{U_{0,\gamma_1}^{*}(0, \beta_{sex})U_{0,\gamma_2}^{*}(0, \beta_{sex})\}/\sqrt{\Sigma_{\gamma_1}^{*}(0, \beta_{sex})\Sigma_{\gamma_2^{*}}(0, \beta_{sex})}.$$

It follows from [36] that

$$U(\beta_0) = \sum_{i=1}^{n} \tilde{U}_i(\beta_0) + o_p(n^{1/2}),$$

with $\tilde{U}_i(\beta_0) = \tilde{\eta}_i(\beta_0) - \tilde{\phi}_i(\beta_0)$, $\tilde{\eta}(\beta_0) = \int_0^{\tau} \left\{ Z - \tilde{e}(t, \beta_0) \right\} dM(t, \beta_0), \tilde{\phi} = \int_0^{\tau} \tilde{q}(t)/E\{\mathbf{1}(Y \geq t)\}\, dM^c(t, \beta_0)$, $\tilde{q}(s)(\beta_0) = \int_0^{\tau} \left\{ Z - \tilde{e}(t, \beta_0) \right\} \mathbf{1}(t \geq s \geq Y)\, dM(t, \beta_0)$, $\tilde{e}(t, \beta_0)(\beta_0) = E\{\tilde{w}(t)\exp(Z^{\top}\beta_0)Z\}/E\{\tilde{w}(t)\exp(Z^{\top}\beta_0)\}$.

Since the biological process is totally unobserved, we propose the test statistic as:

$$\text{Score}_{\text{UMP}} = \max(\text{Score}_{\text{XCI-E}}, \text{Score}_{\text{XCI-S}}).$$

The empirical reject region can be obtained through the resampling approach. Specifically, the resampling approach is by generating a large of $\text{Score}^*_{\text{UMP}}$, where $\text{Score}^*_{\text{UMP}} = \max(\text{Score}^*_{\text{XCI-S}}, \text{Score}^*_{\text{XCI-E}})$. Here

$$\text{Score}^*_{\text{XCI-E}} = \frac{\{\sum_{i=1}^n \varepsilon_i \tilde{U}^*_{i0}(0, \tilde{\beta}_{\text{sex}})\}^2}{n\Sigma^*_0(0, \tilde{\beta}_{\text{sex}})},$$

where $\varepsilon_i, i = 1, 2, \ldots, n$ are i.i.d standard normal random variables independent of the data, and

$$\tilde{U}^*_{i0}(0, \tilde{\beta}_{\text{sex}}) = \tilde{U}_{i,\beta_{\text{snp}}}(0, \tilde{\beta}_{\text{sex}}) - \Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})\Sigma^{-1}_{\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})\tilde{U}_{i,\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}}).$$

Here $\tilde{U}_{i,\beta_{\text{snp}}}(0, \tilde{\beta}_{\text{sex}})$,and $\tilde{U}_{i,\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})$ are the elements of $\tilde{U}_i(0, \tilde{\beta}_{\text{sex}})$ according to the β_{snp} and β_{sex} parts, while $\tilde{U}_{i,\beta_{\text{snp}}}(0, \tilde{\beta}_{\text{sex}})$, $\Sigma_{\beta_{\text{snp}}\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})$, $\Sigma_{\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})$ and $\tilde{U}_{i,\beta_{\text{sex}}}(0, \tilde{\beta}_{\text{sex}})$ derived under the XCI-E mechanism. For

$$\text{Score}^*_{\text{XCI-S}} = \max_{\gamma\in[0,1]} \frac{\{\sum_{i=1}^n \varepsilon_i \tilde{U}^*_{i0,\gamma}(0, \tilde{\beta}_{\text{sex}})\}^2}{n\Sigma^*_\gamma(0, \tilde{\beta}_{\text{sex}})},$$

$\tilde{U}^*_{i0,\gamma}(0, \tilde{\beta}_{\text{sex}})$ has the same formula as $\tilde{U}^*_{i0}(0, \tilde{\beta}_{\text{sex}})$ with replacing Z as Z_γ, $\text{Score}^*_{\text{XCI-S}}$ and $\text{Score}^*_{\text{XCI-E}}$ share the same ε_i. By generating a large of $\text{Score}^*_{\text{UMP}}$, such as 1000 times, we can get the upper αth quantile of the empirical distribution, denoted as $C_{\alpha,\text{UMP}}$. The α-level test reject region is $\{\text{Score}_{\text{UMP}} > C_{\alpha,\text{UMP}}\}$.

2.2. Random X-chromosome Inactivation

The biological process of X-chromosome inactivation is quite complex and sometimes it is inherently subject specific. For a specific gene, it may go through different biological processes in different subjects [45]. We introduce ρ as the indictor: $\rho = 1$ when the gene goes through XCI; $\rho = 2$ when it goes through XCI-E; $\rho = 0$ when it goes through XCI-S. The log pseudo-partial likelihood is: $lp(\beta) = \sum_{i=1}^n l_i(\beta)$, where

$$l_i(\beta) = \mathbf{1}(\Delta_i = 1)\sum_{k=0}^{2} \mathbf{1}(\rho_i = k)\big\{{Z_i^k}^\top\beta - \log[\sum_{j=1}^{n}\sum_{m=0}^{2} \mathbf{1}(\rho_j = m)w_j(Y_i)\exp({Z_j^m}^\top\beta)]\big\},$$

Z^m is the coding for SNPs under the $m(m = 0, 1, 2)$ biological process. The estimating function is

$$Up(\beta) = \sum_{i=1}^{n} \mathbf{1}(\Delta_i = 1) \sum_{k=0}^{2} \mathbf{1}(\rho_i = k) \times \Big\{ Z_i^k - \frac{\sum_{j=1}^{n} \sum_{m=0}^{2} \mathbf{1}(\rho_j = m) w_j(Y_i) \exp({Z_j^m}^\top \beta) Z_j^m}{\sum_{j=1}^{n} \sum_{m=0}^{2} \mathbf{1}(\rho_j = m) w_j(Y_i) \exp({Z_j^m}^\top \beta)} \Big\} = 0.$$

We call it an oracle model as every biological process is known. However, the biological process of X-chromosome inactivation is complex and the true subject-level biological process is totally unobserved. If we assume $P(\rho = m) = p_m, m = 0, 1, 2$, we can get the mean coding value of genotypes Aa and AA, denoted as u_1 and u_2, respectively. It follows from $u_2 = 1 + p_2$ that, if $u_2 > 1$, it means some subjects are undergoing XCI-E. Further, it follows from $u_1 = p_0\gamma + 0.5p_1 + p_2$ that with $p_2 < 1$, $(u_1 - p_2)/(1 - p_2)$ can indicate the direction of skewness. Specifically, $(u_1 - p_2)/(1 - p_2) < 0.5$ shows that the skewness direction is normal allele, while $(u_1 - p_2)/(1 - p_2) > 0.5$ indicates that it skews towards the deleterious allele. Denote $o = 1$ being the genotype of Aa, $o = 2$ being the genotype of AA and $o = 0$ being the other genotypes. Denote $Z_u(u_1, u_2) = ((o = 0)X_{\text{snp}} + (o = 1)u_1 + (o = 2)u_2, X_{\text{sex}})^\top$, $\tilde{Z}_u(\beta_{\text{snp}}, u_1, u_2) = ((o = 0)X_{\text{snp}} + (o = 1)u_1 + (o = 2)u_2, X_{\text{sex}}, (o = 1)\beta_{\text{snp}}, (o = 2)\beta_{\text{snp}})^\top$, where X_{snp} is coded under the XCI,

$$S_u^{(k)}(t, \beta, u_1, u_2) = \frac{1}{n} \sum_{j=1}^{n} w_j(t) \exp\{Z_{ju}(u_1, u_2)^\top \beta\} \tilde{Z}_{ju}(\beta_{\text{snp}}, u_1, u_2)^{\otimes k}, k = 0, 1, 2,$$

$$\tilde{S}_u^{(k)}(t, \beta, u_1, u_2) = \frac{1}{n} \sum_{j=1}^{n} \tilde{w}_j(t) \exp\{Z_{ju}(u_1, u_2)^\top \beta\} \tilde{Z}_{ju}(\beta_{\text{snp}}, u_1, u_2)^{\otimes k}, k = 0, 1, 2.$$

Then the estimating function is

$$eUp(\beta, u_1, u_2) = \sum_{i=1}^{n} \mathbf{1}(\Delta_i = 1) \Big\{ \tilde{Z}_{iu}(\beta_{\text{snp}}, u_1, u_2) - \frac{S_u^{(1)}(Y_i, \beta, u_1, u_2)}{S_u^{(0)}(Y_i, \beta, u_1, u_2)} \Big\} = 0.$$

Define $\hat{\beta}$, $\hat{u}_1$, $\hat{u}_2$ as the solution of $eUp(\hat{\beta}, \hat{u}_1, \hat{u}_2) = 0$. To state the asymptotic properties, we define $s_u^{(k)}(t, \beta, u_1, u_2) = E\{\tilde{S}_u^{(k)}(t, \beta, u_1, u_2)\}$ for $k = 0, 1, 2$,

$$e_u(t,\beta,u_1,u_2) = s_u^{(1)}(t,\beta,u_1,u_2)/s_u^{(0)}(t,\beta,u_1,u_2),$$

$$A_u = E\left[\int_0^\tau \left\{\frac{s_u^{(2)}(t,\beta_0,u_{10},u_{20})}{s_u^{(0)}(t,\beta_0,u_{10},u_{20})} - e_u(t,\beta_0,u_{10},u_{20})^{\otimes 2}\right\} dN(t)\right],$$

$$\Sigma_u = E\{(\eta_u - \phi_u)^{\otimes 2}\},$$

$$\eta_u = \int_0^\tau \left\{\tilde{Z}(\beta_{0,\mathrm{snp}},u_1,u_2) - e_u^{(0)}(t,\beta_0,u_{10},u_{20})\right\} dM(t,\beta_0),$$

$$\phi_u = \int_0^\tau \frac{q(t)_u}{E\{\mathbf{1}(Y \geq t)\}}\, dM^c(t,\beta_0),$$

$$q(s)_u = \int_0^\tau \left\{\tilde{Z}(\beta_{0,\mathrm{snp}},u_1,u_2) - e_u^{(0)}(t,\beta_0,u_{10},u_{20})\right\}\mathbf{1}(t \geq s \geq Y)\, dM(t,\beta_0).$$

Theorem 3. *(Asymptotic Normality) Under the regularity conditions* $(C1)-(C2)$ *and* $(C3')$ *in the Appendix A of the Supplementary Material, we have that* $\sqrt{n}\{\hat{\beta}^\top - \beta_0^\top, \hat{u_1} - u_{10}, \hat{u_2} - u_{20}\}^\top$ *converges to a normal distribution with mean zero and variance* $A_u^{-1}\Sigma_u A_u^{-1}$.

Remark 2: Theorem 3 implies that the convergence rate for the estimates of β, u_1 and u_2 is $n^{-1/2}$. Besides, through the Karush-Kuhn-Tucker (KKT) conditions, the estimating equation is:

$$\begin{cases} eUp(\beta,u_1,u_2) = (0,0,\mu_1(2u_1-1),\mu_2(2u_2-3))^\top, \\ \mu_1 u_1(u_1-1) = 0, \\ \mu_2(u_2-1)(u_2-2) = 0, \\ \mu_i \geq 0, i = 1,2. \end{cases} \tag{3}$$

For finite sample performance, it satisfies that

$$\sqrt{n}\{\hat{\beta}^\top - \beta_0^\top, \hat{u_1} - u_{10}, \hat{u_2} - u_{20}\}^\top - \hat{A}_u^{-1}\{n^{-1/2}(0,0,\mu_1(2\hat{u}_1-1),\mu_2(2\hat{u}_2-3))^\top\} \asymp \hat{A}_u^{-1}\hat{\Sigma}_u\hat{A}_u,$$

with $\hat{A}_u$ being the estimate of A_u.

To test whether $\beta_{\mathrm{snp}} = 0$ or not, we can use the U-score test. As u_1, u_2 are not identifiable, the null model is not an interior point in the alternative space. Thus, the U-score test is defined as:

$$\mathrm{Score}_{\mathrm{RMP}} = \max_{u_1\in[0,1],u_2\in[1,2]} \frac{\{\sum_{i=1}^n U_{i0,u_1,u_2}^*(0,\tilde{\beta}_{\mathrm{sex}})\}^2}{\Sigma_{u_1,u_2}^*(0,\tilde{\beta}_{\mathrm{sex}})} = \frac{\{\sum_{i=1}^n U_{i0,\hat{u}_1,\hat{u}_2}^*(0,\tilde{\beta}_{\mathrm{sex}})\}^2}{\Sigma_{\hat{u}_1,\hat{u}_2}^*(0,\tilde{\beta}_{\mathrm{sex}})},$$

where $U_{i0,u_1,u_2}^*(0,\tilde{\beta}_{\mathrm{sex}})$ and $\Sigma_{u_1,u_2}^*(0,\tilde{\beta}_{\mathrm{sex}})$ have the same formula as $U_{i0}^*(0,\tilde{\beta}_{\mathrm{sex}})$ and $\Sigma_0^*(0,\tilde{\beta}_{\mathrm{sex}})$ replacing Z with $Z_{(u_1,u_2)}$.

Theorem 4. *Under the regularity conditions* $(C1)-(C2)$ *and* $(C3')$ *in the Appendix A of the supplementary material,* $Score_{RMP}$ *converges in distribution to* $\sup_{u_1\in[0,1],u_2\in[1,2]}\mathcal{G}^2(u_1,u_2)$ *under* $\beta_{snp}=0$ *as* n *goes to infinity, where* $\{\mathcal{G}(u_1,u_2),u_1\in[0,1],u_2\in[1,2]\}$ *is a mean zero Gaussian process with the covariance function*

$$\Sigma(\boldsymbol{u}_1,\boldsymbol{u}_2)=E\{U^*_{0,\boldsymbol{u}_1}(0,\beta_{sex})U^*_{0,\boldsymbol{u}_2}(0,\beta_{sex})\}/\sqrt{\Sigma^*_{\boldsymbol{u}_1}(0,\beta_{sex})\Sigma^*_{\boldsymbol{u}_2}(0,\beta_{sex})},$$

where $\boldsymbol{u}_i=(u_{i1},u_{i2})^\top, i=1,2.$

To get critical values for $\text{Score}_{\text{RMP}}$, we use the following resampling approach similar as $\text{Score}_{\text{UMP}}$. Specifically, define

$$\text{Score}^*_{\text{RMP}}=\max_{u_1\in[0,1],u_2\in[1,2]}\frac{\{\sum_{i=1}^n\varepsilon_i\tilde{U}^*_{i0,u_1,u_2}(0,\tilde{\beta}_{\text{sex}})\}^2}{n\Sigma^*_{u_1,u_2}(0,\tilde{\beta}_{\text{sex}})},$$

where $\tilde{U}^*_{i0,u_1,u_2}(0,\tilde{\beta}_{\text{sex}})$ has the same formula as $\tilde{U}^*_{i0}(0,\tilde{\beta}_{\text{sex}})$ with replacing Z as Z_{u_1,u_2}, and $\varepsilon_i, i=1,2,\ldots,n$ are i.i.d standard normal random variables independent of the data. By generating a large of $\text{Score}^*_{\text{RMP}}$, such as 1000 times, we can get the upper αth quantile of the empirical distribution, denoted as $C_{\alpha,\text{RMP}}$. The α-level test reject region is $\{\text{Score}_{\text{RMP}}>C_{\alpha,\text{RMP}}\}$.

3. Simulation Studies

In this section, we assess the finite sample performance of the proposed methods by using simulations. For brevity, we refer the *unified approach of maximizing the pseudo-partial likelihood* as UMP and the random X-chromosome inactivation with the *unified approach of maximizing the pseudo-partial likelihood* as RMP, the extension of *Clayton's* method as CL and the extension of the *PLINK* method as PL, the method with the true biological process is denoted as oracle. Comparisons are conducted across different methods. The survival function of the censoring time is estimated by the Kaplan-Meier method.

In each simulation setting, the sample size is $n=250$ or 500 and the replication times are $N=1000$. Further, to assess the Type I error and the statistical power at 0.05 significance level, the score test is conducted. Besides, to get the empirical critical value for the UMP and the RMP approaches, we resample 1000 times. The computation times for resampling

1000 times for the UMP and the RMP methods are 347.90 seconds and 570.54 seconds, respectively. Suppose the censoring time follows a uniform distribution $unif(0, c)$, with c chosen to yield 20% censoring rate. The minor allele frequency (MAF) is set at 0.4, the female and male rate is 1 : 1. Let "Bias" be the sample mean of the estimate minus the true value, "SSE" denote the sampling standard error of the estimates, "ESE" denote the sample mean of the estimated standard error. To save space, we mostly demonstrate the results with sample size $n = 250$, and the results with sample size $n = 500$ are in the Appendix B of the Supplementary Material. Besides, as suggested by one anonymous reviewer, we added the simulation results for 40% censoring rate in the Appendix C of the Supplementary Material, and the female: male=3:1 in the Appendix D of the Supplementary Material.

Scenario 1. Under this scenario, we assume all the SNPs in the population following from the same kind of X-chromosome inactivation. To compare the RMP, the UMP, the CL, and the PL approaches, we consider three biological processes: XCI-E, XCI and XCI-S. For the XCI-S model, the value for γ is set at 0.9 as a population level parameter. A proportional subdistribution hazard model can be obtained by defining the subdistribution for the event of interest as:

$$F(t, X_{\text{snp}}, X_{\text{sex}}) = 1 - [1 - q_0\{1 - \exp(-t)\}]^{\exp(c_1 + X_{\text{snp}}\beta_{\text{snp}} + \beta_{\text{sex}}X_{\text{sex}})}. \quad (4)$$

Then we can get the desired subdistribution hazard model as:

$$\lambda(t|X) = \lambda_0(t)\exp(X_{\text{snp}}\beta_{\text{snp}} + \beta_{\text{sex}}X_{\text{sex}}).$$

For the competing event, it is generated from the following model:

$$\lambda_{\text{com}}(t|X) = \lambda_{\text{com},0}(t)\exp(X_{\text{snp}}\beta_{\text{com,snp}} + \beta_{\text{com,sex}}X_{\text{sex}}), \quad (5)$$

where X_{snp} represents the SNP genotype, and X_{sex} represents the gender covariate. Specifically, denote $X_{\text{sex}} = 0$ as female and $X_{\text{sex}} = 1$ as male. $\beta_{\text{sex}} = -0.5$ and $\beta_{\text{snp}} = 1$. Set $\beta_{\text{com,sex}} = 0.5$ and $\beta_{\text{com,snp}} = 0, 0.5$. $q_0 = 0.2$, c_1 and $\lambda_{\text{com},0}(t)$ are chosen to yield 20% competing event rate. We will use S1 to denote the situation with $\beta_{\text{com,snp}} = 0$ and S2 to denote that with $\beta_{\text{com,snp}} = 0.5$ in the following tables.

Table 2 displays the Bias, SSE and ESE of the estimate of β and the percentage of different selected biological models (Sel_Mod). The results indicate that the UMP method has a robust performance in selecting true

biological models through AIC. Besides, the UMP and the RMP methods provide an asymptotically unbiased estimate of β, while both the CL and the PL method are biased with model mis-specification. Furthermore, ESE is very close to SSE, which means the variance estimation is reasonable. Tables 2 and 3 indicate that the estimate of p_2 is very close to the UMP selection percentage of XCI-E, and the estimate of $(u_1 - p_2)/(1 - p_2)$ can show the skewness direction. This result suggests that the UMP and the RMP can provide similar conclusion about the biological mechanism.

To demonstrate the power of test statistics, we do some simulations with $\beta_{\text{snp}} = 0, 0.2, 0.4, 0.6, 0.8$. From the estimated Type I error and power results with different biological models in Table 4, we conclude that the UMP and the RMP are comparable to each other. If the underline models are misspecified, the CL and the PL are generally less powerful than the UMP and the RMP. Moreover, as expected, the UMP method is generally powerful in detecting the significant genetic association under the XCI-S; under the XCI process, the CL has better performance than the PL; while the PL outperforms the CL under the XCI-E process.

Scenario 2. This scenario is for random inactivation. Under this scenario, 20% SNPs follow from XCI-S with $\gamma = 0.1$ or $\gamma = 0.9$, 60% of SNPs undergo XCI while the remaining SNPs undergo XCI-E. Other settings are exactly the same as Scenario 1. Through direct calculations, we get $p_2 = 0.2$ and $(u_1 - p_2)/(1 - p_2) = 0.4$ with $\gamma = 0.1$ and $(u_1 - p_2)/(1 - p_2) = 0.6$ with $\gamma = 0.9$. We denote $\gamma = 0.9$ as Case 1 while $\gamma = 0.1$ as Case 2. Table 5 shows that under the scenario of random inactivation, the RMP method can provide an asymptotically unbiased estimate for β. Besides, the ESE is very close to the SSE, which implies that Theorem 3 is valid. Further, Table 6 are the estimates of p_2 and $(u_1 - p_2)/(1 - p_2)$. The asymptotically unbiased estimation results indicate that our proposed RMP method can infer the true biological inactivation mechanism. Table 7 displays the estimated Type I error and power under the significance level of 5%. The results imply that, the RMP method is comparable to the UMP while they two are more powerful than the CL and the PL. The CL approach is more powerful than the PL method under this scenario since the XCI occupies 60% while the XCI-E only occupies 20%. Furthermore, Scenarios 1 and 2 indicate that the UMP is generally as powerful as the RMP but it cost less time than the RMP, so we suggest to use the UMP to select the significant genes first. If the XCI-E or the XCI-S is the selected biological process of the identified significant gene, we prefer to use the RMP to reestimate the parameters.

Scenario 3. This scenario focuses on parameter estimation. The subdistribution hazard is:

$$\lambda(t|X) = \lambda_0(t)\exp(X_{\text{snp1}}\beta_{\text{snp1}} + X_{\text{snp2}}\beta_{\text{snp2}} + \beta_{\text{sex}}X_{\text{sex}} + \beta_{c1}X_{c1} + \beta_{c2}X_{c2}),$$

where $\beta_{\text{snp1}} = 1.0, \beta_{\text{snp2}} = -1.0, \beta_{\text{sex}} = 0.5, \beta_{c1} = -0.5, \beta_{c2} = 0.5$, $X_{ci}, i = 1, 2$ follow the standard normal distribution. For SNP1, the MAF=0.4, 20% SNPs follow XCI-S with $\gamma = 0.9$, 60% of SNPs undergo XCI while the remaining SNPs are undergoing XCI-E. For SNP2, the MAF=0.25 and 20% SNPs follow XCI-S with $\gamma = 0.1$, 50% of SNPs undergo XCI while the remaining SNPs are undergoing XCI-E. For SNP1, we have $p_2 = 0.2$ and $(u_1 - p_2)/(1 - p_2) = 0.6$. For SNP2, we derive $p_{22} = 0.3$ and $(u_{21} - p_{22})/(1 - p_{22}) = 0.3857$. Other settings are exactly the same as Scenario 1.

Table 8 implies that under the scenario of random inactivation, the RMP method can provide an asymptotically unbiased estimate for β while the CL and the PL approach are biased. Besides, the ESE is very close to the SSE, which is in accordance to Theorem 3. Further, Table 9 displays the estimates of $(p_2, (u_1 - p_2)/(1 - p_2))$ for SNP1 and $(p_{22}, (u_{21} - p_{22})/(1 - p_{22}))$ for SNP2. The estimates are asymptotically unbiased and this demonstrates that our proposed RMP method can indicate the right biological inactivation mechanism.

4. Application

In this section, we apply our proposed methods to analyze several X-linked genes on the recurrence-free survival (RFS) of colorectal cancer patients with metastatic disease (mCRC), which is the second most diagnosed cancer for both males and females ([28]). The main event of interest is recurrence, death without recurrence is treated as the competing risk event. There are a total of 502 mCRC patients in this study, 242 among them were given Brivanib and Cetuximab treatment (TmT=1), while the remaining were given placebo and Cetuximab treatment (TmT=2). The genetic data was genotyped using Chip HumanOmniExpressExome-8v1.2. The Principal Component Analysis (PCA) was conducted by using Eigensoft 5.0.2 on autosomal SNPs. Among the total of 502 patients, the median RFS time is 5.06 (95% CI: 3.71, 5.36) months. In the real data analysis, there exist 169 female subjects ($X_{\text{sex}} = 0$) and 333 male subjects ($X_{\text{sex}} = 1$).

We apply the proposed UMP and RMP methods to assess the genetic association on each X-linked SNP. To get the empirical p-value, we use the

permutation method with 10,000 permutations. The CL and the PL methods are also applied. Besides the gender information, there are three covariates that are included in the genetic association models such as treatment (TmT), PC1 and PC2. The PC1 and PC2 are the two major principal components estimated from the autosomal SNPs.

To save space, we just provide the detailed analysis information for two significant X-linked SNPs in Table 10. From the table, we can find that, for both SNPs, the UMP and the RMP methods have identified the same biological process that may be overlooked by the CL or the PL method with wrong model specification. The table indicates that both the RMP and the UMP methods suggest that "rs1997625" goes through XCI-E while " rs7059265" goes through XCI-S with skewness direction towards normal allele. The conclusion is in accordance to the simulation results, namely, the RMP can derive very similar results to the UMP. Besides, Table 10 shows that the coefficient of the SNP of "rs1997625" is negative while that of " rs7059265" is positive. This indicates subjects with allele A survive longer than that with allele a for the SNP "rs1997625". But for the SNP " rs7059265", the subjects with allele A survive shorter than that with allele a. Furthermore, the coefficient of treatment β_{TmT} being positive shows that the Brivanib and Cetuximab treatment (TmT=1) works better than placebo and Cetuximab treatment (TmT=2). To give an intuitive and clear impression, we have added the estimate of the survival function for the subjects based on our RMP (or UMP) method, see Figures 1 and 2. Specifically, we plot the survival function for the subjects with different genotypes and treatments for $PC1 = PC2 = 0$.

5. Discussion

In this paper, we have proposed four X-chromosome analysis methods on competing risk outcomes. The proposed UMP and RMP methods can detect the genetic association of genetic variants on the X-chromosome and infer the undergoing biological process. The obvious advantage of the UMP and the RMP methods is that they can derive asymptotically unbiased estimates for the association effect and indicate the correct inactivation direction and magnitude. In general, compared to the CL and the PL, the UMP and the RMP are generally more powerful. For the X-chromosome wide genome scan, as the CL and the PL are computationally effective and straightforward to implement, we suggest to use them for preliminary screening and identify-

ing potential genetic association signals, and apply the UMP and the RMP for further assessment including parameter estimation and biological process specification.

The proposed pharmacogenomics methods aim to understand how X-linked genetic variants influence treatment efficacy. Such studies can reveal how genetic variation across individuals affects a drug's pharmacokinetics and pharmacodynamics. If the associations of genotypes with drug-related phenotypes are reproducible and have large effect sizes, clinical use of genetic information can be implemented for patients' benefit. This is particularly important in oncology research because cancer is a leading cause of morbidity and mortality in industrialized nations, and failed treatment is often life-threatening. To predict how a cancer patient will respond to a particular treatment regimen is our next goal of personalized oncology.

Score tests with nuisance parameters present only under the alternatives are proposed, and the empirical rejective region is provided. However, this will result in some computation burden especially in handling the whole X-chromosome genes. Thereby, to provide an exact or sharper bound for the maxima stochastic process is put on the agenda. However, it's a pity that as pointed out in [46], any exact computations about the distribution about $P(\max_{s\in\mathcal{S}}\{\mathcal{G}^2(s)\} \geq t)$ is almost impossible with $\mathcal{S}$ being a topology space, and [47] gives the asymptotic distribution under some regularity conditions. Combing the results and the definition of the upcrossing, the asymptotic shaper bound for the probability of $P(\max_{s\in\mathcal{S}}\{\mathcal{G}^2(s)\} \geq t)$ can be derived at some specific points, which is asymptotic equivalent to that given by [48] with $\mathcal{S} = [0, 1]$.

For the parameterizations of genetic effects on marker genotypes, [49] mainly focused on coding genotypes for genetic markers with multiple alleles on autosome. As suggested by one anonymous reviewer, the model can be extended to model genetic variants on X-chromosome.

Acknowledgements

Xu's research is partly supported by the Canadian Institutes of Health Research (CIHR, Grant No. 145546). Zhao's research is supported in part by the Research Grants Council of Hong Kong (No. PolyU 15301218), the National Natural Science Foundation of China (No. 11771366) and The Hong Kong Polytechnic University. Hao's research is supported by the Program

for Young Excellent Talents, UIBE (No. 19YQ15) and the National Natural Science Foundation of China (No. 11901087).

Supplementary Material

The proof and extensive simulation studies are available at Supplementary Material.

[1] Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *The New England Journal of Medicine*. 2011; 364: 1144-1153.

[2] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Research*. 2002; 12: 996-1006.

[3] Erdmann J, Grosshennig A, Braund PS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22. *Nature Genetics*. 2009; 41: 280-282.

[4] Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *The New England Journal of Medicine*. 2007; 357: 443-453.

[5] Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*. 2009; 41: 56-65.

[6] Lyon MF. Gene action in the X-chromosome of the mouse. (Mus musculus L.). *Nature*. 1961; 190: 372-373.

[7] Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *The American Society of Human Genetics*. 2002; 71: 168-173.

[8] Talebizadeh Z, Bittel DC, Veatch OJ, Kibiryeva N, Butler MG. Brief report: nonrandom X chromosome inactivation in females with autism. *The Journal of Autism and Developmental Disorders*. 2005; 35: 675-681.

[9] Chabchoub G, Uz E, Maalej A, Mustafa CA, Rebai A, Mnif M, Bahloul Z, Farid NR, Ozcelik T, Ayadi H. Analysis of skewed X-chromosome inactivation in females with rheumatoid arthritis and autoimmune thyroid diseases. *Arthritis Research and Therapy*. 2009; 11: 1-8.

[10] Buller RE, Sood AK, Lallas T, Buekers T, Skilling JS. Association between nonrandom X-chromosome inactivation and BRCA1 mutation in germline DNA of patients with ovarian cancer. *The Journal of the National Cancer Institute*. 1999; 91: 339-346.

[11] Kristiansen M, Langerod A, Knudsen GP, Weber BL, Borresen-Dale AL, Orstavik KH. High frequency of skewed X inactivation in young breast cancer patients. *Journal of Medical Genetics*. 2002; 39: 30-33.

[12] Brown CJ, Carrel L, Willard HF. Expression of genes from the human active and inactive X chromosomes. *American Journal of Medical Genetics*. 1997; 60: 1333-1343.

[13] Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005; 434: 400-404.

[14] Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD. Genomic environment predicts expression patterns on the human inactive X chromosome. *PLOS Genetics*. 2006; 2: 1477-1486.

[15] Miller AP, Willard HF. Chromosomal basis of X chromosome inactivation: identification of a multigene domain in Xp11.2p11.22 that escapes X inactivation. *Proceedings of the National Academy of Sciences*. 1998; 95: 8709-8714.

[16] Willard HF. The sex chromosomes and X chromosome inactivation. *The Metabolic and Molecular Bases of Inherited Disease*. 1995; 719-737.

[17] Zheng G, Joo J, Zhang C, Geller NL. Testing association for markers on the X-chromosome. *Genetic Epidemiology*. 2007; 31: 834-843.

[18] Clayton D. Testing for association on the X-chromosome. *Biostatistics*. 2008; 9: 593-600.

[19] Browning SR, Briley JD, Briley LP, Chandra G, Charnecki JH, Ehm MG, Johansson KA, Jones BJ, Karter AJ, Yarnall DP, Wagner MJ.

Case-control single-marker and haplotypic association analysis of pedigree data. *Genetic Epidemiology*. 2005; 28: 110-122.

[20] Thornton T, Zhang Q, Cai X, Ober C, McPeek MS. XM: association testing on the X-Chromosome in case-control samples with related individuals. *Genetic Epidemiology*. 2012; 36: 438-450.

[21] Loley C, Ziegler A, Konig IR. Association tests for X-chromosomal markers-a comparison of different test statistics. *Human Heredity*. 2011; 71: 23-36.

[22] Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, Billing-Ross P, Madar A, Spritz R, Keinan A. Accounting for eXentricities: analysis of the X-chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLOS One*. 2014; 9: 1-31.

[23] Behrens G, Winkler TW, Gorski M, Leitzmann MF, Heid IM. To stratify or not to stratify: power considerations for population-based genome-wide association studies of quantitative traits. *Genetic Epidemiology*. 2011; 35: 867-879.

[24] Wang J, Yu R, Shete S. X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genetic Epidemiology*. 2014; 38: 483-493.

[25] Ma L, Hoffman G, Keinan A. X-inactivation informs variance-based testing for X-linked association of a quantitative trait. *BMC Genetics*. 2015; 16: 777-780.

[26] Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z, Keinan A. XWAS: a toolset for genetic data analysis and association studies of the X-chromosome. *Journal of Heredity*. 2015; 106: 666-671.

[27] Chen B, Craiu RV, Sun L. Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *Biostatistics*, in press.

[28] Xu W, Hao, M. A unified partial likelihood approach for X-chromosome association studies on time to event outcomes. *Genetic Epidemiology*. 2018; 42: 80-94.

[29] Xu W, Hao, M. Partial likelihood ratio test for X-chromosome association models. *Genetic Epidemiology*. 2018; 42: 846-848.

[30] Han D, Hao M, Qu L, Xu W. (2019). A novel model for the X-chromosome inactivation association on survival data. Statistical methods in medical research, 0962280219859037.

[31] Cheng SC, Fine JP, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998; 54: 219-228.

[32] Shen Y, Cheng SC. Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*. 1999; 55: 1093-1100.

[33] Scheike TH, Zhang MJ. An additive multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics*. 2002; 29: 75-88.

[34] Scheike TH, Zhang MJ. Extensions and applications of the Cox-Aalen survival model. *Biometrics*. 2003; 59: 1036-1045.

[35] Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005; 61: 223-229.

[36] Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing Risk. *The Annals of Statistics*. 1988; 16: 1141-1154.

[37] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94, 496-509.

[38] Sun LQ, Liu JX, Sun JG, Zhang MJ. Modeling the subdistribution of a competing risk. *Statistica Sinica*. 2006; 16: 1367-1385.

[39] Latouche A, Boisson V, Chevret S, Porcher R. Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*. 2007; 26: 965-974.

[40] Beyersmann J, Schumacher M. Time-depedent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics*. 2008; 9: 765-776.

[41] Zhang X, Zhang MJ, Fine J. A proportional hazards regression model for the subdistribution with right-censored and left-truncated competing risks data. *Statistics in Medicine*. 2011; 30: 1933-1951.

[42] Donoghoe MW, Gebski V. The importance of censoring in competing risks analysis of the subdistribution hazard. *BMC Medical Research Methodology*. 2017; 17: 1-11.

[43] Bellach A, Kosorok MR, Rschendorf L, Fine JP. Weighted NPMLE for the subdistribution of a competing risk. *Journal of the American Statistical Association*, in press.

[44] Hall WJ, Mathiason DJ. On large-sample estimation and testing in parametric models. *International Statistical Review*. 1990; 77-97.

[45] Krumlauf R, Chapman VM, Hammer RE, Brinster R, Tilghman S-M. Differential expression of α-fetoprotein genes on the inactive X-chromosome in extraembryonic and somatic tissues of a transgenic mouse line. *Nature*. 1986; 319: 224-226.

[46] Hasofer AM. Upcrossings of random fields. *Advances in Applied Probability*. 1978; 10: 14-21.

[47] Worsley KJ. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*. 1994; 26, 13-42.

[48] Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1987; 74: 33-43.

[49] Wang T. On coding genotypes for genetic markers with multiple alleles in genetic association study of quantitative traits[J]. *BMC Genetics*. 2011; 12: 82.

Table 1: Coding for the genotypes on the X-chromosome

	Female genotypes			Male genotypes	
Biologicl Process	aa	Aa	AA	a	A
XCI-E	0	1	2	0	1
XCI	0	0.5	1	0	1
XCI-S	0	γ	1	0	1

Table 2: Estimation results of parameters for Scenario 1 with censoring rate being 20% and $n = 250$

	Bio_Mod[1]	Methods	Sel_Mod[2]	$\beta_{\text{snp}} = 1$			$\beta_{\text{sex}} = 0.5$			per[3]
				Bias	ESE	SSE	Bias	ESE	SSE	
S1	XCI-E	UMP	XCI-E	0.0104	0.1334	0.1415	0.0027	0.1672	0.1711	88.7
			XCI	0.4356	0.1975	0.2069	-0.4825	0.1668	0.1486	8.2
			XCI-S	0.4792	0.2053	0.2511	-0.3575	0.2005	0.3062	3.1
		RMP		0.0114	0.2497	0.2530	0.0045	0.2691	0.2614	
		CL		0.3891	0.2027	0.2093	-0.4785	0.1703	0.1700	
		PL		0.0050	0.1335	0.1411	0.0017	0.1675	0.1684	
	XCI	UMP	XCI-E	-0.3010	0.1297	0.1351	0.3255	0.1661	0.1710	17.3
			XCI	0.0140	0.1876	0.1911	0.0034	0.1577	0.1587	67.3
			XCI-S	0.0361	0.1880	0.1980	0.0058	0.1912	0.2655	15.4
		RMP		0.0139	0.2354	0.2426	0.0018	0.2574	0.2531	
		CL		0.0074	0.1877	0.1912	0.0031	0.1581	0.1599	
		PL		-0.3504	0.1305	0.1342	0.3099	0.1697	0.1710	
	XCI-S	UMP	XCI-E	-0.2746	0.1285	0.1344	0.1205	0.1658	0.1604	13.0
			XCI	0.0634	0.1896	0.1978	-0.1855	0.1605	0.1606	22.6
			XCI-S	0.0014	0.1939	0.1987	0.0577	0.1960	0.1807	66.4
		RMP		0.0130	0.2435	0.2483	0.0025	0.2656	0.2561	
		CL		0.0357	0.1887	0.192	-0.1956	0.1618	0.1588	
		PL		-0.3436	0.1289	0.1342	0.1198	0.1687	0.1679	
S2	XCI-E	UMP	XCI-E	0.0104	0.1331	0.1418	0.0022	0.1665	0.1698	89.2
			XCI	0.4252	0.1962	0.2119	-0.4776	0.1661	0.1467	7.9
			XCI-S	0.4694	0.2045	0.2162	-0.3465	0.200	0.2994	2.9
		RMP		0.0089	0.2485	0.2509	0.0054	0.2680	0.2584	
		CL		0.3868	0.2019	0.2088	-0.4777	0.1697	0.1689	
		PL		0.0044	0.1330	0.1411	0.00124	0.1668	0.1672	
	XCI	UMP	XCI-E	-0.3065	0.1300	0.1361	0.3223	0.1670	0.1718	17.5
			XCI	0.0141	0.1885	0.1925	0.0044	0.1585	0.1592	67.7
			XCI-S	0.0388	0.1886	0.1932	0.0058	0.1918	0.2694	14.8
		RMP		0.0120	0.2365	0.2431	0.0031	0.2594	0.2538	
		CL		0.0064	0.1885	0.1914	0.0032	0.1589	0.1598	
		PL		-0.3509	0.1311	0.1343	0.3106	0.1706	0.1716	
	XCI-S	UMP	XCI-E	-0.2700	0.1279	0.1392	0.1259	0.1652	0.1708	12.8
			XCI	0.0613	0.1883	0.1974	-0.1824	0.1596	0.1567	22.1
			XCI-S	-0.0004	0.1926	0.1954	0.05459	0.1949	0.1799	65.1
		RMP		0.0101	0.2417	0.2461	0.0043	0.2638	0.254	
		CL		0.0350	0.1875	0.191	-0.1951	0.1608	0.1581	
		PL		-0.3432	0.1283	0.1336	0.1198	0.1677	0.1676	

1: The true biological models; 2: The selected biological models;
3: the percentage of selected biological models.
S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.

Table 3: Estimation of $\big(p_2, (u_1 - p_2)/(1 - p_2)\big)$ for Scenario 2 with censoring rate being 20%

n	Bio_Mod[1]	n=250	$\frac{u_1-p_2}{1-p_2}$	p_2	n=500	$\frac{u_1-p_2}{1-p_2}$	p_2
S1	XCI-E		0.4061	0.7681		0.3405	0.8444
	XCI		0.4856	0.1806		0.4802	0.1343
	XCI-S		0.8900	0.1148		0.8992	0.0875
S2	XCI-E		0.4027	0.7704		0.3402	0.8450
	XCI		0.4862	0.1822		0.4811	0.1340
	XCI-S		0.8902	0.1154		0.8987	0.0879

1: The true biological models.
S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.

Table 4: Estimated size and power for Scenario 1 with censoring rate being 20% and $n = 250$

	Bio_Mod[1]	Methods	$\beta_{\text{snp}} = 0$	$\beta_{\text{snp}} = 0.2$	$\beta_{\text{snp}} = 0.4$	$\beta_{\text{snp}} = 0.6$	$\beta_{\text{snp}} = 0.8$
S1	XCI-E	UMP	0.066	0.361	0.875	0.992	1.000
		RMP	0.064	0.356	0.869	0.994	1.000
		CL	0.057	0.320	0.856	0.992	1.000
		PL	0.052	0.356	0.876	0.996	1.000
	XCI	UMP	0.066	0.205	0.597	0.902	0.993
		RMP	0.064	0.207	0.581	0.898	0.990
		CL	0.057	0.191	0.592	0.894	0.992
		PL	0.052	0.186	0.542	0.869	0.984
	XCI-S	UMP	0.066	0.221	0.672	0.939	0.996
		RMP	0.064	0.211	0.641	0.927	0.995
		CL	0.057	0.206	0.642	0.916	0.996
		PL	0.052	0.203	0.598	0.906	0.992
S2	XCI-E	UMP	0.066	0.362	0.871	0.992	1.000
		RMP	0.062	0.354	0.867	0.994	1.000
		CL	0.059	0.323	0.854	0.991	1.000
		PL	0.067	0.356	0.879	0.996	1.000
	XCI	UMP	0.066	0.211	0.607	0.908	0.993
		RMP	0.065	0.200	0.581	0.894	0.990
		CL	0.060	0.192	0.589	0.899	0.993
		PL	0.065	0.181	0.529	0.864	0.983
	XCI-S	UMP	0.066	0.224	0.670	0.939	0.997
		RMP	0.065	0.212	0.636	0.922	0.996
		CL	0.060	0.209	0.646	0.918	0.996
		PL	0.065	0.208	0.597	0.903	0.992

1: The true biological models.
S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.

Table 5: Estimation results of parameters for Scenario 2 with censoring rate being 20% and $n = 250$

	Bio_Mod[1]	Methods	Sel_Mod[2]	$\beta_{\text{snp}} = 1$			$\beta_{\text{sex}} = 0.5$			per[3]
				Bias	ESE	SSE	Bias	ESE	SSE	
S1	Case 1	UMP	XCI-E	-0.2506	0.1349	0.1392	0.2203	0.1710	0.1673	15.0
			XCI	0.0846	0.1943	0.2034	-0.1235	0.1642	0.1703	56.3
			XCI-S	0.0911	0.1985	0.2053	0.0278	0.2012	0.2335	28.7
		RMP		0.0176	0.2450	0.2488	-0.0060	0.2695	0.2731	
		CL		0.0768	0.1947	0.2005	-0.1284	0.1649	0.1659	
		PL		-0.2959	0.1348	0.1408	0.2171	0.1727	0.1703	
		oracle		0.0187	0.1652	0.1712	-0.0070	0.1624	0.1624	
	Case 2	UMP	XCI-E	-0.2548	0.1356	0.1395	0.3053	0.1714	0.1646	26.9
			XCI	0.0688	0.1938	0.2027	-0.0351	0.1630	0.1659	63.0
			XCI-S	0.1018	0.1962	0.1907	-0.1956	0.1990	0.2567	10.1
		RMP		0.0141	0.2421	0.2464	-0.004	0.2663	0.2712	
		CL		0.0578	0.1945	0.1993	-0.0419	0.1638	0.1644	
		PL		-0.3024	0.1360	0.1410	0.2980	0.1737	0.1705	
		oracle		0.0162	0.1634	0.1704	-0.0068	0.1619	0.1626	
S2	Case 1	UMP	XCI-E	-0.2512	0.1349	0.1390	0.2211	0.1711	0.1682	14.9
			XCI	0.0851	0.1942	0.2032	-0.1243	0.1640	0.1699	56.2
			XCI-S	0.0907	0.1983	0.2050	0.0276	0.2011	0.2341	28.9
		RMP		0.0174	0.2449	0.2483	-0.0060	0.2695	0.2731	
		CL		0.0767	0.1946	0.2002	-0.1285	0.1648	0.1657	
		PL		-0.2959	0.1347	0.1406	0.2170	0.1726	0.1703	
		oracle		0.0186	0.1651	0.1710	-0.0070	0.1623	0.1622	
	Case 2	UMP	XCI-E	-0.2542	0.1357	0.1403	0.3068	0.1714	0.1661	10.4
			XCI	0.0687	0.1937	0.2026	-0.0360	0.1630	0.1655	62.9
			XCI-S	0.1015	0.1960	0.1880	-0.1886	0.1988	0.2566	26.7
		RMP		0.0141	0.2419	0.2461	-0.0036	0.2663	0.2712	
		CL		0.0579	0.1944	0.1992	-0.0419	0.1637	0.1643	
		PL		-0.3023	0.1359	0.1409	0.2980	0.1736	0.1702	
		oracle		0.0163	0.1634	0.1703	-0.0069	0.1618	0.1625	

1: The true biological models; 2: The selected biological models;
3: the percentage of selected biological models.
S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.
Case 1: $\gamma = 0.9$; Case 2: $\gamma = 0.1$.

Table 6: Estimation of $\big(p_2, (u_1 - p_2)/(1 - p_2)\big)$ for Scenario 2 with censoring rate being 20%

n		Methods	$\frac{u_1-p_2}{1-p_2} = 0.6$	$p_2 = 0.2$	$\frac{u_1-p_2}{1-p_2} = 0.4$	$p_2 = 0.2$
n=250	S1	RMP	0.6166	0.2542	0.4004	0.2753
n=500		RMP	0.6018	0.2122	0.3962	0.2206
n=250	S2	RMP	0.6165	0.2542	0.4004	0.2753
n=500		RMP	0.6020	0.2119	0.3965	0.2203

S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.
Case 1: $\gamma = 0.9$; Case 2: $\gamma = 0.1$.

Table 7: Estimated size and power for Scenario 2 with censoring rate being 20%, $n = 250$

	Bio_Mod[1]	Methods	$\beta_{\text{snp}} = 0$	$\beta_{\text{snp}} = 0.2$	$\beta_{\text{snp}} = 0.4$	$\beta_{\text{snp}} = 0.6$	$\beta_{\text{snp}} = 0.8$
S1	Case 1	UMP	0.047	0.244	0.666	0.940	0.990
		RMP	0.052	0.240	0.664	0.937	0.991
		CL	0.045	0.226	0.662	0.940	0.991
		PL	0.048	0.210	0.643	0.918	0.991
		oracle	0.047	0.252	0.733	0.964	0.997
	Case 2	UMP	0.047	0.240	0.643	0.927	0.990
		RMP	0.052	0.242	0.655	0.925	0.991
		CL	0.045	0.216	0.642	0.931	0.987
		PL	0.048	0.199	0.608	0.897	0.984
		oracle	0.047	0.264	0.724	0.965	0.998
S2	Case 1	UMP	0.051	0.240	0.664	0.930	0.991
		RMP	0.056	0.251	0.667	0.927	0.991
		CL	0.048	0.223	0.661	0.929	0.992
		PL	0.055	0.213	0.644	0.908	0.989
		oracle	0.051	0.258	0.729	0.957	0.997
	Case 2	UMP	0.050	0.259	0.672	0.933	0.991
		RMP	0.061	0.269	0.669	0.930	0.992
		CL	0.044	0.213	0.640	0.929	0.985
		PL	0.047	0.201	0.610	0.903	0.985
		oracle	0.046	0.253	0.728	0.961	0.998

1: The true biological models.
S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.
Case 1: $\gamma = 0.9$; Case 2: $\gamma = 0.1$.

Table 8: Estimation results of parameters for Scenario 3 with censoring rate being 20% and $n = 250$

	Methods	$\beta_{\text{SNP1}} = 1$			$\beta_{\text{SNP2}} = -1$			$\beta_{\text{sex}} = 0.5$			$\beta_{c1} = 0.5$			$\beta_{c2} = -0.5$		
		Bias	ESE	SSE	Bias	ESE	SSE	Bias	ESE	SSE	Bias	ESE	SSE	Bias	ESE	SSE
S1	RMP	0.0128	0.2468	0.2510	-0.0220	0.3329	0.3642	0.0001	0.2835	0.2969	-0.0009	0.0876	0.0899	-0.0048	0.0876	0.0942
	CL	0.0756	0.1981	0.2035	-0.0942	0.2574	0.2751	-0.0891	0.1682	0.1810	-0.0091	0.0864	0.0876	0.0031	0.0864	0.0920
	PL	-0.2873	0.1377	0.1466	0.2907	0.1696	0.1805	0.1097	0.1776	0.1880	-0.0117	0.0864	0.0878	0.0061	0.0864	0.0925
	oracle	0.0269	0.1696	0.1757	-0.0192	0.1696	0.2351	-0.0078	0.1665	0.1780	0.0041	0.0863	0.0870	-0.0047	0.0864	0.0915
S2	RMP	0.0125	0.2468	0.2507	-0.0213	0.3327	0.3637	0.0001	0.2835	0.2968	-0.0009	0.0876	0.0899	-0.0047	0.0876	0.0941
	CL	0.0754	0.1980	0.2035	-0.0936	0.2572	0.2751	-0.0891	0.1682	0.1809	-0.0091	0.0864	0.0876	0.0031	0.0864	0.0919
	PL	-0.2874	0.1376	0.1468	0.2911	0.1695	0.1806	0.1097	0.1775	0.1879	-0.0117	0.0864	0.0878	0.0061	0.0864	0.0924
	oracle	0.0267	0.1695	0.1758	-0.0186	0.1695	0.2351	-0.0079	0.1664	0.1779	0.0041	0.0863	0.0870	-0.0100	0.0863	0.0913

S1: $\beta_{\text{com,snp}} = 0$; S2: $\beta_{\text{com,snp}} = 0.5$.

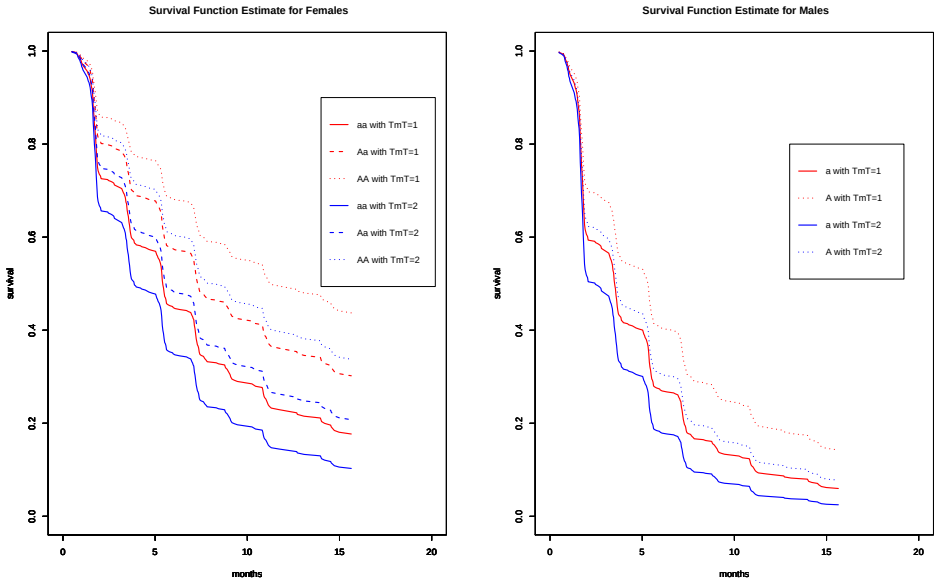
Table 9: Estimation of $\big(p_2, (u_1-p_2)/(1-p_2)\big)$ and $\big(p_{22}, (u_{21}-p_{22})/(1-p_{22})\big)$ for Scenario 3 with censoring rate being 20%

n		Methods	$\frac{u_1-p_2}{1-p_2}=0.6$	$p_2=0.2$	$\frac{u_{21}-p_{22}}{1-p_{22}}=0.3857$	$p_{22}=0.3$
n=250	S1	RMP	0.5903	0.2790	0.3618	0.3818
n=500		RMP	0.6122	0.2260	0.3761	0.3520
n=250	S2	RMP	0.5905	0.2790	0.3609	0.3822
n=500		RMP	0.6120	0.2259	0.3762	0.3521

S1: $\beta_{\text{com,snp}}=0$; S2: $\beta_{\text{com,snp}}=0.5$.

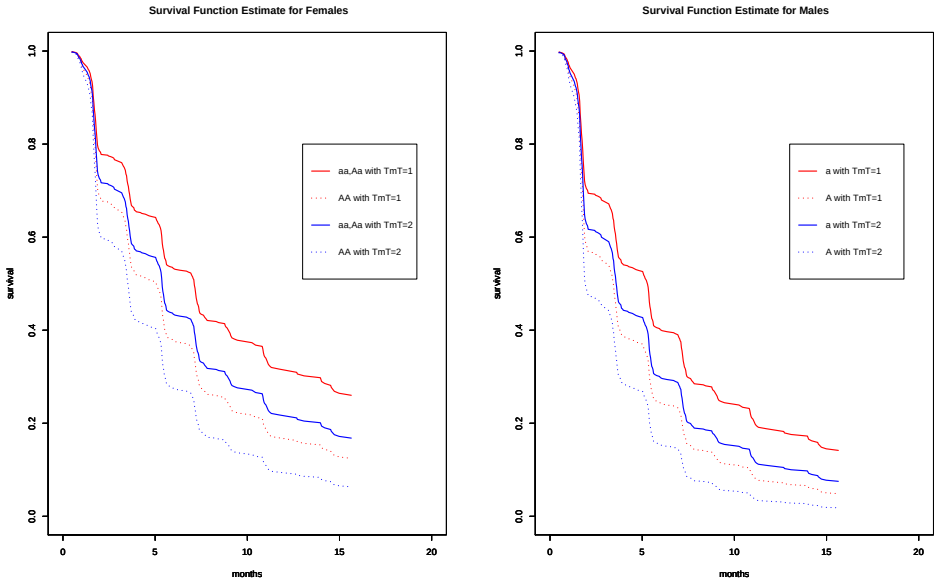
Table 10: The analysis information of top two X-linked SNPs

SNP	MAF	model	p-value	γ	u_1	u_2	β_{snp}	β_{sex}	β_{PC1}	β_{PC2}	β_{TmT}
rs1997625	0.212	RMP	0.00040	NA	1	2	-0.3695	-0.4866	-1.3486	1.0471	0.2724
		UMP(XCI-E)	0.00030	NA	NA	NA	-0.3695	-0.4865	-1.3490	1.0480	0.2722
		PL	0.00010	NA	NA	NA	-0.3695	-0.4865	-1.3490	1.0480	0.2722
		CL	0.00060	0.5	NA	NA	-0.4019	-0.3304	-1.2250	1.0280	0.2698
rs7059265	0.405	RMP	0.00090	NA	0	1	0.4362	-0.3725	-0.5831	0.5951	0.2811
		UMP(XCI-S)	0.00050	0	NA	NA	0.4362	-0.3725	-0.5831	0.5951	0.2811
		PL	0.01577	0	NA	NA	0.2252	-0.2422	-0.7132	0.5599	0.2803
		CL	0.00050	0.5	NA	NA	0.3868	-0.2964	-0.6405	0.5025	0.2844



(a) Female (b) Male

Figure 1: Survival function estimate for SNP rs1997625 with different genotypes and treatments: the blue color stands for treatment 2 while the red color stands for treatment 1. The number of patients in each genotypic group with the treatment 1 in females and males separately: 24 for "aa", 35 for "Aa", 14 for "AA"; 102 for "a", 66 for "A". The number of patients in each genotypic group with the treatment 2 in females and males separately: 33 for "aa", 46 for "Aa", 16 for "AA"; 101 for "a", 64 for "A".



(a) Female (b) Male

Figure 2: Survival function estimate for SNP rs7059265 with different genotypes and treatments: the blue color stands for treatment 2 while the red color stands for treatment 1. The number of patients in each genotypic group with the treatment 1 in females and males separately: 73 for "aa" & "Aa", 1 for "AA"; 132 for "a", 35 for "A". The number of patients in each genotypic group with the treatment 2 in females and males separately: 90 for "aa" &"Aa", 5 for "AA"; 123 for "a", 42 for "A".