Research Article

Statistics in Medicine

Received 6 December 2013, Accepted 17 December 2014 Published online 14 January 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6415

Rank-based estimating equations with general weight for accelerated failure time models: an induced smoothing approach

S. Chiou,[a*†] S. Kang[b] and J. Yan[c]

The induced smoothing technique overcomes the difficulties caused by the non-smoothness in rank-based estimating functions for accelerated failure time models, but it is only natural when the estimating function has Gehan's weight. For a general weight, the induced smoothing method does not provide smooth estimating functions that can be easily evaluated. We propose an iterative-induced smoothing procedure for general weights with the estimator from Gehan's weight initial value. The resulting estimators have the same asymptotic properties as those from the non-smooth estimating equations with the same weight. Their variances are estimated with an efficient resampling approach that avoids solving estimating equations repeatedly. The methodology is generalized to incorporate an additional weight to accommodate missing data and various sampling schemes. In a numerical study, the proposed estimators were obtained much faster without losing accuracy in comparison to those from non-smooth estimating equations, and the variance estimators provided good approximation of the variation in estimation. The methodology was applied to two real datasets, the first one from an adolescent depression study and the second one from a cancer study with missing covariates by design. The implementation is available in an R package `aftgee`.

Keywords: Gehan weight; G^ρ class; log-rank; Prentice–Wilcoxon; survival analysis; weighted log-rank test

1. Introduction

The semiparametric accelerated failure time (AFT) model and proportional hazards (PH) model [1] are two major regression models for censored failure time data. In an AFT model, failure times are directly related to covariates through a log linear formulation providing a more intuitive interpretation of the regression parameters than those in a PH model. An important estimation procedure for semiparametric AFT models is the rank-based estimating equations with a general weight. This rank-based approach inverts the weighted log-rank test [2] and its asymptotic properties have been well established [3–5]. Nonetheless, solving the weighted estimating equations is computationally challenging because the estimating functions are non-smooth step functions of regression parameters. Moreover, with a general weight, the monotonicity of the estimating functions is not guaranteed, which may lead to multiple roots. Fygenson and Ritov [5] shows that the estimating functions with Gehan's weight [6] gives monotone estimating functions that are the gradient of a convex objective function. This fact was exploited by Jin, *et al.* [7] to obtain the Gehan estimator through minimizing the convex objective function with a standard linear programming technique. The linear programming method, however, is still computationally demanding especially with larger sample sizes or more covariates.

A more recent approach that bypasses the computational difficulty is to consider approximating the non-smooth estimating functions with some suitable smooth functions. In this paper, we employ the induced smoothing procedure of [8, 9]. The idea is to smooth the non-smooth estimating functions by adding a continuous normal noise to the regression coefficients and then taking expectations with respect to the noise. Unlike other common smoothing method, such as kernel smoothing, induced smoothing is

[a] *Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN, U.S.A.*
[b] *Department of Applied Statistics, Yonsei University, Korea*
[c] *Department of Statistics, University of Connecticut, Storrs, CT, U.S.A.*
**Correspondence to: S. Chiou, Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN, U.S.A.*
†*E-mail: schiou@d.umn.edu*

practically convenient because it does not require selecting tuning parameters. For this reason, induced smoothing method is suitable for problems where unsmoothed estimating functions have jumps. In particular, notable applications of induced smoothing in different contexts includes semiparametric accelerated hazards model [10], univariate AFT model [11], clustered AFT model [12–15], and quantile regression model [16, 17]. When induced smoothing is applied to the rank-based estimating equations in AFT model, little has been studied in general weight setting. This procedure is natural for Gehan's weight, which makes the expectation with respect to the noise very easy to compute. The smoothed estimating equations are continuously differentiable with respect to the parameter and thus can be solved with standard numerical routines. The resulting estimator preserves the same asymptotic properties as those from the non-smooth estimating equations. For a non-Gehan weight, however, the current methodology does not apply and no approaches have yet been proposed.

It is often desirable to obtain rank-based estimators with general weights for efficiency concerns [3]. In general, the variance of the rank-based estimator is minimized if the limit of the weight is proportional to $\dot{\lambda}(t)/\lambda(t)$ where $\lambda(\cdot)$ is the common hazard function of the errors and $\dot{\lambda}(t) = \mathrm{d}\lambda(t)/\mathrm{d}t$ [3]. Despite this observation, the rank-based estimator with general weights has not been widely used in practice mainly because of the lack of efficient and reliable computational methods. Jin *et al.* [7] proposed an iterative method, with Gehan's estimator as initial value, to approximate a class of non-smooth estimating equations with general weight functions. Within each iteration, the estimate is obtained by linear programming with a fixed weight computed from the last iteration. This method is computationally very demanding, especially for the variance estimates obtained through bootstrap.

We generalize the induced smoothing technique to handle rank-based estimating functions with a general weight for AFT modeling. Our method is the counterpart of the iterative procedure in [7] applied to induced smoothing. For each weight fixed at values computed from the last iteration, an approximate induced smoothing procedure is proposed. The asymptotic equivalence between the non-smooth and smooth estimating functions is established, and an efficient resampling-based sandwich estimator of the asymptotic variance of the estimator is developed. Furthermore, to accommodate data from various sampling schemes and missing observations, we extend the method to include a sampling weight. For example, when the outcome of interest is rare and main covariates of interest are expensive to measure, study designs such as the case-cohort design [18] that focus on a subset of subjects in the cohort, are often considered. The sampling weight in these situations can adjust for the fact that the sample is not obtained by simple random sampling.

The rest of the paper is organized as follows. The weighted rank-based inferences for AFT models and the induced smoothing technique are presented in Section 2. An iterative-induced smoothing approach is proposed and asymptotically justified in Section 3. An efficient resampling-based sandwich estimator of the asymptotic variance is proposed in Section 4. The method is extended to incorporate an additional sampling weight in Section 5. A large-scale simulation study on performance of the estimators is reported in Section 6. The methods are then applied to two real datasets in Section 7, one from a stressor and coping study in young adults and the other from the National Wilms Tumor Study. A discussion concludes Section 8.

2. Rank-based inference and induced smoothing

For $i = 1, \ldots, n$, let T_i, C_i and X_i denote the log-transformed failure time, log-transformed censoring time and a $p \times 1$ covariate vector, respectively. It is assumed that, for subject i, T_i and C_i are independent conditional on X_i. A semiparametric AFT model is

$$T_i = X_i^{\top}\beta_0 + \epsilon_i, i = 1, \ldots, n,$$

where β_0 is a $p \times 1$ regression parameter and ϵ_is are independent and identically distributed random variables but with an unspecified distribution. The error term ϵ_i is assumed to be independent of X_i. In the presence of censoring, the observed data consist of $\{Y_i, \Delta_i, X_i\}$, $i = 1, \ldots, n$, where $Y_i = \min\left(T_i, C_i\right)$, $\Delta_i = I\left[T_i < C_i\right]$, and $I[\cdot]$ is the indicator function.

The regression parameters β can be estimated with weighted rank-based estimating function

$$U_{n,\varphi}(\beta) = \sum_{i=1}^{n} \Delta_i \varphi_i(\beta) \left[X_i - \frac{\sum_{j=1}^{n} X_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]}{\sum_{j=1}^{n} I\left[e_j(\beta) \geqslant e_i(\beta)\right]} \right], \tag{1}$$

where $e_i(\beta) = Y_i - X_i^\top \beta$ and $\varphi_i(\beta)$ is a possibly data-dependent nonnegative weight function. Let $\hat{F}_{e_i(\beta)}$ be the estimated cumulative distribution function based on the censored residual $e_i(\beta)$s. Common choices of $\varphi_i(\beta)$ are 1, $\sum_{j=1}^n I\left[e_j(\beta) \geqslant e_i(\beta)\right]$, $1-\hat{F}_{e_i(\beta)}$, and $\left(1-\hat{F}_{e_i(\beta)}\right)^\rho$, $\rho \geqslant 0$, corresponding to the log-rank weight [2], Gehan weight [6], Prentice–Wilcoxon weight [2] and general G^ρ class weight [19], respectively. For a given $\varphi_i(\beta)$, solving $U_{n,\varphi}(\beta) = 0$ for β yields a consistent estimator for the true parameter β_0, $\hat{\beta}_{n,\varphi}$, and random vector $n^{1/2}(\hat{\beta}_{n,\varphi} - \beta_0)$ is asymptotically normal for any positive monotone function $\varphi_i(\beta)$ [3,4,7]. Nevertheless, solving the estimating equation is challenging because $U_{n,\varphi}(\beta)$ is a step function. Further, $U_{n,\varphi}(\beta)$ is not necessarily a monotone function of β, so it may lead to multiple solutions. Even after one obtains an estimator numerically, estimation of the covariance matrices of the estimators can still be problematic because the slope matrix of the non-smooth estimating function is difficult to estimate.

The induced smoothing method smooths the non-smooth estimating functions in a proper way such that the resulting estimator has the same asymptotic property as that from the non-smooth version [8,9]. Let Z be a $p \times 1$ standard normal random vector and Γ_n be some $p \times p$ matrix such that $\Gamma_n^2 = \Omega_n$ with element $\Gamma_{n,ij} = O_p(1)$, where Ω_n is a symmetric positive definite matrix. With the Gehan weight, the denominator of the ratio in (1) gets canceled, and the induced smoothing procedure replaces $U_{n,G}(\beta)$ with $E_Z\left[U_{n,G}(\beta + n^{-1/2}\Gamma_n Z)\right]$, where the expectation is taken with respect to Z [9]. The resulting smooth estimating function is

$$\tilde{U}_{n,G}(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i \left(X_i - X_j\right) \Phi\left[\frac{e_j(\beta) - e_i(\beta)}{r_{ij}}\right], \tag{2}$$

where $r_{ij}^2 = n^{-1}\left(X_i - X_j\right)^\top \Omega_n \left(X_i - X_j\right)$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. A choice of Ω_n is the identity matrix [8,9]. Some other forms of Ω_n might be considered but these choices generally have minimal impact on the bias and standard error [11,12]. The estimating function in (2) is monotone and continuously differentiable with respect to β, hence its root can be found with standard numerical methods. The solution to $\tilde{U}_{n,G}(\beta) = 0$, denoted by $\tilde{\beta}_{n,G}$, is consistent to β_0 and has the same asymptotic distribution as the solution to the non-smooth version [9,12]. Given an initial choice of Ω_n, [12] obtained $\tilde{\beta}_{n,G}$ by optimizing an objective function whose gradient is (2). Alternatively, nonlinear equation solvers can be used. For example, the R package `BB` [20] implements the Barzilai–Borwein spectral methods and is used by References [11] and [21]. In general, β can be estimated with a given choice of Ω_n. Let Σ_G be the asymptotic covariance matrix of $n^{1/2}(\tilde{\beta}_{n,G} - \beta_0)$. Johnson and Strawderman [12] considered an iterative algorithm to estimate β and Σ_G alternatively by updating $\tilde{\beta}_{n,G}$ and Ω_n until convergence. This algorithm comes with a computation cost and has no obvious improvements in estimation.

3. Induced smoothing with general weights

We start with the log-rank weight $\varphi_i(\beta) = 1$, $i = 1, \ldots, n$. The corresponding estimating function is

$$U_{n,L}(\beta) = \sum_{i=1}^n \Delta_i \left[X_i - \frac{\sum_{j=1}^n X_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]}{\sum_{j=1}^n I\left[e_j(\beta) \geqslant e_i(\beta)\right]}\right].$$

Deriving the smoothed log-rank estimating equations is challenging because $E_Z\left[U_{n,L}(\beta + n^{-1/2}\Gamma_n Z)\right]$ involves the expectation of the ratio of two random quantities. We propose an approximation which replaces the expectation of the ratio with the ratio of the expectations of the two terms. Specifically, the approximated smooth estimating function is

$$\tilde{U}_{n,L}(\beta) = \sum_{i=1}^n \Delta_i \left[X_i - \frac{\sum_{j=1}^n X_j \Phi\left(\kappa_{ij}(\beta)\right)}{\sum_{j=1}^n \Phi\left(\kappa_{ij}(\beta)\right)}\right], \tag{3}$$

where $\kappa_{ij}(\beta) = \left[e_j(\beta) - e_i(\beta)\right]/r_{ij}$. The smoothed log-rank estimator, $\tilde{\beta}_{n,L}$, is the solution to $\tilde{U}_{n,L}(\beta) = 0$. In the Appendix A.1, some steps to establish the asymptotic equivalence between (3) and $E\left[U_{n,L}(\beta + n^{-1/2}\Gamma_n Z)\right]$ are provided.

For a general weight function $\varphi_i(\beta)$, obtaining the smoothed version of (1) is even more challenging because, in addition to the aforementioned ratio, $\varphi_i(\beta)$ also depends on β. We propose an iteratively reweighed induced smoothing strategy similar to that used by Jin *et al.* [7]. When $\varphi_i(b)$ is evaluated at some fixed initial estimator b of β, define

$$\tilde{U}_{n,\varphi}(b,\beta) = \sum_{i=1}^{n} \Delta_i \varphi_i(b) \left[X_i - \frac{\sum_{j=1}^{n} X_j \Phi\left(\kappa_{ij}(\beta)\right)}{\sum_{j=1}^{n} \Phi\left(\kappa_{ij}(\beta)\right)} \right]. \tag{4}$$

The proposed procedure consists of the following steps:

(1) Obtain an initial estimate $\tilde{\beta}_{n,\varphi}^{(0)} = b_n$ of β and initialize with $m = 1$.
(2) Update $\tilde{\beta}_{n,\varphi}^{(m)}$ by solving $\tilde{U}_{n,\varphi}\left(\tilde{\beta}_{n,\varphi}^{(m-1)}, \tilde{\beta}_{n,\varphi}^{(m)}\right) = 0$.
(3) Increase m by one and repeat step 2 until $|\tilde{\beta}_{n,\varphi}^{(m)} - \tilde{\beta}_{n,\varphi}^{(m-1)}| < \tau$ for a prefixed tolerance τ.

A simple choice of initial estimator is the easy-to-compute Gehan's estimator $\tilde{\beta}_{n,G}$. Using the arguments in [7], the consistency and asymptotic normality of resulting estimator $\tilde{\beta}_{n,\varphi}$ can be established. At convergence, the estimator from the smoothed estimating functions has the same asymptotic distribution as that from the unsmoothed ones. The convergence of the iterative procedure is usually very fast. In all the simulation and real data we have tested, it converged within ten steps.

A limitation of estimating function (4) is that it is not necessarily monotone, which might cause numerical problems in solving the estimating equations. Inspired by a discussion in [7], we consider reweighing the monotone estimating function (1) with $\varphi_i(\beta)$ being Gehan's weight. In particular, rewrite (1) as

$$U_{n,\phi}(\beta) = \sum_{i=1}^{n} \Delta_i \phi_i(\beta) \sum_{j=1}^{n} \left(X_i - X_j\right) I\left[e_j(\beta) \geqslant e_i(\beta)\right], \tag{5}$$

where $\phi_i(\beta) = \varphi_i(\beta)/\sum_{j=1}^{n} I\left[e_j(\beta) \geqslant e_i(\beta)\right]$. Fixing the weight $\phi_i(b)$ evaluated at b and smoothing the estimating function lead to

$$\tilde{U}_{n,\phi}(b,\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_i \phi_i(b) \left(X_i - X_j\right) \Phi\left[\kappa_{ij}(\beta)\right]. \tag{6}$$

This is the same as (2) except for the weight $\phi_i(b)$, which is free from β. Compared with (4), it is monotone in β. Hence, in each iteration, a root is guaranteed. The consistency and asymptotic normality of the resulting estimator $\tilde{\beta}_{n,\phi}$ at convergence follows from the arguments in [7]. In all the simulation and real data we have tested, convergence occurred within five steps.

4. Sandwich variance estimation

To estimate the asymptotic variance of $\tilde{\beta}_{n,\varphi}$ and $\tilde{\beta}_{n,\phi}$, one could adapt the multiplier bootstrap approach proposed by Jin *et al*. [7]. Consider $\tilde{\beta}_{n,\varphi}$ for illustration. Let η_i, $i = 1, \dots, n$, be independent and identically distributed positive random variables with $\mathrm{E}\left(\eta_i\right) = \mathrm{Var}\left(\eta_i\right) = 1$. Define a perturbed version of the estimating function (4) from induced smoothing

$$\tilde{U}^*_{n,\varphi}(b,\beta) = \sum_{i=1}^{n} \eta_i \Delta_i \varphi_i(b) \left[X_i - \frac{\sum_{j=1}^{n} \eta_j X_j \Phi\left(\kappa_{ij}(\beta)\right)}{\sum_{j=1}^{n} \eta_j \Phi\left(\kappa_{ij}(\beta)\right)} \right].$$

For a realization of $\left(\eta_1, \dots, \eta_n\right)$, and a given b, the solution to $\tilde{U}^*_{n,\varphi}(b,\beta) = 0$, $\tilde{\beta}^*_{n,\varphi}$, provides one draw of $\tilde{\beta}_{n,\varphi}$ from its asymptotic distribution. The weight $\varphi(b)$ is evaluated at $b = \tilde{\beta}_{n,\varphi}$. By repeating this process a large number B times, the variance matrix of $\tilde{\beta}_{n,\varphi}$ can be estimated by the sample variance matrix of the bootstrap sample of $\tilde{\beta}^*_{n,\varphi}$. This process is straightforward in principle, but the computation is very time consuming even with our iterative-induced smoothing procedure. It requires to solve the perturbed estimating equations B times.

We consider a more computing efficient sandwich variance estimator that extends the estimators recommended by Chiou, Kang, and Yan [11] from a performance study of several sandwich estimators. The method applies to both $\tilde{\beta}_{n,\varphi}$ and $\tilde{\beta}_{n,\phi}$, and we continue using $\tilde{\beta}_{n,\varphi}$ in the following presentation. Under some regularity conditions [22], uniformly in a neighborhood of β_0, (1) can be expressed as

$$n^{-1/2}U_{n,\varphi}(\beta) = n^{-1/2}\sum_{i=1}^{n} S_i(\beta_0) + A_n n^{1/2}(\beta-\beta_0) + o_p(1+n^{1/2}\|\beta-\beta_0\|), \tag{7}$$

where $S_i(\beta_0)$ are independent zero-mean random vectors and $A = n^{-1}\partial U_{n,\varphi}(\beta)/\partial\beta$ is the asymptotic slope matrix of $n^{-1}\tilde{U}_{n,\varphi}(\beta_0)$. The closed-form of $S_i(\beta_0)$ exists and has the following form

$$\begin{aligned} S_i(\beta_0) &= \int_{-\infty}^{\infty} \varphi_i(\beta_0)\left[X_i - \frac{\omega^{(1)}(\beta_0)}{\omega^{(0)}(\beta_0)}\right]\mathrm{d}M_i(t) \\ &= \Delta_i\varphi_i(\beta_0)\left[X_i - \frac{\omega^{(1)}(\beta_0)}{\omega^{(0)}(\beta_0)}\right] - \int_{-\infty}^{e_i(\beta)} \varphi_i(\beta_0)\left[X_i - \frac{\omega^{(1)}(\beta_0)}{\omega^{(0)}(\beta_0)}\right]\lambda(t)\mathrm{d}t, \end{aligned}$$

where

$$\omega^{(k)} = \lim_{n\to\infty}\sum_{j=1}^{n} X_j^k I\left(e_j(\beta) \geqslant e_i(\beta)\right), \text{ for } k = 0, 1,$$

$$M_i(t) = N_i(\beta;t) - \int_0^t I\left(e_i(\beta) \geqslant u\right)\lambda(u)\mathrm{d}u,$$

$N_i(\beta;t) = \Delta_i I\left(e_i(\beta) \leqslant t\right)$ and $\lambda(u)$ is the common hazard function of ϵ_i. The asymptotic variance matrix of $n^{1/2}\left(\tilde{\beta}_{n,\varphi} - \beta_0\right)$ can be estimated from $\Sigma = A^{-1}V\left(A^{-1}\right)^\top$, where $V = \lim_{n\to\infty}\mathrm{Var}\left(n^{-1/2}\sum_{i=1}^n S_i(\beta_0)\right)$ and $A = \lim_{n\to\infty} A_n$. One can estimate Σ by estimating V and A.

To estimate V, one can evaluate $S_i(\beta)$ directly by replacing the unknown quantities in its explicit expression with their sample estimators. When these sample estimators have complicated expressions, it is more convenient and perhaps more accurate to estimate V via bootstrap [22]. For that reason, we consider an efficient multiplier bootstrap approach to estimate V. First, we generate multipliers, η_i, $i = 1, \ldots, n$, with unit mean and variance independently from the data. Given a realization of $(\eta_1, \ldots, \eta_n)$, a bootstrap replicate of $\tilde{U}_n^*(b,\beta)$ evaluated at $\tilde{\beta}_{n,\varphi}$ is obtained. With large sample number B replicates, the bootstrap estimate of V is the sample variance of the B bootstrap replicates of $\tilde{U}_n^*\left(\tilde{\beta}_{n,\varphi}, \tilde{\beta}_{n,\varphi}\right)$. Unlike the aforementioned full-multiplier bootstrap approach, this bootstrap approach only requires evaluating B estimating equations rather than solving them to obtain the bootstrap replicates. Therefore, it is much less computationally demanding.

To estimate A, one might consider directly taking derivatives of the smoothed estimating functions. For the log-rank weight, the slope matrix A can be estimated by $\tilde{A}_n = n^{-1}\partial\tilde{U}_{n,L}(\beta)/\partial\beta$. This derivative can be obtained numerically or analytically. For a general weight, however, one need to compute $\tilde{A}_n = n^{-1}\partial\tilde{U}_{n,\varphi}(\beta,\beta)/\partial\beta$, which is a challenging task. For instance, nonparametric density estimation based on the censored residual is required when the estimating function is a general G^ρ weight. To avoid direct evaluation of $\tilde{A}_n$, we consider a stochastic version of the finite difference approach [22]. For a large number R, let Z_r be realizations of a p-dimensional standard normal random vector, $r = 1, \ldots, R$. It follows from (7) that

$$n^{-1/2}\tilde{U}_{n,\varphi}\left(\beta + n^{-1/2}Z_r\right) - n^{-1/2}\tilde{U}_{n,\varphi}(\beta) = A_n n^{1/2}\left(n^{-1/2}Z_r\right) + o_p(1).$$

Because $n^{-1/2}\tilde{U}_{n,\varphi}\left(\tilde{\beta}_{n,\varphi}\right) = 0$, we have

$$n^{-1/2}\tilde{U}_{n,\varphi}\left(\beta + n^{-1/2}Z_r\right) = A_n Z_r + o_p(1).$$

Let $\tilde{A}_{nj}$ be the jth row of $\tilde{A}_n$ and $\tilde{U}_{nj,\varphi}(\cdot)$ be the jth component of $\tilde{U}_{n,\varphi}(\cdot)$, $j = 1, \ldots, p$. We approximate $\tilde{A}_{nj}$ by the least squares estimate of the regression coefficients when regressing $n^{-1/2}\tilde{U}_{nj,\varphi}\left(\tilde{\beta}_{n,\varphi} + n^{-1/2}Z_r\right)$

on Z_r, $r = 1, \ldots, R$. This method can be viewed as a stochastic version of numerical derivative. It is computing efficient because, again, it only involves evaluation of the estimating functions and p least squares fits.

The justification for the aforementioned sandwich variance estimation is provided in Appendix A.2. Estimation of the asymptotic variance for $\tilde{\beta}_{n,\phi}$ can be similarly constructed but only simpler because of the perturbed weighted estimating function of Gehan's form.

5. Incorporating sampling weight

The proposed methods can be extended to accommodate various sampling schemes and missing data. One example is the case-cohort design [18], which is cost-effective when a large percentage of outcomes are censored or some of covariates are expensive to measure. In a case-cohort design, covariates are missing by design because they are measured only for the members in a case-cohort sample. The case-cohort sample consists of a subcohort, which is a random subset of the full cohort, and all the remaining non-censored members, known as cases, outside the subcohort. This design is a special case of the more general stratified case-cohort design where a subcohort is selected via a stratified random sampling from S mutually exclusive strata in the original full cohort. Let $\tilde{n}_s$ denote the numbers of subjects sampled from the sth stratum consisting of n_s subjects for $s = 1, \ldots, S$. Statistical methods do not account for these sampling schemes could result in biased estimates because the stratified case-cohort sample is not a simple random sample.

A typical method to adjust for biases is to weigh each observation by the inverse of its inclusion probability. Let ψ_{is} be the strata indicator ($\psi_{is} = 1$ if the ith subject is in the sth stratum and $\psi_{is} = 0$ otherwise) and ξ_i be the sampling indicator ($\xi_i = 1$ if the ith subject is sampled and $\xi_i = 0$ otherwise). The inclusion probability of subject i is $h_i = \sum_{s=1}^{S} \xi_i \psi_{is}/p_{n,s}$, where $p_{n,s} = \tilde{n}_s/n_s$. With the sampling weight, the non-smooth estimating functions (1) and (5) become,

$$U^c_{n,\varphi}(\beta) = \sum_{i=1}^{n} \Delta_i h_i \varphi_i(\beta) \left[X_i - \frac{\sum_{j=1}^{n} h_j X_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]}{\sum_{j=1}^{n} h_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]} \right],$$

and

$$U^c_{n,\phi}(\beta) = \sum_{i=1}^{n} \Delta_i h_i \phi^c_i(\beta) \sum_{j=1}^{n} h_j (X_i - X_j) I\left[e_j(\beta) \geqslant e_i(\beta)\right], \text{respectively}$$

where $\phi^c_i(\beta) = \varphi_i(\beta)/\sum_{j=1}^{n} h_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]$.

Applying the iterative procedure and induced smoothing leads to

$$\tilde{U}^c_{n,\varphi}(b, \beta) = \sum_{i=1}^{n} h_i \Delta_i \varphi_i(b) \left[X_i - \frac{\sum_{j=1}^{n} h_j X_j \Phi\left(\kappa_{ij}(\beta)\right)}{\sum_{j=1}^{n} h_j \Phi\left(\kappa_{ij}(\beta)\right)} \right], \tag{8}$$

and

$$\tilde{U}^c_{n,\phi}(b, \beta) = \sum_{i=1}^{n} \sum_{i=1}^{n} h_i h_j \Delta_i \phi^c_i(b) \left(X_i - X_j\right) \Phi\left[\kappa_{ij}(\beta)\right]. \tag{9}$$

where $\phi^c_i(b) = \varphi_i(b)/\sum_{j=1}^{n} h_j I\left(\kappa_{ij}(\beta)\right)$. For full- cohort data, $\xi_i = p_{n,s} = 1$ for $i = 1, \cdots, n$ and $s = 1, \cdots, S$, (8) and (9) reduce to (4) and (6), respectively. This method extends the induced smoothing approach for case-cohort data with Gehan's weight [11]. Let $\tilde{\beta}^c_{n,\varphi}$ and $\tilde{\beta}^c_{n,\phi}$ be resulting estimator at the convergence of the iterative procedure. The consistency and asymptotic normality of $\tilde{\beta}^c_{n,\varphi}$ and $\tilde{\beta}^c_{n,\phi}$ are the same as those from the corresponding non-smooth estimating equations. The asymptotic variance of $\tilde{\beta}^c_{n,\varphi}$ and $\tilde{\beta}^c_{n,\phi}$ can be derived similar to the case with Gehan's weight [11]. The computing efficient sandwich estimator of the asymptotic variance in Section 4 can be adapted accordingly.

6. Simulation

An extensive simulation study was conducted to assess the finite-sample properties of the proposed methods. Failure time T was generated from the AFT model

$$\log(T) = 2 + X_1 + X_2 + X_3 + \epsilon,$$

where X_1 was Bernoulli with rate 0.5, X_2 and X_3 were uncorrelated standard normal variables. The error term ϵ followed standard normal, standard Gumbel, or standard logistic distributions, abbreviated by N, G, and L, respectively. The censoring times were generated from uniform distributions $[0, c]$ where c was adjusted to yield the desirable censoring rates, C_p. Two censoring rates, $C_p \in \{0.25, 0.50\}$, were considered for the full-cohort data. Three sample sizes were considered, $n \in \{100, 200, 400\}$. We considered the following three types of weight functions, log-rank (LR), Prentice–Wilcoxon (PW), and the general G^ρ class (G^ρ), in addition to the Gehan weight. When the general G^ρ class was used, ρ was chosen to be $1/3$ (one over the number of the covariates). For each weight function, three point estimators were considered: the non-smoothed (NS) version that solves (1) directly, the induced smoothing (IS) version from (4), and the monotone induced smoothing (MIS) version from (6). Smooth estimator from Gehan's weight, $\tilde{\beta}_{n,G}$ was used as the initial value for both the IS and MIS methods. An absolute convergence criteria is considered with a prefixed tolerance of $\tau = 10^{-4}$. Using the perturbation variables, η, generated from independent standard exponential distribution, the variances estimator for the IS and MIS methods were estimated with the proposed sandwich estimators with bootstrap sample size $B = 100$ in obtaining V_n and $R = 100$ in obtaining A_n. To the variance estimator under NS, the proposed sandwich variance estimation was applied to the perturbed estimating function of NS

Table I. Comparison of the MIS, IS, and NS estimators under Gumbel error margin and $n = 100$. MIS is the monotone induced smoothing estimator; IS is the induced smoothing estimator; NS is the non-smoothed estimator; PE is the point estimator; ESE is the empirical standard errors; ASE is the average of the standard errors of the estimator; CP is the coverage percentage.

		PE			ESE			ASE			CP(%)		
C_p	β	MIS	IS	NS	MIS	IS	NS	MIS	IS	NS	MIS	IS	NS
						Log–rank							
0.25	β_1	1.023	1.017	0.981	0.299	0.313	0.292	0.315	0.285	0.285	95.3	91.9	94.2
	β_2	1.007	1.004	1.004	0.152	0.160	0.157	0.164	0.146	0.144	95.2	92.3	92.6
	β_3	1.015	1.013	1.002	0.148	0.154	0.157	0.162	0.144	0.144	96.5	93.4	92.4
0.50	β_1	1.023	1.013	1.007	0.305	0.318	0.324	0.335	0.306	0.294	97.0	93.0	92.2
	β_2	1.020	1.023	1.003	0.172	0.179	0.168	0.182	0.162	0.157	94.8	91.6	92.6
	β_3	1.031	1.030	1.002	0.176	0.182	0.173	0.181	0.163	0.156	93.3	90.3	90.4
						Prentice–Wilcoxon							
0.25	β_1	1.007	1.006	1.012	0.251	0.249	0.250	0.248	0.242	0.237	93.7	93.2	92.2
	β_2	1.001	1.001	0.996	0.127	0.126	0.132	0.128	0.125	0.123	93.4	93.7	91.1
	β_3	1.005	1.006	1.004	0.125	0.124	0.124	0.127	0.124	0.121	94.5	93.8	93.5
0.50	β_1	1.009	1.009	1.000	0.268	0.269	0.269	0.273	0.265	0.256	94.7	94.8	92.9
	β_2	1.009	1.009	1.000	0.147	0.147	0.144	0.146	0.143	0.137	92.9	92.2	93.3
	β_3	1.014	1.013	1.002	0.154	0.153	0.148	0.146	0.143	0.138	92.8	91.9	92.4
						General G^ρ with $\rho = 1/3$							
0.25	β_1	1.020	1.006	1.015	0.298	0.249	0.277	0.320	0.241	0.261	95.2	94.4	92.3
	β_2	1.007	1.001	0.996	0.152	0.136	0.148	0.164	0.135	0.133	95.4	93.0	91.1
	β_3	1.015	1.006	1.005	0.148	0.134	0.138	0.162	0.135	0.133	96.8	94.1	93.4
0.50	β_1	1.016	1.008	0.996	0.293	0.301	0.296	0.308	0.287	0.278	95.5	93.9	92.9
	β_2	1.019	1.017	1.001	0.164	0.157	0.158	0.162	0.152	0.148	95.0	92.6	93.5
	β_3	1.021	1.019	1.002	0.160	0.154	0.162	0.162	0.154	0.148	95.0	93.7	92.3

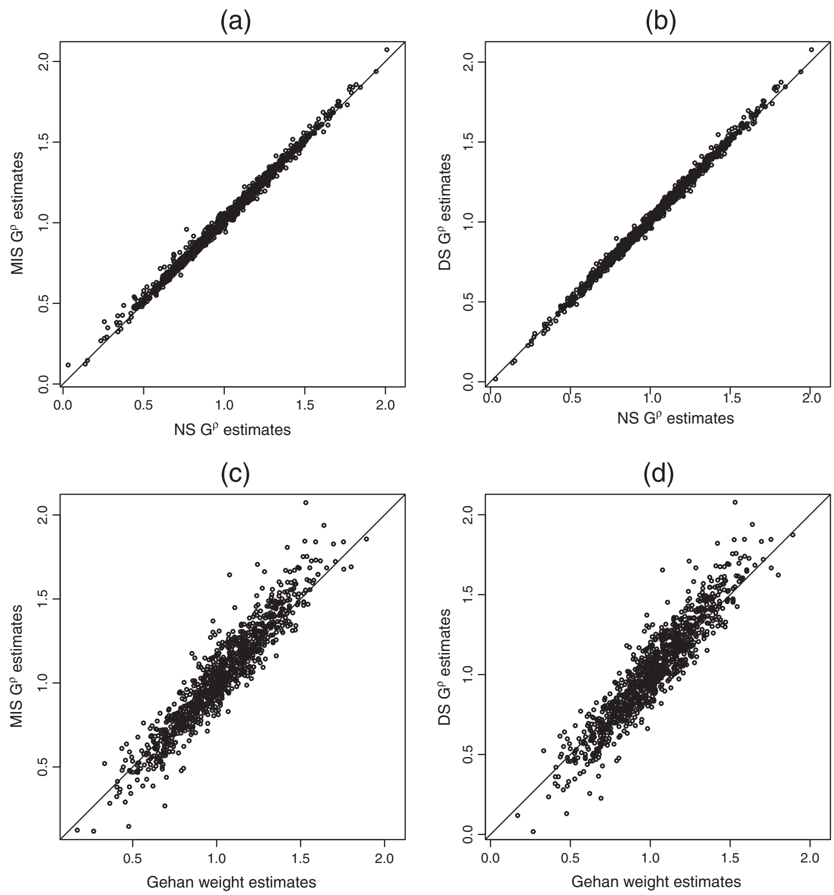

Figure 1. Comparisons of different estimators for β_1 under 50% censoring rate, $n = 100$ and Gumbel error distribution. (a) and (b): Non-smooth G^ρ estimates (NS) versus smoothed G^ρ estimates (MIS); (c) and (d): Gehan estimates versus smoothed G^ρ estimates (MIS).

$$U^*_{n,\varphi}(\beta) = \sum_{i=1}^{n} \eta_i \Delta_i \varphi_i(\beta) \left[X_i - \frac{\sum_{j=1}^{n} \eta_j X_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]}{\sum_{j=1}^{n} \eta_j I\left[e_j(\beta) \geqslant e_i(\beta)\right]} \right].$$

For each scenario, we obtained 1000 replicates.

We first investigated the difference between the MIS, IS, and NS estimators. To save space, only the results with $n = 100$ and Gumbel error distribution are summarized in Table I; results under other scenarios are similar. All estimators appear to be virtually unbiased. Our iterative estimators, IS and MIS, agreed with the NS closely on a 45° line as shown in Figure 1(a and b) for β_1. For the NS and IS estimators, the standard errors appear to slightly underestimate the empirical variation, which is typical for sandwich estimators with smaller sample sizes. The underestimation is not observed for the MIS estimator. As a result, the 95% confidence intervals associated with the NS and IS estimators have slightly lower empirical coverage percentages whereas that of the MIS estimator has proper coverage. The coverage improves as the sample size increases.

Now that all the estimators are close to each other, we next focus on the computing efficiency. Table II summarizes the timing results of the full-scale simulation using our implementation in R package `aftgee` [21]. Each entry represents the time in seconds, averaged from 1000 replicates, to obtain the point estimate under the corresponding scenario on a Linux machine with a 2GHz AMD Opteron Processor (Advanced Micro Devices, Sunnyvale, CA, USA). For log-rank weight, the IS estimator was most efficient, up to 1.8 times faster than the NS estimator (with $n = 100$, normal margin, $C_p = 50\%$). Because no iterative process is needed, both the IS and MIS estimators were obtained fairly quickly, with a speed comparable to that of the NS estimator. For the PW weight and G^ρ weight, however, the advantage of the IS and MIS estimators is much more obvious. The IS estimator was up to 46 times faster

Table II. Timing results in seconds for point estimation. MIS is the monotone induced smoothing estimator; IS is the induced smoothing estimator; NS is the non-smoothed estimator; ED is the error distribution; N is the standard normal distribution; G is the standard Gumbel distribution; L is the standard logistic distribution.

		$n = 100$			$n = 200$			$n = 400$		
ED	C_p	MIS	IS	NS	MIS	IS	NS	MIS	IS	NS
					Log-rank					
N	25%	2.4	0.8	1.3	4.5	2.2	2.9	12.2	7.6	9.2
	50%	2.4	0.6	1.1	4.0	1.6	2.1	9.6	5.2	6.0
G	25%	2.6	0.8	1.2	4.8	2.2	2.7	12.8	7.3	8.8
	50%	2.4	0.6	1.1	4.0	1.6	2.0	9.5	5.2	5.9
L	25%	2.8	0.8	1.3	4.9	2.0	2.5	12.2	6.6	8.6
	50%	2.6	0.6	1.1	4.0	1.5	1.8	9.5	4.5	6.1
					Prentice–Wilcoxon					
N	25%	1.5	18.8	103.4	2.9	34.2	178.4	8.0	40.3	403.0
	50%	1.6	12.0	101.6	2.7	14.8	174.2	6.9	21.5	389.6
G	25%	1.5	18.5	103.5	2.8	33.1	175.4	8.0	37.1	407.8
	50%	1.6	11.9	101.5	2.6	13.1	169.5	6.7	20.2	390.9
L	25%	1.6	18.9	99.5	2.8	41.8	170.8	7.5	47.1	405.4
	50%	1.7	10.8	100.0	2.7	13.4	166.6	6.7	19.9	385.0
					General G^ρ weight					
N	25%	2.0	3.1	104.9	3.8	7.1	180.1	10.3	22.5	390.2
	50%	2.0	2.7	103.3	3.4	5.3	173.6	8.4	15.8	384.4
G	25%	2.1	2.8	103.7	3.9	7.1	174.0	10.6	22.7	394.6
	50%	2.0	2.3	103.5	3.4	5.2	169.6	8.2	13.5	379.5
L	25%	2.2	3.0	103.0	4.0	6.8	171.8	10.2	21.2	395.7
	50%	2.2	2.2	101.5	3.4	4.9	166.4	8.3	14.5	373.8

than the NS estimator (with $n = 100$, logistic margin, $C_p = 50\%$, G^ρ weight), and the MIS estimator was up to 69 times faster than the NS estimator (with $n = 100$, normal margin, $C_p = 25\%$, PW weight). Between the two induced smoothing estimators, the MIS estimator was up to 15 times faster than the IS estimator (with $n = 200$, logistic margin, $C_p = 25\%$, PW weight). Note that these timing results are for point estimation only. Given that the three estimators have similar accuracy, the MIS estimator clearly has an edge in practice. Therefore, only the results of the MIS estimator are reported in the sequel for each weight.

Table III summarizes the simulation results for all scenarios with sample size $n \in \{200, 400\}$. All estimators appear unbiased under all scenarios. The average of the standard errors and the empirical standard errors are very close, suggesting that the proposed sandwich estimator has good performance for the reported sample sizes. Consequently, the empirical coverage percentages of the 95% confidence intervals are close to the nominal level. Among the three weights considered, the PW estimator seems to have the smallest standard errors and the closest coverage percentage for the 95% confidence intervals under all the scenarios considered in this simulation study. The differences in standard errors among the three weights are more visible with Gumbel errors, with PW weight the smallest, followed by the G^ρ weight, and then the LR weight. Figure 1(c and d) show the comparison of the G^ρ estimator versus initial value, the Gehan estimator, for β_1. The estimator with Gehan's weight seems to have smaller variation.

Lastly, we modify the simulation setting to mimic a case-cohort study and investigate the performance of the method with sampling weights. The full-cohort size was set to be 1500. The censoring rate was set to be $C_p = 0.9$ mimicking a heavy censoring rate for a rare disease. The subjects were stratified into two strata: stratum 1 formed by cases ($\Delta_i = 1$) and stratum 2 formed by controls ($\Delta_i = 0$). A stratified sample was obtained by sampling all the subjects from the cases and sampling a number of subjects from the controls so that the average case-cohort size was 300. This sampling scheme means that, on average, the weights for subjects in stratum 1 and 2 are 1 and 9, respectively. The results are summarized in Table IV. All point estimators appear to be virtually unbiased. The empirical standard errors and average of the sandwich standard errors closely agree with each other. The coverage percentages of the 95% confidence intervals based on the sandwich variance were reasonably close to the nominal level.

Table III. Summary of simulation studies on assessing the performance of the monotone induced smoothing estimator. PE is the point estimator; ESE is the empirical standard errors; ASE is the average of the standard errors of the estimates; CP is the coverage percentage; ED is the error distribution; N is the standard normal distribution; G is the standard Gumbel distribution; L is the standard logistic distribution; LR is the log-rank weight; PW is the Prentice–Wilcoxon weight; G^ρ is the general G^ρ weight.

				PE			ESE			ASE			CP (%)		
ED	n	C_p	β	LR	PW	G^ρ	LR	PW	G^ρ	LR	PW	G^ρ	LR	PW	G^ρ
N	200	0.25	β_1	1.014	1.003	1.008	0.165	0.159	0.160	0.175	0.159	0.164	96.2	94.3	95.6
			β_2	1.013	1.007	1.009	0.088	0.085	0.086	0.091	0.083	0.085	94.5	93.5	94.6
			β_3	1.007	1.000	1.003	0.080	0.078	0.078	0.090	0.083	0.085	96.7	95.9	95.7
		0.50	β_1	1.020	1.003	1.012	0.189	0.180	0.183	0.202	0.183	0.190	96.1	94.6	95.4
			β_2	1.016	1.007	1.011	0.103	0.099	0.101	0.109	0.099	0.103	94.6	94.1	94.9
			β_3	1.018	1.007	1.012	0.106	0.101	0.103	0.110	0.100	0.103	93.8	94.2	94.4
	400	0.25	β_1	1.005	0.998	1.001	0.119	0.114	0.116	0.121	0.111	0.114	94.7	93.4	94.4
			β_2	1.006	1.002	1.004	0.061	0.059	0.059	0.063	0.058	0.060	95.3	94.0	95.3
			β_3	1.005	1.002	1.004	0.057	0.055	0.056	0.063	0.058	0.059	96.3	96.0	96.4
		0.50	β_1	1.010	1.002	1.006	0.132	0.130	0.130	0.140	0.128	0.133	95.7	94.1	94.9
			β_2	1.008	1.003	1.006	0.073	0.069	0.071	0.076	0.070	0.072	94.8	94.5	94.7
			β_3	1.010	1.006	1.008	0.070	0.069	0.069	0.076	0.070	0.072	96.1	95.0	96.2
G	200	0.25	β_1	1.010	0.997	1.003	0.209	0.173	0.191	0.223	0.173	0.194	95.9	93.7	94.6
			β_2	1.006	1.003	1.004	0.105	0.090	0.097	0.112	0.089	0.100	95.4	94.6	95.2
			β_3	1.005	0.999	1.002	0.102	0.086	0.094	0.113	0.089	0.100	96.1	94.1	95.8
		0.50	β_1	1.013	1.002	1.008	0.213	0.184	0.198	0.233	0.189	0.209	96.8	95.8	95.7
			β_2	1.013	1.006	1.010	0.115	0.100	0.107	0.124	0.101	0.112	95.6	94.7	95.2
			β_3	1.014	1.005	1.009	0.113	0.098	0.106	0.124	0.101	0.112	95.6	95.7	96.5
	400	0.25	β_1	1.001	0.994	0.997	0.142	0.118	0.130	0.153	0.120	0.135	96.0	94.5	95.1
			β_2	1.005	1.001	1.003	0.071	0.060	0.066	0.077	0.062	0.069	95.6	95.2	95.4
			β_3	1.008	1.003	1.005	0.072	0.061	0.067	0.077	0.062	0.069	96.1	94.2	95.0
		0.50	β_1	1.007	1.000	1.003	0.144	0.126	0.136	0.161	0.132	0.146	96.0	94.9	95.8
			β_2	1.007	1.003	1.005	0.077	0.069	0.073	0.085	0.070	0.077	96.7	95.4	96.4
			β_3	1.007	1.003	1.005	0.081	0.070	0.076	0.086	0.071	0.078	95.4	94.9	94.9
L	200	0.25	β_1	1.020	1.002	1.012	0.260	0.245	0.249	0.293	0.262	0.271	96.6	96.4	96.8
			β_2	1.016	1.010	1.012	0.147	0.138	0.141	0.149	0.134	0.139	94.6	93.6	93.4
			β_3	1.004	0.995	0.999	0.144	0.139	0.140	0.151	0.136	0.140	95.5	93.6	93.2
		0.50	β_1	1.013	1.003	1.008	0.291	0.281	0.284	0.322	0.297	0.306	96.8	96.5	96.4
			β_2	1.017	1.008	1.012	0.161	0.155	0.157	0.171	0.157	0.163	95.9	95.9	95.8
			β_3	1.020	1.012	1.016	0.170	0.162	0.165	0.171	0.157	0.162	94.8	93.7	94.1
	400	0.25	β_1	1.021	1.014	1.017	0.190	0.179	0.183	0.203	0.183	0.188	95.6	94.5	94.8
			β_2	1.010	1.005	1.008	0.098	0.092	0.094	0.103	0.094	0.097	95.4	95.0	95.4
			β_3	1.004	1.000	1.002	0.100	0.094	0.096	0.103	0.094	0.097	95.2	95.0	94.8
		0.50	β_1	1.019	1.014	1.016	0.212	0.204	0.206	0.223	0.207	0.212	95.0	95.1	94.5
			β_2	1.005	1.001	1.003	0.109	0.106	0.107	0.118	0.110	0.113	96.4	96.3	96.4
			β_3	1.013	1.007	1.010	0.113	0.110	0.111	0.117	0.110	0.112	94.8	95.3	94.9

To investigate the effect of bootstrapping size used in variance estimation, we also obtained the results for $B = 200$ and $R = 200$. These results did not provide much improvement over $B = 100$ and $R = 100$, thus, are not reported. Considering the computational cost comes with larger bootstrap size, we believe choosing $B = 100$ and $R = 100$ is sufficient, at least for the scenarios considered.

7. Applications

7.1. Stressor and coping in young adults

As part of a longitudinal study of the stress process during the adolescent and early adult period, data on event duration and coping were collected in one of the waves with 829 participants in the greater Boston metropolitan area in 1996 [23]. The sample contained more women (60%) than men and was predominantly Caucasian (94%), which were consistent with the demographics of the original sample. All respondents were asked to describe in detail a single major stressor (an event, situation, or problem that had been the most difficult or stressful) during the past year. Among the 829 participants, 22 were dropped because of incompleteness. The duration of the stressor was censored for 451 (56%) out of the effective sample size of 807 because their stressors were still ongoing at the time of the interview. As Harnish, Aseltine, and Gore [23] suggested, durations over 27 months were overly long. These durations were truncated to 27 months and considered as censored events.

The severity of the stressor was coded on a four-point scale (1 to 4) based on descriptive, contextual information [24]. The coping strategies employed during the course of the stressor were constructed from their responses to each of the 24 coping behaviors on a four-point scale [23]. The final measure consisted of six types of coping strategies: avoidance (avoid), positive reappraisal (reapp), religion (relig), active cognitive (actcog), active behavioral (actbeh), and social support (socsup). The score of each coping strategy was the sum of four items, with possible scores on each strategy ranging from 4 to 16.

We fitted AFT models to the stressor duration with nine covariates: gender (male = 0, female = 1), race (Caucasian = 0, other = 1), severity, and six coping strategy scores. The weight functions considered were GE, LR, PW, and the general G^ρ weight (G^ρ) with $\rho = 1/9$. Point estimate is estimated by the MIS estimator initiated at $\tilde{\beta}_{n,G}$ and standard error is estimated by the proposed sandwich variance estimator with random perturbation generated from $B = R = 100$ independent standard exponential distribution. Absolute convergence criteria with a prefixed tolerance of $\tau = 10^{-4}$ was used for the iteration. The results are summarized in Table V. All the point estimates are similar to one another. The standard errors from the GE and PW weights appear to be smaller than those with the LR and G^ρ weights, especially for those covariates that had more significant effects. In fact, the gender effect was only significant with the GE

Table IV. Summary of simulation studies for data from a stratified case-cohort study with full-cohort size 1500, case-cohort size 300, and censoring rate $C_p = 0.9$. PE is the point estimator; ESE is the empirical standard errors; ASE is the average of the standard errors of the estimates; CP is the coverage percentage; ED is the error distribution; N is the standard normal distribution; G is the standard Gumbel distribution; L is the standard logistic distribution; LR is the log-rank weight; PW is the Prentice–Wilcoxon weight; G^ρ is the general G^ρ weight.

		PE			ESE			ASE			CP (%)		
ED	β	LR	PW	G^ρ	LR	PW	G^ρ	LR	PW	G^ρ	LR	PW	G^ρ
N	β_1	1.015	1.013	1.014	0.199	0.182	0.192	0.220	0.194	0.207	94.9	95.8	95.4
	β_2	1.018	1.014	1.016	0.107	0.098	0.103	0.118	0.103	0.110	95.8	95.9	94.7
	β_3	1.021	1.016	1.019	0.109	0.103	0.106	0.117	0.104	0.110	95.6	94.8	95.1
G	β_1	1.016	1.008	1.012	0.205	0.174	0.191	0.229	0.186	0.210	95.7	96.1	97.4
	β_2	1.017	1.011	1.014	0.109	0.096	0.104	0.123	0.102	0.113	96.6	95.9	96.1
	β_3	1.022	1.016	1.019	0.112	0.100	0.107	0.122	0.102	0.113	94.4	94.8	95.3
L	β_1	1.028	1.024	1.026	0.307	0.296	0.302	0.326	0.291	0.305	93.7	93.8	92.9
	β_2	1.016	1.011	1.014	0.154	0.148	0.152	0.167	0.153	0.160	95.3	95.2	94.9
	β_3	1.020	1.016	1.018	0.156	0.151	0.154	0.165	0.155	0.159	94.0	94.6	95.6

Table V. Point estimates (PE) and standard errors (SE) of accelerated failure time models fitted to the stressor and coping data with four weights: Gehan (GE), log-rank (LR), Prentice–Wilcoxon (PW), and the general G^ρ weight (G^ρ) with $\rho = 1/9$.

	GE		LR		PW		G^ρ	
Covariate	PE	SE	PE	SE	PE	SE	PE	SE
Gender	0.396	0.180	0.397	0.220	0.397	0.197	0.385	0.231
Race	0.059	0.466	−0.082	0.554	−0.038	0.441	−0.060	0.464
Severity	0.216	0.091	0.207	0.099	0.210	0.098	0.200	0.097
Avoid	0.109	0.035	0.115	0.044	0.107	0.040	0.103	0.042
Reapp	0.008	0.035	−0.005	0.047	0.001	0.042	−0.003	0.046
Relig	0.022	0.023	0.015	0.031	0.018	0.027	0.015	0.029
Actcog	0.015	0.050	0.034	0.060	0.027	0.054	0.031	0.054
Actbeh	−0.027	0.042	−0.074	0.062	−0.057	0.053	−0.067	0.054
Socsup	−0.024	0.032	−0.044	0.056	−0.034	0.040	−0.036	0.049

Note: Avoid, avoidance; Reapp, positive reappraisal; Relig, religion; Actcog, active cognitive; Actbeh, active behavioral; Socsup, social support.

and PW weights; women tended to have longer stressor duration than men. The stressor severity was also significant; more severe stressors tend to have longer durations. After controlling for gender, race, and severity, the only coping strategy that was found to have a significant effect was avoidance. Coping with avoidance tended to prolong the duration of the stressor. These results are consistent with those obtained with the PH model [23] but might be more intuitive to understand because the effects of the covariates are on the duration itself rather than its hazards.

7.2. *National Wilms Tumor Study*

We illustrate the incorporation of sampling weights in the proposed methods with the data from a case-cohort study. The National Wilms Tumor Study (NWTS) was first created in 1969 to address the need to study and compare treatments in different periods of time [25, 26]. The subjects considered here were from two study groups, NWTSG-3 and NWTSG-4, recorded between May 1979 and August 1995. The specific research interest was to assess the relationship between the tumor histology and time to tumor relapse. Tumor histology was a dichotomous variable classified depending on whether a histological diagnosis was favored. The diagnosis performed at a central lab was accurate but expensive. There were a total of 4028 subjects in the full cohort. Among them, 571 were the cases who experienced relapse, and 3457 were the controls who did not experience relapse. The censoring rate was about 86% reflecting the fact that Wilms tumor is a rare kidney cancer in young children. Although a central histological diagnosis was eventually available for each subject, the data had a case-cohort version where only those included in the case-cohort sample had central lab diagnosis. The case-cohort version of the data had 1154 subjects, including 571 cases and 583 controls, which gave sampling weights 1 and 5.93, respectively. It has been analyzed with Cox models [27, 28] and additive hazards models [29], respectively. Other covariates included were subject age, study group, and tumor stage, which indicated the tumor spread. The staging system consisted of four stages, with Stage IV being the latest and most severe stage. The data are available in the R package `survival` [30].

We fitted AFT models to both the case-cohort data with four weights: GE, LR, PW, and G^ρ with $\rho = 1/6$. The covariates included central histology measurement (1 = favorable, 0 = unfavorable), age (measured in years) at diagnosis, three tumor stages indicators (Stage I as reference) and a study group indicator (NWTSG-3 as reference). Point estimate is estimated by the MIS estimator initiated at $\tilde{\beta}_{n,G}$ and standard error is estimated by the proposed sandwich variance estimator with random perturbation generated from $B = R = 100$ independent standard exponential distribution. Absolute convergence

Table VI. Point estimates (PE) and standard errors (SE) of accelerated failure time models fitted to the data from the National Wilms Tumor Study with four weights: Gehan (GE), log-rank (LR), Prentice–Wilcoxon (PW), and the general G^ρ weight (G^ρ) with $\rho = 1/6$. Timing results are presented in seconds.

	GE		LR		PW		G^ρ	
Covariate	PE	SE	PE	SE	PE	SE	PE	SE
Case-cohort analysis:								
(Time)	(66.8)		(58.1)		(163.8)		(167.3)	
Histology	−2.743	0.213	−3.291	0.294	−3.223	0.129	−3.280	0.189
Age (years)	−0.127	0.038	−0.153	0.055	−0.149	0.055	−0.152	0.051
Stage2	−1.334	0.264	−1.501	0.398	−1.460	0.315	−1.494	0.357
Stage3	−1.340	0.312	−1.418	0.370	−1.389	0.322	−1.413	0.338
Stage4	−2.201	0.324	−2.611	0.382	−2.526	0.490	−2.596	0.456
Study	−0.145	0.227	−0.228	0.286	−0.219	0.243	−0.226	0.292
Full-cohort analysis:								
(Time)	(565.9)		(210.4)		(627.1)		(552.6)	
Histology	−2.749	0.202	−3.758	0.162	−3.614	0.143	−3.731	0.160
Age (years)	−0.127	0.037	−0.177	0.039	−0.172	0.029	−0.176	0.038
Stage2	−1.335	0.280	−1.466	0.233	−1.414	0.200	−1.458	0.238
Stage3	−1.341	0.286	−1.808	0.251	−1.694	0.195	−1.789	0.253
Stage4	−2.203	0.319	−2.627	0.294	−2.404	0.239	−2.584	0.281
Study	−0.106	0.226	−0.361	0.197	−0.304	0.191	−0.350	0.214

criteria with a prefixed tolerance of $\tau = 10^{-4}$ was used for the iteration. The results are summarized in Table VI. All estimates led to the same conclusion, the coefficients of central histological diagnosis, age, and all stages had a significant effect on the time to relapse. Favorable central histological diagnosis, older subjects, and higher tumor stage were associated with a shorter time to relapse.

The point estimates from the full-cohort data were close to those from the case-cohort data, considering their standard errors. All the standard errors decreased in the full-cohort data analysis. The GE weight gave the smallest standard errors in the case-cohort analysis, but the PW weight gave the smallest standard errors in the full-cohort analysis. For the sample size in this example, the fitting with all four weights completed within 8 minutes for the case-cohort data and 33 minutes for the full-cohort data. The time results suggest that the proposed method can be realistically applied to routine survival analysis with studies of thousands of subjects.

8. Discussion

Rank-based inferences for AFT modeling are seldom applied with general weight functions mainly because of the lack of efficient and reliable computing algorithm. With the recently developed induced smoothing technique, we have proposed two iterative procedures for rank-based inferences with general weight function and further extended it to allow sampling weights. The smoothed estimating functions were shown to have the same asymptotic properties as those non-smooth ones and our computationally efficient variance estimators provided good approximations to the true variation. All these methods are publicly available in an R package `aftgee` [21, 31]. An alternative method for the variance computation is to consider a Monte Carlo method proposed by Jin, Shao, and Ying [32], where the explicit form of the slope matrix is implicated from simple integration by parts argument, and the covariance matrix is estimated via an iterative algorithm. This approach also does not involve solving estimating equations but it requires evaluating the estimating functions a large number of times at each iteration. Thus, it is expected to take longer than our method.

The proposed tools make it possible to compare the performance of different weight functions with different values of ρ in G^ρ. It also facilitates combining estimators obtained with different weights in some optimal or suboptimal way. For rare events, it might be worth to consider the weight function proposed by Buyske, Fagerstrom, and Ying [33]. When the censoring rate is high, the G^ρ class of weight does not have much flexibility because $\left(1 - \hat{F}_{e_i(\beta)}\right)^\rho$ is close to 1 for all $e_i(\beta)$ and $0 \leqslant \rho \leqslant 1$, making it very similar to the log-rank weight. The modified G^ρ weight in [33], $\varphi_i(\beta) = \left[\max_i \left(\hat{F}_{e_i(\beta)}\right) - \hat{F}_{e_i(\beta)}\right]^\rho$, can be used with the iterative-induced smoothing procedure. Another interesting aspect in the application of the proposed methods is to consider an optimal choice for the weight function. For example, to obtain an asymptotically efficient estimate, one might choose $\varphi_i(\beta) = \dot{\lambda}\left[e_i(\beta)\right] / \lambda\left[e_i(\beta)\right]$ and replace $\dot{\lambda}(\cdot)$ and $\lambda(\cdot)$ with their corresponding nonparametric counterparts. Provided that the iterative-induced smoothing procedure converges, its performance can be studied.

Appendix A

A.1. Asymptotic equivalence

Here we provide a sketched proof to establish the asymptotic equivalence of $\tilde{U}_{n,L}(\beta)$ and $E_Z\left[U_{n,L}(\beta + n^{-1/2}\Gamma_n Z)\right]$. For $k \in \{0, 1\}$, define $W_{n,i}^{(k)}(\beta) = n^{-1}\sum_{j=1}^n X_j^k I\left[e_j(\beta) \geqslant e_i(\beta)\right]$, $\omega_{n,i}^{(k)}(\beta, Z) = W_{n,i}^{(k)}\left(\beta + n^{-1/2}\Gamma_n Z\right)$, and $\tilde{W}_{n,i}^{(k)}(\beta) = E_Z\left[\omega_{n,i}^{(k)}(\beta, Z)\right]$. We assume the following conditions:

A1. $\Pr\left[W_{n,i}^{(0)}(\beta) \neq 0\right] = 1$ for all i and n.

A2. $\lim_{n\to\infty} W_{n,i}^{(k)}(\beta)$ and $\lim_{n\to\infty} \omega_{n,i}^{(k)}(\beta, Z)$ exist for $k \in \{0, 1\}$, and $\lim_{n\to\infty} W_{n,i}^{(0)}(\beta)$ and $\lim_{n\to\infty} \omega_{n,i}^{(0)}(\beta, Z)$ are nonzero.

Note that Conditions A1 and A2 also implies both $\tilde{W}_{n,i}^{(0)}(\beta)$ and $\lim_{n\to\infty} \tilde{W}_{n,i}^{(0)}(\beta)$ are nonzero.

To establish the desired asymptotic equivalence, it is sufficient to show that $E_Z\left[\omega_{n,i}^{(1)}(\beta,Z)/\omega_{n,i}^{(0)}(\beta,Z)\right]$ is asymptotically equivalent to $\tilde{W}_{n,i}^{(1)}(\beta)/\tilde{W}_{n,i}^{(0)}(\beta)$. We first have

$$E_Z\left[\frac{\omega_{n,i}^{(1)}(\beta,Z)}{\omega_{n,i}^{(0)}(\beta,Z)}\right] = \int_{\mathbb{R}^p}\frac{\omega_{n,i}^{(1)}(\beta,z)}{\omega_{n,i}^{(0)}(\beta,z)}\phi(z)\mathrm{d}z = \frac{\tilde{W}_{n,i}^{(1)}(\beta)}{\tilde{W}_{n,i}^{(0)}(\beta)} + \int_{\mathbb{R}^p}\left[\frac{1}{\omega_{n,i}^{(0)}(\beta,z)} - \frac{1}{\tilde{W}_{n,i}^{(0)}(\beta)}\right]\omega_{n,i}^{(1)}(\beta)\phi(z)\mathrm{d}z \quad \text{(A1)}$$

where $\phi(\cdot)$ is the density function of the p-dimensional standard normal random vector. The first term in (A1) is the ratio of two expectations. In order to show that the second term in (A1) is asymptotically negligible, it suffices to establish that $\omega_{n,i}^{(0)}(\beta)$ and $\tilde{W}_{n,i}^{(0)}(\beta)$ converges to the same limit.

By the triangular inequality, we have

$$\left|\omega_{n,i}^{(0)}(\beta,Z) - \tilde{W}_{n,i}^{(0)}(\beta)\right| \leqslant \left|\omega_{n,i}^{(0)}(\beta,Z) - W_{n,i}^{(0)}(\beta)\right| + \left|W_{n,i}^{(0)}(\beta) - \tilde{W}_{n,i}^{(0)}(\beta)\right|.$$

One can show $\left|\omega_{n,i}^{(0)}(\beta) - W_{n,i}^{(0)}(\beta)\right| \to 0$ as $n\to\infty$ by the uniform strong law of large numbers [34, Section 8]. On the other hand, because $W_{n,i}^{(0)}(\beta)$ is a monotone function, we have $\left|W_{n,i}^{(0)}(\beta) - \tilde{W}_{n,i}^{(0)}(\beta)\right| \to 0$ as $n\to\infty$ by Brown and Wang [8].

A.2. Validation of the sandwich variance estimation

To justify the proposed sandwich variance estimator, it suffices to show that, as $n\to\infty$,

(i) $\left\|n^{-1}\,\partial\tilde{U}_{n,\varphi}(\beta)/\partial\beta\Big|_{\beta=\beta_0} - A\right\| \to 0;$

(ii) $\left\|n^{-1/2}\{\tilde{U}_{n,\varphi}(\beta) - U_{n,\varphi}(\beta)\}\right\| \to 0.$

First, (i) can be shown using the arguments similar to those in [12, Lemma 3]. Let $\tilde{U}_{n,\varphi}(\beta,\beta) = \tilde{U}_{n,\varphi}(\beta)$, we have

$$U_{n,\varphi}(\beta) = \sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left[X_i - \frac{W_{n,i}^{(1)}(\beta)}{W_{n,i}^{(0)}(\beta)}\right] \text{ and } \tilde{U}_{n,\varphi}(\beta) = \sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left[X_i - \frac{\tilde{W}_{n,i}^{(1)}(\beta)}{\tilde{W}_{n,i}^{(0)}(\beta)}\right].$$

To verify (ii), we observe the following:

$$\begin{aligned}
&\left\|n^{-1/2}\left\{\tilde{U}_{n,\varphi}(\beta) - U_{n,\varphi}(\beta)\right\}\right\| \\
&= \left\|n^{-1/2}\sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left\{\left(\frac{W_{n,i}^{(1)}(\beta)}{W_{n,i}^{(0)}(\beta)} - \frac{W_{n,i}^{(1)}(\beta)}{\tilde{W}_{n,i}^{(0)}(\beta)}\right) + \left(\frac{W_{n,i}^{(1)}(\beta)}{\tilde{W}_{n,i}^{(0)}(\beta)} - \frac{\tilde{W}_{n,i}^{(1)}(\beta)}{\tilde{W}_{n,i}^{(0)}(\beta)}\right)\right\}\right\| \\
&\leqslant \left\|n^{-1/2}\sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left\{W_{n,i}^{(1)}(\beta)\left(\frac{\tilde{W}_{n,i}^{(0)}(\beta) - W_{n,i}^{(0)}(\beta)}{W_{n,i}^{(0)}(\beta)\tilde{W}_{n,i}^{(0)}(\beta)}\right) + \frac{n^{-1}\sum_{j=1}^{n}X_j\left\{I\left[e_j(\beta)\geqslant e_i(\beta)\right] - \Phi_i\left[\kappa_{ij}(\beta)\right]\right\}}{\tilde{W}_{n,i}^{(0)}(\beta)}\right\}\right\| \\
&\leqslant \left\|n^{-1/2}\sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left\{\frac{W_{n,i}^{(1)}(\beta)}{W_{n,i}^{(0)}(\beta)\tilde{W}_{n,i}^{(0)}(\beta)} - \frac{n^{-1}\sum_{j=1}^{n}X_j}{\tilde{W}_{n,i}^{(0)}(\beta)}\right\}\right\|\cdot\left|I\left[e_j(\beta)\geqslant e_i(\beta)\right] - \Phi_i\left[\kappa_{ij}(\beta)\right]\right| \\
&\leqslant \left\|n^{-1/2}\sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\left\{\frac{W_{n,i}^{(1)}(\beta)}{W_{n,i}^{(0)}(\beta)\tilde{W}_{n,i}^{(0)}(\beta)} - \frac{n^{-1}\sum_{j=1}^{n}X_j}{\tilde{W}_{n,i}^{(0)}(\beta)}\right\}\right\|\cdot\left|\Phi_i\left[-|\kappa_{ij}(\beta)|\right]\right| \\
&= \left\|n^{-1}\sum_{i=1}^{n}\Delta_i\varphi_i(\beta)\frac{n^{1/2}}{\kappa_{ij}(\beta)}\left\{\frac{W_{n,i}^{(1)}(\beta)}{W_{n,i}^{(0)}(\beta)\tilde{W}_{n,i}^{(0)}(\beta)} - \frac{n^{-1}\sum_{j=1}^{n}X_j}{\tilde{W}_{n,i}^{(0)}(\beta)}\right\}\right\|\cdot\left|\kappa_{ij}(\beta)\Phi\left[-|\kappa_{ij}(\beta)|\right]\right|.
\end{aligned} \quad \text{(A2)}$$

By the strong law of large numbers for U-statistics [35, Section 5.4], the first term in (A2) converges in probability to some finite limit. Because $\kappa_{ij}(\beta)$ is $O_p\left(n^{1/2}\right)$, $\Phi(-x) \leqslant \left(\sqrt{2\pi x}\right)^{-1} \exp\left(-x^2/2\right)$ and $\lim_{x\to\infty} x\Phi(-x) = 0$, the second term of (A2) also converges to 0 in probability. Therefore, (ii) follows from Slutsky's theorem.

The asymptotic properties of the proposed estimator with general weights from using a non-smooth rank-based estimating equations have been established in [7]. The asymptotic equivalence between the non-smooth and induced smoothed estimator can be readily shown using the arguments in [12] or [15].

Acknowledgements

Part of the computation was done on the Beowulf cluster of the Department of Statistics, University of Connecticut, partially financed by the National Science Foundation Scientific Computing Research Environments for the Mathematical Sciences (SCREMS) grant number 0723557. This work was supported by the National Science Foundation [DMS1209022 to J. Y. and S. K.] and National Research Foundation of Korea [NRF-2014R1A1A2055898 to S. K.].

References

1. Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B: Methodological* 1972; **34**:187–220.
2. Prentice RL. Linear rank tests with right censored data. *Biometrika* 1978; **65**:167–180.
3. Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* 1990; **18**:354–372.
4. Ying Z. A large sample study of rank estimation for censored regression data. *The Annals of Statistics* 1993; **21**:76–99.
5. Fygenson M, Ritov Y. Monotone estimating equations for censored data. *The Annals of Statistics* 1994; **22**(2):732–746.
6. Gehan EA. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 1965; **52**:203–223.
7. Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika* 2003; **90**(2): 341–353.
8. Brown BM, Wang YG. Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* 2005; **92**(1): 149–158.
9. Brown BM, Wang YG. Induced smoothing for rank regression with censored survival times. *Statistics in Medicine* 2007; **26**(4):828–836.
10. Li H, Zhang J, Tang Y. Induced smoothing for the semiparametric accelerated hazards model. *Computational Statistics and Data Analysis* 2012; **56**(4):4312–4319.
11. Chiou SH, Kang S, Yan J. Fast accelerated failure time modeling for case-cohort data. *Statistics and Computing* 2014; **24**(4):559–568.
12. Johnson LM, Strawderman RL. Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika* 2009; **96**(3):577–590.
13. Fu L, Wang YG, Bai Z. Rank regression for analysis of clustered data: a natural induced smoothing approach. *Computational Statistics and Data Analysis* 2010; **54**:1036–1050.
14. Wang YG, Fu L. Rank regression for accelerated failure time model with clustered and censored data. *Computational Statistics and Data Analysis* 2011; **55**:2334–2343.
15. Chiou SH, Kang S, Yan J. Semiparametric accelerated failure time modeling for clustered failure times from stratified sampling. *Journal of the American Statistical Association* 2014. DOI:10.1080/01621459.2014.917978.
16. Fu L, Wang YG. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics and Data Analysis* 2012; **56**:2526–2538.
17. Pang L, Lu W, Wang HJ. Variance estimation in censored quantile regression via induced smoothing. *Computational Statistics and Data Analysis* 2012; **56**(4):785–796.
18. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**:1–11.
19. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika* 1982; **69**:133–143.
20. Varadhan R, Gilbert P. BB: an R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software* 2009; **32**(4):1–26. [Accessed on 17 December 2014].
21. Chiou SH, Kang S, Yan J. Fitting accelerated failure time models in routine survival analysis with R package aftgee. *Journal of Statistical Software* 2014; **61**(11):1–23. [Accessed on 17 December 2014].
22. Zeng D, Lin DY. Efficient resampling methods for nonsmooth estimating functions. *Biostatistics* 2008; **9**(2):355–363.
23. Harnish JD, Aseltine RH, JR., Gore S. Resolution of stressful experiences as an indicator of coping effectiveness in young adults: an event history analysis. *Journal of Health and Social Behavior* 2000; **41**(2):121–136.
24. Brown GW, Harris T. *Social Origins of Depression. A Study of Psychiatric Disorder in Women*. Tavistock Publications: London, 1978.
25. D'Angio GJ, Breslow N, Beckwith JB, Evans A, Baum E, De Lorimier A, Fernbach D, Hrabovsky E, Jones B, Kelalis P, Othersen HB, Tefft M, Thomas PRM. Treatment of Wilms' tumor. Results of the Third National Wilms' Tumor Study. *Cancer* 1989; **64**:349–360.

26. Green D, Breslow N, Beckwith JB, Finklestein J, Grundy P, Thomas P, Kim T, Shochat S, Haase G, Ritchey M, Kelalis P, D'Angio GJ. Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: a report from the National Wilms' Tumor Study Group. *Journal of Clinical Oncology* 1998; **16**(1):237–245.
27. Kulich M, Lin D. Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association* 2004; **99**(467):832–844.
28. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Statistics in Biosciences* 2009; **1**(1):32–49.
29. Kulich M, Lin D. Additive hazards regression for case-cohort studies. *Biometrika* 2000; **87**(1):73–87.
30. Therneau T. *A package for survival analysis in S* 2012. R package version 2.36-12.
31. Chiou SH, Kang S, Yan J. *aftgee: accelerated failure time model with generalized estimating equations* 2012. R package version 0.2-27.
32. Jin Z, Shao Y, Ying Z. A Monte Carlo method for variance estimation for estimators based on induced smoothing. *Biostatistics* 2015; **16**(1):179–188.
33. Buyske S, Fagerstrom R, Ying Z. A class of weighted log-rank tests for survival data when the event is rare. *Journal of the American Statistical Association* 2000; **95**(449):249–258.
34. Pollard D. *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics: Hayward, 1990.
35. Serfling RJ. *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics. John Wiley & Sons: New York, 2001.