



Published in final edited form as:

Biometrics. 2023 December ; 79(4): 3111–3125. doi:10.1111/biom.13895.

A Semiparametric Cox-Aalen Transformation Model with Censored Data

Xi Ning¹, Yinghao Pan¹, Yanqing Sun^{1,*}, Peter B. Gilbert^{2,3}

¹Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.

²Department of Biostatistics, University of Washington, Seattle, Washington, U.S.A.

³Vaccine and Infectious Disease and Public Health Sciences Divisions, Fred Hutchinson Cancer Center, Seattle, Washington, U.S.A.

Summary:

We propose a broad class of so-called Cox-Aalen transformation models that incorporate both multiplicative and additive covariate effects on the baseline hazard function within a transformation. The proposed models provide a highly flexible and versatile class of semiparametric models that include the transformation models and the Cox-Aalen model as special cases. Specifically, it extends the transformation models by allowing potentially time-dependent covariates to work additively on the baseline hazard and extends the Cox-Aalen model through a predetermined transformation function. We propose an estimating equation approach and devise an Expectation-Solving (ES) algorithm that involves fast and robust calculations. The resulting estimator is shown to be consistent and asymptotically normal via modern empirical process techniques. The ES algorithm yields a computationally simple method for estimating the variance of both parametric and nonparametric estimators. Finally, we demonstrate the performance of our procedures through extensive simulation studies and applications in two randomized, placebo-controlled HIV prevention efficacy trials. The data example shows the utility of the proposed Cox-Aalen transformation models in enhancing statistical power for discovering covariate effects.

Keywords

Cox-Aalen model; ES algorithm; Estimating equations; Time-dependent covariates; Transformation models

* yasun@charlotte.edu .

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 3, 4 and 5, along with the simulated datasets and R code, are all available with this paper at the Biometrics website on Wiley Online Library. In addition, the dataset and R code are also available at <https://github.com/xining1/CoxAalen-transformation>.

1. Introduction

Censored failure time data are frequently encountered in epidemiological and biomedical studies. In the literature, the multiplicative and additive hazards models provide two principal frameworks for analyzing such data. The most popular multiplicative hazards model is the proportional hazards model (Cox, 1972), where the covariates are assumed to act multiplicatively on an unknown baseline hazard function. In contrast, the additive hazards models furnish an additive effect between the covariates and the baseline hazard function, enabling a direct reflection of the risk increase or decrease (Aalen, 1980; Huffer and McKeague, 1991; Lin and Ying, 1994). Without prior domain knowledge, it is hard to determine which approach is preferable among multiplicative and additive hazards models. In fact, both models may often be used to complement each other and provide more complete insights. Therefore, various multiplicative-additive hazard models have been proposed to capture both multiplicative and additive effects (Lin and Ying, 1995; Martinussen and Scheike, 2002). In particular, Scheike and Zhang (2002) suggested a Cox-Aalen model by replacing the baseline hazard function in the Cox model with Aalen's additive model. The Cox-Aalen model has been studied for various types of censored data, e.g., right-censored (Scheike and Zhang, 2002), interval-censored (Boruvka and Cook, 2015), left-truncated and right-censored (Shen and Weng, 2018), left-truncated and mixed interval-censored (Shen and Weng, 2019), and recurrent-event (Qu and Sun, 2019).

Transformation models have also received wide attention in survival analysis. Dabrowska and Doksum (1988) introduced the class of linear transformation models, which includes the proportional hazards and proportional odds models (Pettitt, 1982; Bennett, 1983). Estimators for this class of models were proposed by Dabrowska and Doksum (1988), Cheng et al. (1995), Fine et al. (1998), Chen et al. (2002), among others. Zeng and Lin (2006) extended the linear transformation models to allow time-dependent covariates. Hereafter, we refer to this class of transformation models as Zeng and Lin's model to avoid confusion. There is rich literature investigating Zeng and Lin's model. Zeng and Lin (2006) proposed a nonparametric maximum likelihood estimator (NPMLE) in the presence of right-censored data. Zeng and Lin (2007) derived a system of self-consistent equations for the jump sizes of the baseline cumulative hazard function at exact failure times through an expectation-maximization (EM) algorithm. Chen (2009) showed that the self-consistent estimator derived in Zeng and Lin (2007) is asymptotically equivalent to a weighted Breslow-type estimator, which can be solved by a computationally-efficient iterative reweighting algorithm. Liu and Zeng (2013) investigated variable selection procedures by minimizing a weighted negative partial log-likelihood function plus an adaptive lasso penalty. More recently, Zeng et al. (2016) and Zhou et al. (2021) studied the nonparametric maximum likelihood estimation of Zeng and Lin's model with interval-censored and partly interval-censored data, respectively.

However, one limitation of Zeng and Lin's model is that all covariate effects are assumed to be multiplicative within the transformation function. This assumption is too restrictive in some applications. For example, in an analysis of risk factors on mortality among patients with myocardial infarction, Scheike and Zhang (2003) showed that some covariates (e.g., ventricular fibrillation and congestive heart failure) have additive effects while others

(e.g., age and sex) have multiplicative effects. In addition, they pointed out that naively treating all covariates as multiplicative led to incorrect results when predicting survival probabilities. Another example is HIV prevention efficacy trials, for which HIV incidence varies across geographic regions and by sex/gender; thus, the different regions/sex/gender subgroups have different baseline hazard functions (Corey et al., 2021). Moreover, a Kaplan-Meier plot shows that survival curves for different regions cross, potentially suggesting an additive region effect. To the best of our knowledge, no existing work considers a class of semiparametric transformation models in which the baseline hazard function is allowed to depend on some potentially time-varying covariates additively. Therefore, it is desirable to provide a larger class of semiparametric transformation models that can accommodate both multiplicative and additive covariate effects under one unified framework.

The EM algorithm is a powerful tool for performing maximum likelihood estimation in the presence of latent variables or missing data (Dempster et al., 1977). In particular, various EM-type algorithms have been proposed to find NPMLE for semiparametric transformation models (Zeng and Lin, 2007; Liu and Zeng, 2013; Zeng et al., 2016; Zhou et al., 2021). In analogy to EM, Elashoff and Ryan (2004) proposed an expectation-solving (ES) algorithm that handles missing data for general estimating equations, greatly facilitating its application to a broader framework. When the complete-data estimating equations correspond to the score functions from the likelihood, the ES algorithm essentially reduces to the EM. The ES algorithm dramatically improves computational efficiency for solving estimating equations involving frailty or latent variables. For example, Johnson and Strawderman (2012) developed a smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. Henderson and Rathouz (2018) considered an approximate EM procedure for a longitudinal latent class model for count data.

In this article, we propose a broad class of so-called Cox-Aalen transformation models that incorporate both multiplicative and additive covariate effects upon the baseline hazard function within a transformation. The proposed class of models is very flexible and contains Zeng and Lin's model and the Cox-Aalen model as special cases. However, the multiplicative-additive structure within the transformation and the need to estimate several nonparametric functions simultaneously impose additional challenges for model estimation. To alleviate such difficulties, we devise an ES-type algorithm, which iterates between an E-step wherein functions of complete data are replaced by their expectations and an S-step where these expected values are substituted into the complete-data estimating equations, which are then solved. More specifically, within the S-step, the high-dimensional parameters are calculated explicitly, while the low-dimensional parameters are updated via the Newton-Raphson method. Consequently, the proposed ES algorithm is fast and stable even under a high percentage censoring rate, as evidenced by our simulation studies and real data applications. Another attraction of our approach is that we provide simple variance estimators for both parametric and nonparametric estimates. Furthermore, the theoretical properties of the proposed estimators are rigorously studied via modern empirical process techniques.

The rest of the article is organized as follows. In Section 2, we present the proposed Cox-Aalen transformation models. In Section 3, we formally describe the estimation procedure

and establish the asymptotic properties of the proposed estimators. In Section 4, simulation studies are conducted to evaluate the finite-sample performance of the proposed method. In Section 5, we apply our method to two randomized HIV prevention efficacy trials. Section 6 concludes with a discussion.

2. Semiparametric Cox-Aalen Transformation Models

Let $X(\cdot) = (X_1(\cdot), \dots, X_q(\cdot))^T$ and $Z(\cdot) = (Z_1(\cdot), \dots, Z_d(\cdot))^T$ denote $q \times 1$ and $d \times 1$ vectors of potentially time-varying covariates, and T denote the failure time of interest. We propose a broad class of so-called Cox-Aalen transformation models such that the cumulative hazard function for T conditional on $X(\cdot)$ and $Z(\cdot)$ takes the form

$$\Lambda(t|X(\cdot), Z(\cdot)) = G \left[\int_0^t \exp\{\beta^T Z(s)\} d\Lambda_X(s) \right], \quad (1)$$

where β is a $d \times 1$ vector of unknown regression coefficients, $\Lambda_X(s) = \int_0^s \{X^T(v)\alpha(v)\} dv$ is an unknown increasing function with $\alpha(v) = (\alpha_1(v), \dots, \alpha_q(v))^T$, and $G(\cdot)$ is a pre-specified transformation function that is strictly increasing and thrice continuously differentiable with $G(0) = 0$, $G'(0) > 0$ and $G(\infty) = \infty$. In addition, let $A(t) = \int_0^t \alpha(s) ds = (A_1(t), \dots, A_q(t))^T$, where $A_j(t) = \int_0^t \alpha_j(v) dv$ for $j = 1, \dots, q$. With $X_1(\cdot)$ fixed at 1, $\alpha_1(t)$ can be interpreted as a reference level of the risk. Generally, it is not meaningful to have $X(\cdot)$ equal or proportional to $Z(\cdot)$. For the choices of G , it is useful to consider the class of frailty-induced transformations

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi, \quad (2)$$

where $f(\xi)$ is a density function of a nonnegative random variable ξ with support $[0, \infty)$. The choice of the gamma density with unit mean and variance r for $f(\xi)$ yields the logarithmic transformations $G(x) = r^{-1} \log(1 + rx)$ ($r \geq 0$) with $r = 0$ specifying $G(x) = x$. The choice of the positive stable distribution with parameter $0 < \rho < 1$ yields the class of Box-Cox transformations $G(x) = \{(1 + x)^\rho - 1\} / \rho$. Note that $G(x) = \log(1 + x)$ is often considered to be a member of the above class with $\rho = 0$. By treating the latent variable ξ as missing, the frailty-induced transformations are particularly useful in deriving EM type algorithms (Zeng and Lin, 2007; Liu and Zeng, 2013; Zeng et al., 2016; Zhou et al., 2021). Some remarks regarding the Cox-Aalen transformation models are as follows:

REMARK 1: When $\alpha_j(t) = 0$ ($j = 2, \dots, q$) for any t , the right-hand side of (1) reduces to

$$G \left[\int_0^t \exp\{\beta^\top Z(s)\} dA_1(s) \right]. \quad (3)$$

Hence, Zeng and Lin's model is a special case of the proposed model. Moreover, when Z is time-invariant, (3) further reduces to a class of linear transformation models with the form $\log A_1(T) = -\beta^\top Z + \log G^{-1}(-\log \epsilon_0)$, where ϵ_0 has a uniform distribution (Chen et al., 2002). The choices $G(x) = x$ and $G(x) = \log(1 + x)$ yield the proportional hazards model and proportional odds model, respectively.

REMARK 2: When $G(x) = x$, according to (1), the cumulative hazard function on the left-hand side can be written as

$$\int_0^t \{X^\top(s)\alpha(s)\} \exp\{\beta^\top Z(s)\} ds. \quad (4)$$

Thus, the conditional hazard function of T is $X^\top(t)\alpha(t)\exp\{\beta^\top Z(t)\}$. Therefore, the Cox-Aalen model is a special case of the proposed models. In particular, when X_2, \dots, X_q represent levels in a set of factors, model (4) further reduces to the stratified Cox model (Kalbfleisch and Prentice, 2002, Section 4.4).

REMARK 3: for $G(x) = \log(1 + x)$, the odds of surviving beyond time t based on (1) are

$$\gamma(t|X, Z) = \frac{\Pr(T > t|X, Z)}{\Pr(T \leq t|X, Z)} = \left\{ \int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s) \right\}^{-1}.$$

In the special case where Z are time-independent covariates,

$$\gamma(t|X, Z) = \gamma(t|X, Z_0) \exp\{-\beta^\top (Z - Z_0)\},$$

which is a stratified proportional odds model when X is a random variable indicating strata.

As illustrated above, our proposed class of semiparametric models is very flexible and contains many popular models in survival analysis. To motivate our approach, we first set up the observed data likelihood and derive the NPMLE for a special case. Then, for more general situations, we propose estimating the parameters β and $A(\cdot)$ using estimating equations along with an easily-implemented ES algorithm. Finally, the asymptotic properties

of the resulting estimators are derived via modern empirical process theory (van der Vaart and Wellner, 1996).

3. Methods

3.1 Notations

For the i th individual, let T_i and C_i be the failure time and censoring time, respectively. Let $\tilde{T}_i = \min(T_i, C_i)$ be the observed time and define $\Delta_i = I(T_i \leq C_i)$. Thus, $\Delta_i = 1$ indicates that the exact failure time for the i th individual was observed while $\Delta_i = 0$ implies censoring. For a random sample of n participants, the observed data consist of $\mathcal{O}_i = \{\Delta_i, \tilde{T}_i, X_i(t), Z_i(t), t \in [0, \tau]\}$ for $i = 1, \dots, n$, where τ denotes the duration of the study. Moreover, we define $Y_i(t) = I(\tilde{T}_i \geq t)$ and $N_i(t) = \Delta_i I(\tilde{T}_i \leq t)$.

3.2 Nonparametric Maximum Likelihood Estimation

Assume that T_i and C_i are conditionally independent given $X_i(\cdot)$ and $Z_i(\cdot)$. Under the proposed Cox-Aalen transformation model (1), the likelihood for the observed data is

$$L_n(\beta, \Lambda_X) = \prod_{i=1}^n \left\{ \Lambda'_{X_i}(\tilde{T}_i) \exp\{\beta^T Z_i(\tilde{T}_i)\} G' \left[\int_0^{\tilde{T}_i} \exp\{\beta^T Z_i(s)\} d\Lambda_{X_i}(s) \right]^{\Delta_i} \times \exp \left(-G \left[\int_0^{\tilde{T}_i} \exp\{\beta^T Z_i(s)\} d\Lambda_{X_i}(s) \right] \right) \right\}, \tag{5}$$

where $\Lambda'_X(\cdot)$ and $G'(\cdot)$ are the derivatives of $\Lambda_X(\cdot)$ and $G(\cdot)$, respectively. The likelihood (5) involves β and q infinite dimensional parameters $A_j (j = 1, \dots, q)$, and it may not be concave in these parameters. Thus, the nonparametric maximum likelihood techniques are usually employed to restrict the parameter space.

To establish a simple and efficient estimation procedure, we adopt the idea in Zeng and Lin (2007) by treating ξ as a latent variable in the class of frailty-induced transformations (2). Note that model (1) is equivalent to the survival time T with cumulative hazard function

$$\Lambda(t|X(\cdot), Z(\cdot), \xi) = \xi \int_0^t \exp\{\beta^T Z(s)\} d\Lambda_X(s), \tag{6}$$

because

$$\begin{aligned}
 \Pr(T > t | X(\cdot), Z(\cdot)) &= E[\Pr\{T > t | X(\cdot), Z(\cdot), \xi\} | X(\cdot), Z(\cdot)] \\
 &= E\left[\exp\left[-\xi \int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\right] \middle| X(\cdot), Z(\cdot)\right] \\
 &= \int_0^\infty \exp\left[-\xi \int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\right] f(\xi) d\xi \\
 &= \exp\left(-G \int_0^t \exp\{\beta^\top Z(s)\} d\Lambda_X(s)\right).
 \end{aligned}$$

Based on (2), it can be shown that the likelihood (5) is equivalent to

$$\prod_{i=1}^n \int_{\xi_i} \left[\xi_i \Lambda'_{X_i}(\tilde{T}_i) \exp\{\beta^\top Z_i(\tilde{T}_i)\} \right]^{\Delta_i} \exp\left[-\xi_i \int_0^{\tilde{T}_i} \exp\{\beta^\top Z_i(s)\} d\Lambda_{X_i}(s)\right] f(\xi_i) d\xi_i.$$

Now we consider nonparametric maximum likelihood estimation of β and $A(\cdot)$. Let $0 = t_0 < t_1 < \dots < t_m < \infty$ denote the uniquely observed event times among the n observations. Assume that the estimator for $A_j(j = 1, \dots, q)$ is a step function with jump size a_{jk} at t_k . By the observation that $d\Lambda_X(t) = X^\top(t) dA(t)$, the estimator for Λ_X is a step function with jump size $X^\top(t_k) a_k$ at t_k where $a_k = (a_{1k}, \dots, a_{qk})^\top$ for $k = 1, \dots, m$. Let $\mathcal{O}_i^C = \{\Delta_i, \tilde{T}_i, X_i(t), Z_i(t), \xi_i, t \in [0, \tau]\}$ be the complete data for the i th participant. The complete-data log-likelihood function can be written as

$$\sum_{i=1}^n \left\{ \sum_{k=1}^m (\Delta_i I(\tilde{T}_i = t_k) [\log\{\xi_i(X_{ik}^\top a_k)\} + \beta^\top Z_{ik}]) - \xi_i \sum_{t_k \leq \tilde{T}_i} \exp(\beta^\top Z_{ik})(X_{ik}^\top a_k) + \log f(\xi_i) \right\}, \tag{7}$$

where $Z_{ik} = Z_i(t_k)$ and $X_{ik} = X_i(t_k)$.

To obtain the NPMLE of β and $A(\cdot)$, we propose an EM-type algorithm by treating ξ as missing data. In the E-step, we evaluate the posterior mean of ξ_i given the observed data, denoted by $\hat{E}(\xi_i)$. The detailed calculations are given in the next subsection. In the M-step, we maximize the expectation of (7) conditional on the observed data. More specifically, we set the derivatives of the conditional expectation of (7) with respect to $a_k(k = 1, \dots, m)$ and β to zeros, respectively. Then one can solve for the estimates through the following equations:

$$\sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_k) \frac{X_{ik}}{X_{ik}^\top a_k} - I(\tilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) X_{ik} \right\} = 0, \quad \text{for } k = 1, \dots, m, \tag{8}$$

$$\sum_{i=1}^n \sum_{k=1}^m \left\{ \Delta_i I(\tilde{T}_i = t_k) - I(\tilde{T}_i \geq t_k) \hat{E}(\xi_i) (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0.$$

(9)

The dimension of the unknown parameters (a_1, \dots, a_m, β) depends on m , which could be a large number when n is large or the censoring rate is low. Therefore, (8) and (9) are a system of high-dimensional nonlinear equations that is notoriously difficult to solve due to the curse of dimensionality. For a special case, i.e., X is a vector of design variables for categories, there exist explicit formulae for calculating the high-dimensional parameters $a_k (k = 1, \dots, m)$. See Web Appendix A for details. However, such explicit formulae do not exist for more general scenarios; hence we consider an alternative estimating equation approach to overcome the aforementioned computational difficulties.

3.3 Estimating Equations

Following Elashoff and Ryan (2004), we develop an expectation-solving (ES) algorithm for model (1) in this section. We begin by constructing a system of complete-data estimating equations based on model (6), which is equivalent to the proposed model (1) under the frailty-induced transformations (2). The connection between the proposed ES algorithm and the EM algorithm is discussed at the end of this section. Note that the intensity for $N_i(t)$ is $Y_i(t)\xi_i \exp\{\beta^\top Z_i(t)\} X_i^\top(t) \alpha(t)$ if ξ_i is known. Let

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \xi_i \exp\{\beta_0^\top Z_i(s)\} X_i^\top(s) dA_0(s),$$

where (β_0, A_0) are the true values of (β, A) . It is clear that $E\{X_i(t) dM_i(t)\} = 0$ for any $0 \leq t \leq \tau$ and $E\{\int_0^\tau Z_i(t) dM_i(t)\} = 0$. By treating ξ_i as missing, we consider the following complete-data estimating equations

$$\sum_{i=1}^n X_i(t) \left[dN_i(t) - Y_i(t) \xi_i \exp\{\beta^\top Z_i(t)\} X_i^\top(t) dA(t) \right] = 0 \quad (0 \leq t \leq \tau),$$

(10)

$$\sum_{i=1}^n \int_0^\tau Z_i(t) \left[dN_i(t) - Y_i(t) \xi_i \exp\{\beta^\top Z_i(t)\} X_i^\top(t) dA(t) \right] = 0.$$

(11)

By the previous arguments that the nonparametric estimator for Λ_X is a step function with jump size $X^\top(t_k)a_k$ at $t_k(k = 1, \dots, m)$, it follows that (10) and (11) can be written as

$$\begin{cases} \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_1) - I(\tilde{T}_i \geq t_1) \xi_i (X_{i1}^\top a_1) \exp(\beta^\top Z_{i1}) \right\} X_{i1} = 0 \\ \dots \\ \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_m) - I(\tilde{T}_i \geq t_m) \xi_i (X_{im}^\top a_m) \exp(\beta^\top Z_{im}) \right\} X_{im} = 0 \\ \sum_{i=1}^n \sum_{k=1}^m \left\{ \Delta_i I(\tilde{T}_i = t_k) - I(\tilde{T}_i \geq t_k) \xi_i (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0. \end{cases} \tag{12}$$

Write $\theta = (a_1^\top, \dots, a_m^\top, \beta^\top)^\top$. We propose to estimate θ through an ES-type algorithm by treating ξ_i as missing. The ES algorithm iterates between an E-step wherein the functions of the complete data are replaced by their expectations, and an S-step where these expected values are substituted into the complete-data estimating equations (12), which are then solved. After specifying initial values of the unknown parameters θ , say $\theta^{(0)}$, the proposed ES algorithm iterates between the following two steps until convergence:

E-step. Evaluate the posterior means $\hat{E}(\xi_i)$. When $\Delta_i = 1$, the posterior density of ξ_i given the observed data $(\Delta_i = 1, T_i, X_i, Z_i)$ is proportional to $\xi_i \exp(-\xi_i S_{i1}) f(\xi_i)$, where $S_{i1} = \Delta_i \sum_{t_k \leq T_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. Hence we obtain

$$\hat{E}(\xi_i) = G'(S_{i1}) - \frac{G''(S_{i1})}{G'(S_{i1})},$$

by taking the derivative twice of the equation $\exp\{-G(x)\} = \int_0^\infty \exp(-x\xi) f(\xi) d\xi$, where $G'(\cdot)$ and $G''(\cdot)$ are the first and second derivatives of $G(\cdot)$, respectively. When $\Delta_i = 0$, the posterior density of ξ_i given the observed data $(\Delta_i = 0, C_i, X_i, Z_i)$ is proportional to $\exp(-\xi_i S_{i2}) f(\xi_i)$, where $S_{i2} = (1 - \Delta_i) \sum_{t_k \leq C_i} (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik})$. One can obtain $\hat{E}(\xi_i) = G'(S_{i2})$. Therefore, the E-step can be summarized as

$$\hat{E}(\xi_i) = \Delta_i \left\{ G'(S_{i1}) - \frac{G''(S_{i1})}{G'(S_{i1})} \right\} + (1 - \Delta_i) G'(S_{i2}).$$

S-step. After replacing ξ by $\hat{E}(\xi_i)$, we solve (12) for θ . To this end, we propose the following nonlinear Gauss-Seidel method (Ortega and Rheinboldt, 1970; Ortega, 1972).

Step 1. Fix β , update $a_k(k = 1, \dots, m)$ by solving

$$\begin{cases} \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_1) - I(\tilde{T}_i \geq t_1) \hat{E}(\xi_i)(X_{i1}^\top a_1) \exp(\beta^\top Z_{i1}) \right\} X_{i1} = 0 \\ \dots \\ \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_m) - I(\tilde{T}_i \geq t_m) \hat{E}(\xi_i)(X_{im}^\top a_m) \exp(\beta^\top Z_{im}) \right\} X_{im} = 0. \end{cases} \tag{13}$$

Note that for fixed β , (13) is a system of linear equations with respect to $a_k (k = 1, \dots, m)$ and updating a_k is independent of updating a_j for $k \neq j (j = 1, \dots, m)$. In particular, we have explicit formulae for updating $a_k (k = 1, \dots, m)$, i.e.,

$$a_k = \left\{ \sum_{i=1}^n I(\tilde{T}_i \geq t_k) \hat{E}(\xi_i) \exp(\beta^\top Z_{ik}) X_{ik} X_{ik}^\top \right\}^{-1} \left\{ \sum_{i=1}^n \Delta_i I(\tilde{T}_i = t_k) X_{ik} \right\}, \tag{14}$$

for $k = 1, \dots, m$.

Step 2. Fix a_1, \dots, a_m , we update β by solving the following equation using the Newton-Raphson method:

$$\sum_{i=1}^n \sum_{k=1}^m \left\{ \Delta_i I(\tilde{T}_i = t_k) - I(\tilde{T}_i \geq t_k) \hat{E}(\xi_i)(X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} = 0.$$

Note that within the S-step, we iterate between Step 1 and Step 2 until convergence. The S-step is declared convergent when the sum of the absolute differences of the estimates at two successive iterations is less than a small positive number, say 10^{-3} .

We iterate between the E- and S-steps until convergence and denote the final estimates by $\hat{\theta} = (\hat{a}_1^\top, \dots, \hat{a}_m^\top, \hat{\beta}^\top)^\top$. A natural estimator of $A(t)$ is $\hat{A}(t) = \sum_{t_k \leq t} \hat{a}_k$ for $0 \leq t \leq \tau$. Moreover, recall that $A(t) = \int_0^t \alpha(s) ds$, hence we can estimate $\alpha(t)$, $0 \leq t \leq \tau$ via a kernel estimator

$$\hat{\alpha}(t) = \sum_{k=1}^m h^{-1} K\left(\frac{t-t_k}{h}\right) \hat{a}_k,$$

where $K(x)$ is the kernel function and h is the bandwidth. Throughout this article, we choose the Epanechnikov kernel function, i.e., $K(x) = \frac{3}{4} \max\{1 - x^2, 0\}$.

The proposed ES algorithm has several desirable features. First, a closed-form formula for computing $\hat{E}(\xi_i)$ is obtained in the E-step. Second, it avoids solving a large system of nonlinear equations in the S-step because the high-dimensional parameters $a_k(k = 1, \dots, m)$ are calculated explicitly while the low-dimensional parameter β is updated via the Newton-Raphson method. Accordingly, the proposed ES algorithm performs stably and satisfactorily without calculating the inverse of any high-dimensional matrices. Third, when X is a vector of design variables for categories, the corresponding ES algorithm coincides with the EM algorithm proposed in Section 3.2 by observing that for fixed β , (8) and (13) share the same solution in terms of $a_k(k = 1, \dots, m)$. This implies that the proposed ES estimator is also efficient under this special case. See Web Appendix B for detailed justifications. Similarly, it can be shown that the ES algorithm coincides with the EM algorithm When $X = 1$, i.e., $q = 1$. Finally, we remark that (13) can be considered as a weighted version of (8), where each participant i is assigned weight $X_{ik}^\top a_k$.

3.4 Variance Estimator

In this section, we provide easy-to-compute variance estimators for both the parametric estimates $\hat{\beta}$ and the nonparametric estimates $\hat{A}(t)$, $\hat{a}(t)$. Note that $\hat{E}(\xi_i)$ is a function of the observed data \mathcal{O}_i and the unknown parameter θ : $\hat{E}(\xi_i) = g(\mathcal{O}_i, \theta)$. With \mathcal{O} the collection of $\mathcal{O}_i(i = 1, \dots, n)$, the proposed ES estimator is intrinsically equivalent to solving the following observed-data estimating equation: $U(\mathcal{O}, \theta) = 0$, where $U(\mathcal{O}, \theta) = (U_{a_1}, \dots, U_{a_m}, U_\beta)$,

$$\begin{cases} U_{a_1} = \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_1) - I(\tilde{T}_i \geq t_1) g(\mathcal{O}_i, \theta) (X_{i1}^\top a_1) \exp(\beta^\top Z_{i1}) \right\} X_{i1} \\ \dots \\ U_{a_m} = \sum_{i=1}^n \left\{ \Delta_i I(\tilde{T}_i = t_m) - I(\tilde{T}_i \geq t_m) g(\mathcal{O}_i, \theta) (X_{im}^\top a_m) \exp(\beta^\top Z_{im}) \right\} X_{im} \\ U_\beta = \sum_{i=1}^n \sum_{k=1}^m \left\{ \Delta_i I(\tilde{T}_i = t_k) - I(\tilde{T}_i \geq t_k) g(\mathcal{O}_i, \theta) (X_{ik}^\top a_k) \exp(\beta^\top Z_{ik}) \right\} Z_{ik} . \end{cases} \tag{15}$$

Note that $U_{a_1}, \dots, U_{a_m}, U_\beta$ also depend on the observed data \mathcal{O} and the unknown parameter θ . Here, we compress the notation when there is no confusion. From (15), one can easily note that $U(\mathcal{O}, \theta)$ can be expressed as the sum of independent terms:

$$U(\mathcal{O}, \theta) = \sum_{i=1}^n U_i(\mathcal{O}_i, \theta).$$

Let $D(\mathcal{O}, \theta)$ be the derivative of $U(\mathcal{O}, \theta)$ with respect to θ . The covariance matrix of $\hat{\theta}$ is consistently estimated by

$$D(\mathcal{O}, \theta)^{-1} \left\{ \sum_{i=1}^n U_i(\mathcal{O}_i, \theta) U_i^\top(\mathcal{O}_i, \theta) \right\} \left\{ D(\mathcal{O}, \theta)^{-1} \right\}^\top \Big|_{\theta = \hat{\theta}} .$$

(16)

Therefore, the variance covariance matrix of $\hat{\beta}$ can be consistently estimated by the $d \times d$ lower right-hand corner of (16). The variance covariance matrix of $\hat{a}_k (k = 1, \dots, m)$ can be consistently estimated by the $(qm) \times (qm)$ upper left-hand corner of (16).

In addition, recall that $\hat{A}(t) = \sum_{t_k \leq t} \hat{a}_k$ and $\hat{\alpha}(t) = \sum_{k=1}^m h^{-1} K\left(\frac{t-t_k}{h}\right) \hat{a}_k$ for $0 \leq t \leq \tau$, such that the variance for $\hat{A}(t)$ and $\hat{\alpha}(t)$ are

$$\text{Var}\{\hat{A}(t)\} = \sum_{t_k \leq t} \sum_{t_j \leq t} \text{Cov}(\hat{a}_k, \hat{a}_j),$$

$$\text{Var}\{\hat{\alpha}(t)\} = \sum_{k=1}^m \sum_{j=1}^m h^{-2} K\left(\frac{t-t_k}{h}\right) K\left(\frac{t-t_j}{h}\right) \text{Cov}(\hat{a}_k, \hat{a}_j).$$

Variance estimators are obtained by replacing $\text{Cov}(\hat{a}_k, \hat{a}_j)$ by $\widehat{\text{Cov}}(\hat{a}_k, \hat{a}_j)$ in the above expressions.

3.5 Asymptotic Properties

In this subsection, we present the asymptotic properties of the proposed ES estimator. let

$$\phi(t) = G'(t), \psi(t) = G''(t)/G'(t) \text{ and } \rho(t; \beta, A) = \int_0^t Y(s) \exp\{\beta^\top Z(s)\} X^\top(s) dA(s).$$

Hence, the posterior mean of ξ can be written as $g(\tau; \beta, A) = \phi(\rho(\tau; \beta, A)) - \Delta\psi(\rho(\tau; \beta, A))$.

Let P denote the true probability measure and \mathbb{P}_n denote the empirical measure. In addition, let $\theta = (\beta, A)$ be the parameters of interest and $\theta_0 = (\beta_0, A_0)$ be the true values of the parameters. Then the proposed ES estimator $\hat{\theta} = (\hat{\beta}, \hat{A})$ is a Z-estimator solving the following observed-data estimating equation

$$\mathbb{P}_n \Phi(\beta, A)(t) \equiv \mathbb{P}_n \begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)(t) \end{pmatrix} = 0,$$

for $0 \leq t \leq \tau$, where

$$\Phi_1(\beta, A) = \int_0^\tau [Z(t) dN(t) - Y(t) \exp\{\beta^\top Z(t)\} g(\tau; \beta, A) Z(t) X^\top(t) dA(t)],$$

$$\Phi_2(\beta, A)(t) = X(t)dN(t) - Y(t)\exp\{\beta^\top Z(t)\}g(\tau; \beta, A)X(t)X^\top(t)dA(t).$$

Let h be a function in $BV_1[0, \tau]$, where $BV_1[0, \tau]$ denotes the set of functions with total variation bounded by 1 on $[0, \tau]$. Define

$$\Phi_2(\beta, A)[h] = \int_0^\tau h(t)\left[X(t)dN(t) - Y(t)\exp\{\beta^\top Z(t)\}g(\tau; \beta, A)X(t)X^\top(t)dA(t)\right].$$

Similar to Gao et al. (2017) and van der Vaart and Wellner (1996, Section 3.3.1), the proposed ES estimator $(\hat{\beta}, \hat{A})$ is equivalent to the root of the estimating equation

$$\mathbb{P}_n\Phi(\beta, A)[h] \equiv \mathbb{P}_n\begin{pmatrix} \Phi_1(\beta, A) \\ \Phi_2(\beta, A)[h] \end{pmatrix} = 0,$$

for all $h \in BV_1[0, \tau]$.

Write $\Psi(\theta) = P\Phi(\beta, A)[h]$ and $\Psi_n(\theta) = \mathbb{P}_n\Phi(\beta, A)[h]$. Note that $\Psi(\theta)$ and $\Psi_n(\theta)$ are actually h -dependent. Rigorously speaking, we should write $\Psi(\theta)[h] = P\Phi(\beta, A)[h]$ and $\Psi_n(\theta)[h] = \mathbb{P}_n\Phi(\beta, A)[h]$, but in the rest of the article, we suppress the letter h when there is no confusion. The proposed ES estimator is a Z-estimator that satisfies $\Psi_n(\hat{\theta}) = 0$. To establish the asymptotic properties, we assume the following regularity conditions:

Condition 1. With probability one, $X(\cdot)$ and $Z(\cdot)$ have bounded total variation in $[0, \tau]$.

Condition 2. Let \mathcal{B} be a compact set of \mathcal{R}^d and $BV[0, \tau]$ be the class of functions with bound variation over $[0, \tau]$. The true parameter (β_0, A_0) belongs to $\mathcal{B} \times BV^q[0, \tau]$ with β_0 an interior point of \mathcal{B} and $A_0(t) = (A_{01}(t), \dots, A_{0q}(t))^\top$ is continuous over $[0, \tau]$ with $A_0(0) = 0$. Here $BV^q[0, \tau]$ denotes the product space $BV[0, \tau] \times \dots \times BV[0, \tau]$.

Condition 3. With probability one, there exists a positive constant ε such that $P(Y(\tau) = 1|X(\cdot), Z(\cdot)) > \varepsilon$ and $PN^2(\tau) < \infty$. If there exists a vector γ and a deterministic function $\gamma_0(t)$ such that $\gamma_0(t) + \gamma^\top X(t) = 0$ with probability one, then $\gamma_0(t) = 0$ and $\gamma = 0$ for any $t \in [0, \tau]$.

Condition 4. The transformation function G is thrice continuously differentiable on $[0, \infty)$ with $G(0) = 0$, $G'(x) > 0$ and $G(\infty) = \infty$.

Condition 5. Let $\dot{\Psi}_{\theta_0}$ be the Fréchet derivative of $\Psi(\theta)$ with respect to θ at $\theta = \theta_0$. See Web Appendix C for detailed expressions of $\dot{\Psi}_{\theta_0}$. We assume that $\dot{\Psi}_{\theta_0}$ is an invertible map.

REMARK 4: Conditions 1 and 2 state the boundedness of the covariates and the compactness of the Euclidean parameter space, which are conventional conditions used in most regression

analyses. Condition 3 ensures the existence and uniqueness of the jump sizes in (14). Condition 4 ensures that the transformation function G is strictly increasing on $[0, \infty)$. Condition 5 is a classical condition for Z-estimators.

THEOREM 1: *Under Conditions 1 – 5, the proposed ES estimator $(\hat{\beta}, \hat{A})$ is strongly consistent to (β_0, A_0) .*

THEOREM 2: *Under Conditions 1 – 5, $\sqrt{n}(\hat{\beta} - \beta_0, \hat{A} - A_0)$ converges weakly to a zero-mean Gaussian process in the metric space $\mathcal{R}^d \times \overline{\text{lin}}^q(BV_1[0, \tau])$.*

Here, we let $\overline{\text{lin}}(BV_1[0, \tau])$ be the closed linear span for linear functionals of $BV_1[0, \tau]$. For each $j(j = 1, \dots, q)$, A_j is contained in the Banach space $\overline{\text{lin}}(BV_1[0, \tau])$, where $A_j[h] = \int h(t)dA_j(t)$ for $h \in BV[0, \tau]$. Thus, $A = (A_1, \dots, A_q)^\top$ is contained in the Banach space $\overline{\text{lin}}^q(BV_1[0, \tau])$. Here, $\overline{\text{lin}}^q(BV_1[0, \tau])$ stands for the product space $\overline{\text{lin}}(BV_1[0, \tau]) \times \dots \times \overline{\text{lin}}(BV_1[0, \tau])$. Detailed proofs of the above theorems are presented in Web Appendix C.

4. Simulation Studies

We carried out extensive simulation studies to evaluate the finite-sample performance of the proposed estimation and inference procedures. Suppose that the failure time T follows the Cox-Aalen transformation model with the cumulative hazard function

$$\Lambda(t|X(\cdot), Z(\cdot)) = G \left[\int_0^t \exp\{\beta_1 Z_1(s) + \beta_2 Z_2\} d\Lambda_X(s) \right].$$

Here, $Z_1(t) = B_1 I(t \leq V) + B_2 I(t > V)$ is a time-dependent covariate where B_1 and B_2 are independent $\text{Ber}(0.5)$, $V \sim \text{Unif}(0, 3)$, and $Z_2 \sim \text{Unif}(0, 1)$ is a time-independent covariate. We set $\beta_1 = 0.5$, $\beta_2 = -0.5$, and consider four different configurations for $\Lambda_X(s) = \int_0^s X^\top(v)dA(v)$, $A(t) = (A_1(t), \dots, A_q(t))^\top$:

Scenario 1. $X = (1, X_2)^\top$ with $X_2 \sim \text{Ber}(0.4)$, $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 2. $X = (1, X_2)^\top$ with $X_2 \sim \text{Unif}(0, 1)$, $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 3. $X(t) = (1, X_2(t))^\top$ with $X_2(t) = B_3 + B_4 t$, where $B_3 \sim \text{Unif}(1, 2)$ and $B_4 \sim \text{Unif}(0.1, 0.5)$, $A_1(t) = \log(1 + t/4)$ and $A_2(t) = 0.1t$.

Scenario 4. Let D be a categorical variable that takes values in $\{1, 2, 3\}$ with equal probability. $X = (1, X_2, X_3)^\top$, where $X_2 = I(D = 2)$, $X_3 = I(D = 3)$, $A_1(t) = \log(1 + t/4)$, $A_2(t) = 0.1t$ and $A_3(t) = 0.05t$.

For the transformation functions, we consider the class of logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ with $r = 0, 0.5$ and 1 , where $r = 0$ specifies the Cox-Aalen model. For all setups, we let $\tau = 1$ be the duration of the study. For each study participant, we generate one censoring time $C \sim \text{Exponential}(0.5)$. We set $\Delta = 1$ if $T \leq \min(C, \tau)$, and 0 otherwise. This process yields about 75 – 85% right-censored observations for $r = 0, 0.5$ and 1 . For each dataset, we applied the proposed ES algorithm by setting the initial value of β to 0 and the initial value of a_k to be $(1/m, 0, \dots, 0)$ for each $k = 1, \dots, m$. We also tried other initial values, yielding almost identical results. We set $n = 200, 500$ or 800 , and all simulation results are based on 1000 replicates.

Table 1 summarizes the results for estimation of β_1 and β_2 for all scenarios. Despite the high censoring percentage, from Table 1, one can see that the proposed procedures perform well in several ways: (i) the estimators are virtually unbiased; (ii) the estimated standard error is fairly close to the empirical standard error; (iii) the empirical coverage probability of 95% confidence intervals are all close to the nominal 95% level; (iv) when the sample size increases, the bias, and the variability of the parameter estimator, decreases. Thus, our proposed estimation procedures are reliable for various Cox-Aalen transformation models.

Figure 1 shows the estimation results for the cumulative regression functions $A(\cdot)$ in Scenario 1. The proposed estimators are again virtually unbiased and the estimated curves are able to capture the shapes of the true cumulative regression functions very well; the estimated standard errors are close to the empirical standard errors; and the confidence intervals have reasonably accurate coverage probabilities. To save space, estimation results for $\alpha(\cdot)$ via the kernel smoothing approach with bandwidth $h = 0.1$ are provided in Web Appendix D for Scenario 1. In addition, estimation results for $A(\cdot)$ and $\alpha(\cdot)$ under Scenarios 2 and 4 are also presented in Web Appendix D. These results further confirm the satisfactory performance of our proposed method in various numerical settings. We also conducted simulation studies to investigate the robustness of the proposed estimator under the misspecification of the G function. The setups were the same as Scenario 3, and simulation results are displayed in Web Appendix D. The results suggested that the misspecification of the transformation function led to biased estimates and lower coverage probabilities than the nominal levels.

Moreover, we demonstrate the superiority of our proposed model over Zeng and Lin's model in one simulation example. Specifically, we generated the data from our proposed Cox-Aalen transformation model where one covariate has a multiplicative effect, and the other has an additive effect. If we falsely assume that both covariates have multiplicative effects and fit Zeng and Lin's model, we will obtain biased estimates of the survival function and cumulative hazard. Thus, our proposed model can better capture complex hazard functions. See Web Appendix D for details.

5. An HIV Prevention Study Example

In this section, we apply the proposed model and methods to two harmonized randomized trials, HIV Vaccine Trials Network (HVTN) 704/HIV Prevention Trials Network (HPTN) 085 and HVTN 703/HPTN 081 (Corey et al., 2021), designed to determine whether a

broadly neutralizing monoclonal antibody (bnAb) can prevent the acquisition of human immunodeficiency virus type 1 (HIV). The HVTN 704/HPTN 085 trial enrolled 2687 men who have sex with men and transgender persons in the Americas and Europe, and HVTN 703/HPTN 081 enrolled 1924 females in sub-Saharan Africa. For each trial, HIV uninfected participants were randomly assigned in 1:1:1 ratio to receive infusions of a bnAb (VRC01) at a dose of 10 mg per kilogram of body weight (low-dose group), VRC01 at 30 mg per kilogram (high-dose group) or saline placebo, administered at 8-week intervals for 10 total infusions. The primary efficacy endpoint was diagnosis of HIV infection by the week 80 trial visit, and HIV testing was conducted at each 4-week trial visit starting at week 0. For participants acquiring HIV infection, the diagnosis date was determined by the adjudicated diagnosis date based on validated assays (Corey et al., 2021). Participant follow-up is right-censored by the minimum of their last negative HIV sample collection date and $\tau = 85.9$ weeks (Corey et al., 2021). Therefore, the observed data consist of exact and right-censored observations.

Among the 4559 HIV negative participants from both trials, 1401 are in the U.S. and Switzerland, 1249 in Brazil and Peru, 1009 in South Africa, and 900 in other sub-Saharan African countries (Switzerland was pooled with the U.S. given few participants in Switzerland). We analyze the two trials pooled together, which is valid given the harmonized protocols such that essentially the study is one trial in two distinct study populations. There were a total of 174 HIV infection diagnosis endpoints in the two trials pooled, including 60 out of 1520 participants in the low-dose group, 47 out of 1520 in the high-dose group, and 67 out of 1519 in the placebo group. The numbers of HIV infection diagnosis endpoints by region are reported in Web Appendix E. Participants were categorized by age (in years old) into four groups, [17, 20], [21, 30], [31, 40] and [41, 52], with 540, 2651, 1102 and 266 participants, respectively.

Figure 2 reveals that the risk of HIV infection diagnosis in different regions crosses over. Therefore, without imposing proportional hazards for different regions, we consider the following Cox-Aalen transformation model to assess the association between treatment assignment, age, and region with the time since the first infusion to HIV infection diagnosis:

$$\Lambda(t|X, Z) = G \left[\int_0^t \exp(\beta^T Z) d\Lambda_X(s) \right],$$

where β is the unknown regression coefficients and $\Lambda_X(s) = \int_0^s \{X^T \alpha(v)\} dv = X^T A(s)$ with $A(s) = (A_1(s), \dots, A_4(s))^T$. Here, $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)^T$, where Z_1 and Z_2 are indicators of being assigned to the low-dose and high-dose group, respectively, with the placebo group as the reference group; Z_3, Z_4, Z_5 are indicators of the age groups [21, 30], [31, 40] and [41, 52], respectively, with [17, 20] as the reference age group. In addition, let $X = (1, X_2, X_3, X_4)^T$, where X_2, X_3, X_4 are indicators of participants from Brazil and Peru, South Africa, and other sub-Saharan African countries, respectively. The participants from USA and Switzerland are considered as the reference group.

We conducted the analysis using the class of logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$, with r values ranging from 0 to 3 with an increment of 0.1. The r value that maximizes the log-likelihood function evaluated at the final parameter estimates was selected. The log-likelihood is maximized at $r = 0$, though the values do not change greatly for different values of r due to a high censoring rate (about 96.2%); this phenomenon is verified in our simulation studies (see Web Appendix E for details).

The lower panel of Table 2 shows the regression parameter estimates for the selected transformation function ($r = 0$). High-dose VRC01 significantly lowers the risk of HIV infection diagnosis, while low-dose VRC01 does not. The model fit also shows a significant association between older age and a lower risk of HIV infection diagnosis. Figure 3 displays the estimated baseline cumulative hazard function $\hat{\Lambda}(t|X, Z = 0)$ for the four different regions. The risk of HIV infection diagnosis is the highest in Brazil and Peru and lowest in the U.S. and Switzerland. The estimates for South Africa and other sub-Saharan African countries cross; in particular, South Africa has a lower risk at early times after the first infusion but a higher risk at later times. In addition, Figure 3 shows that the HIV infection diagnosis hazards are not proportional across geographic regions. Figure 4 plots the estimates of conditional survival functions at sixteen different combinations of covariates: four age groups crossed with four regions. This figure further confirms our findings above. In Web Appendix E, we also report the analysis results under other values of r and observe the same patterns.

The four other panels of Table 2 (upper panels) show results from Zeng and Lin's model fit to each of the four geographic regions separately; this method was not applied to the full cohort (pooled) data because it cannot flexibly model the differences in baseline cumulative hazards and the diagnostics support lack of fit. In these results, the p-values for the effect of high-dose VRC01 markedly increase, and the coefficient estimates for the age group [41, 52] are unstable because there are very few HIV infection diagnosis endpoints in this age group in the three regions Brazil and Peru, South Africa, and other sub-Saharan African countries. Therefore, the results from the Cox-Aalen transformation modeling – which could be based on the full cohort data through flexible specifications of the baseline cumulative hazard functions – provide new insights with improved precision and power beyond insights achieved from the application of Zeng and Lin's model.

6. Discussion

In this article, we proposed a class of semiparametric Cox-Aalen transformation models that includes Zeng and Lin's model (Zeng and Lin, 2006; Zeng et al., 2016) and the Cox-Aalen model (Scheike and Zhang, 2002) as special cases. By considering the class of frailty-induced transformations, we successfully developed a fast and stable ES algorithm to estimate the parametric and nonparametric components of the proposed model along with easy-to-compute variance estimators. In addition, the asymptotic properties of our proposed estimators are rigorously studied. Elashoff and Ryan (2004) pointed out that an ES algorithm can be regarded as a block Newton-Gauss-Seidel algorithm (see Ortega 1972, p.146). Following Ortega (1972, p.147), an ES algorithm converges locally to the solution, $\hat{\theta}$, of $U(\theta) = 0$ if the Jacobian matrix $D = \partial U / \partial \theta$ is nonsingular at $\theta = \hat{\theta}$ and the largest

eigenvalue of $D^{-1}(\hat{\theta})$ is less than 1. For general estimating equations, the two conditions above are difficult to verify in advance, especially for the second condition. Nevertheless, the matrix D is needed to calculate the variance of $\hat{\theta}$ in (16), and hence one can check the required conditions numerically.

In real data applications, we ascertain whether a covariate has a multiplicative or additive effect based on the following criteria. First, we may employ the underlying biological, physical meaning, or other domain knowledge for decision-making. Second, initial data exploration can be performed for each covariate, such as drawing the Kaplan-Meier (KM) plot. If the KM curves cross, this covariate should be modeled additively. Third, similar to Qu and Sun (2019), Yu et al. (2019), we may employ some AIC or BIC-based procedures. In particular, all possible combinations of covariate effects will be examined. However, it is easy to see that this is inefficient when there are many covariates. In addition, Scheike and Zhang (2003) proposed supremum tests to determine the multiplicative and additive parts of the Cox-Aalen model. It is worthwhile to investigate if similar testing procedures can be constructed for our proposed model. More theoretical and numerical studies are needed, which we leave for future work.

Indeed, the outlined procedures for determining the covariates for the multiplicative or additive components are valid for a given class of the Cox-Aalen transformation model. In practice, when the KM curves corresponding to different values or groups of values of a certain covariate, say X , cross, the effects of this covariate are not necessarily additive. In addition to the possibility that the effect of X is additive, there are a number of ways that the proportionality can fail, e.g., X not as a part of Z under the Cox-Aalen transformation model. For example, the effect of X could be time-varying as in the Cox model with time-varying regression coefficients studied by Cai and Sun (2003). In another scenario, the contributions from the additive components and the multiplicative components of the model may not be in the multiplicative form. In addition to choosing the right covariates for the multiplicative and additive components of the model, mis-specifying the transformation function can result in erroneous inferences.

Assessing the adequacy of the proposed model is crucial because model misspecification affects the validity of inference and prediction accuracy. For Zeng and Lin's model, Chen et al. (2012) considered appropriate time-dependent residuals and constructed various graphical and numerical procedures for model assessment. In our analysis of the HIV prevention trial data, we use the log-likelihood to select the transformation function, even though the log-likelihood surface is relatively flat. Similar to Chen et al. (2012), we suggest constructing the cumulative sums of residuals over the argument of the transformation function, i.e., $W(x, t) = n^{-1/2} \sum_{i=1}^n I\left(\int_0^t Y_i(s) e^{\hat{\beta}^\top Z_i(s) X_i^\top(s) d\hat{A}(s)} \leq x\right) dM_i(u; \hat{\beta}, \hat{A})$ to check the transformation form, where $M_i(t; \beta, A) = N_i(t) - G\left\{\int_0^t Y_i(s) e^{\beta^\top Z_i(s) X_i^\top(s) dA(s)}\right\}$. A thorough theoretical and numerical investigation of model misspecification is still needed for the proposed model. We are currently pursuing this direction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research is part of Xi Ning's Ph.D. dissertation. This research was partially supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases [grant numbers UM1 AI68635 and R37 AI054165]. The research of Xi Ning was also supported, in part, by the 2022 Graduate School Summer Fellowship Program from the University of North Carolina at Charlotte (UNC Charlotte). Yinghao Pan's work was partially supported by funds provided by UNC Charlotte. The research of Yanqing Sun was also partially supported by the National Science Foundation [grant number DMS-1915829] and the Reassignment of Duties fund provided by UNC Charlotte. We thank the HIV Vaccine Trials Network (HVTN) for providing the data analyzed in this article, especially Dr. Yunda Huang. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DATA AVAILABILITY STATEMENT

The data that support the findings of this paper are available in the Supporting Information of this article.

References

- Aalen O (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory*, pages 1–25. Springer.
- Bennett S (1983). Analysis of survival data by the proportional odds model. *Statistics in medicine* 2, 273–277. [PubMed: 6648142]
- Boruvka A and Cook RJ (2015). A Cox-Aalen model for interval-censored data. *Scandinavian Journal of Statistics* 42, 414–426.
- Cai Z and Sun Y (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics* 30, 93–111.
- Chen K, Jin Z, and Ying Z (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* 89, 659–668.
- Chen L, Lin D, and Zeng D (2012). Checking semiparametric transformation models with censored data. *Biostatistics* 13, 18–31. [PubMed: 21785165]
- Chen Y-H (2009). Weighted Breslow-type and maximum likelihood estimation in semi-parametric transformation models. *Biometrika* 96, 591–600.
- Cheng S, Wei L, and Ying Z (1995). Analysis of transformation models with censored data. *Biometrika* 82, 835–845.
- Corey L, Gilbert PB, Juraska M, Montefiori DC, Morris L, Karuna ST, Edupuganti S, Mgodini NM, deCamp AC, Rudnicki E, et al. (2021). Two randomized trials of neutralizing antibodies to prevent HIV-1 acquisition. *New England Journal of Medicine* 384, 1003–1014. [PubMed: 33730454]
- Cox DR (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 187–202.
- Dabrowska DM and Doksum KA (1988). Partial likelihood in transformation models with censored data. *Scandinavian journal of statistics* pages 1–23.
- Dempster AP, Laird NM, and Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39, 1–22.
- Elashoff M and Ryan L (2004). An EM algorithm for estimating equations. *Journal of Computational and Graphical Statistics* 13, 48–65.
- Fine J, Ying Z, and Wei L (1998). On the linear transformation model for censored data. *Biometrika* 85, 980–986.
- Gao F, Zeng D, and Lin D-Y (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics* 73, 1161–1168. [PubMed: 28444688]

- Henderson NC and Rathouz PJ (2018). AR (1) latent class models for longitudinal count data. *Statistics in medicine* 37, 4441–4456. [PubMed: 30132947]
- Huffer FW and McKeague IW (1991). Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* 86, 114–129.
- Johnson LM and Strawderman RL (2012). A smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. *Statistics in Medicine* 31, 2335–2358. [PubMed: 22437629]
- Kalbfleisch JD and Prentice RL (2002). *The statistical analysis of failure data* Wiley.
- Lin D and Ying Z (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The annals of Statistics* pages 1712–1734.
- Lin DY and Ying Z (1994). Semiparametric analysis of the additive risk model. *Biometrika* 81, 61–71.
- Liu X and Zeng D (2013). Variable selection in semiparametric transformation models for right-censored data. *Biometrika* 100, 859–876.
- Martinussen T and Scheike TH (2002). A flexible additive multiplicative hazard model. *Biometrika* 89, 283–298.
- Ortega JM (1972). *Numerical analysis: a second course* SIAM.
- Ortega JM and Rheinboldt WC (1970). *Iterative solution of nonlinear equations in several variables* SIAM.
- Pettitt AN (1982). Inference for the linear model using a likelihood based on ranks. *Journal of the Royal Statistical Society: Series B (Methodological)* 44, 234–243.
- Qu L and Sun L (2019). The cox-aalen model for recurrent-event data with a dependent terminal event. *Statistica Neerlandica* 73, 234–255.
- Scheike TH and Zhang M. j. (2002). An additive–multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics* 29, 75–88.
- Scheike TH and Zhang M-J (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics* 59, 1036–1045. [PubMed: 14969483]
- Shen P-S and Weng LN (2018). The Cox–Aalen model for left-truncated and right-censored data. *Communications in Statistics-Theory and Methods* 47, 5357–5368.
- Shen P. s. and Weng LN (2019). The cox-aalen model for left-truncated and mixed interval-censored data. *Statistics* 53, 1152–1167.
- van der Vaart A and Wellner JA (1996). *Weak convergence and empirical processes: with applications to statistics* Springer, New York.
- Yu G, Li Y, Zhu L, Zhao H, Sun J, and Robison LL (2019). An additive– multiplicative mean model for panel count data with dependent observation and dropout processes. *Scandinavian Journal of Statistics* 46, 414–431.
- Zeng D and Lin D (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* 93, 627–640.
- Zeng D and Lin D (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 507–564.
- Zeng D, Mao L, and Lin D (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* 103, 253–271. [PubMed: 27279656]
- Zhou Q, Sun Y, and Gilbert PB (2021). Semiparametric regression analysis of partly interval-censored failure time data with application to an AIDS clinical trial. *Statistics in Medicine* 40, 4376–4394. [PubMed: 34080723]

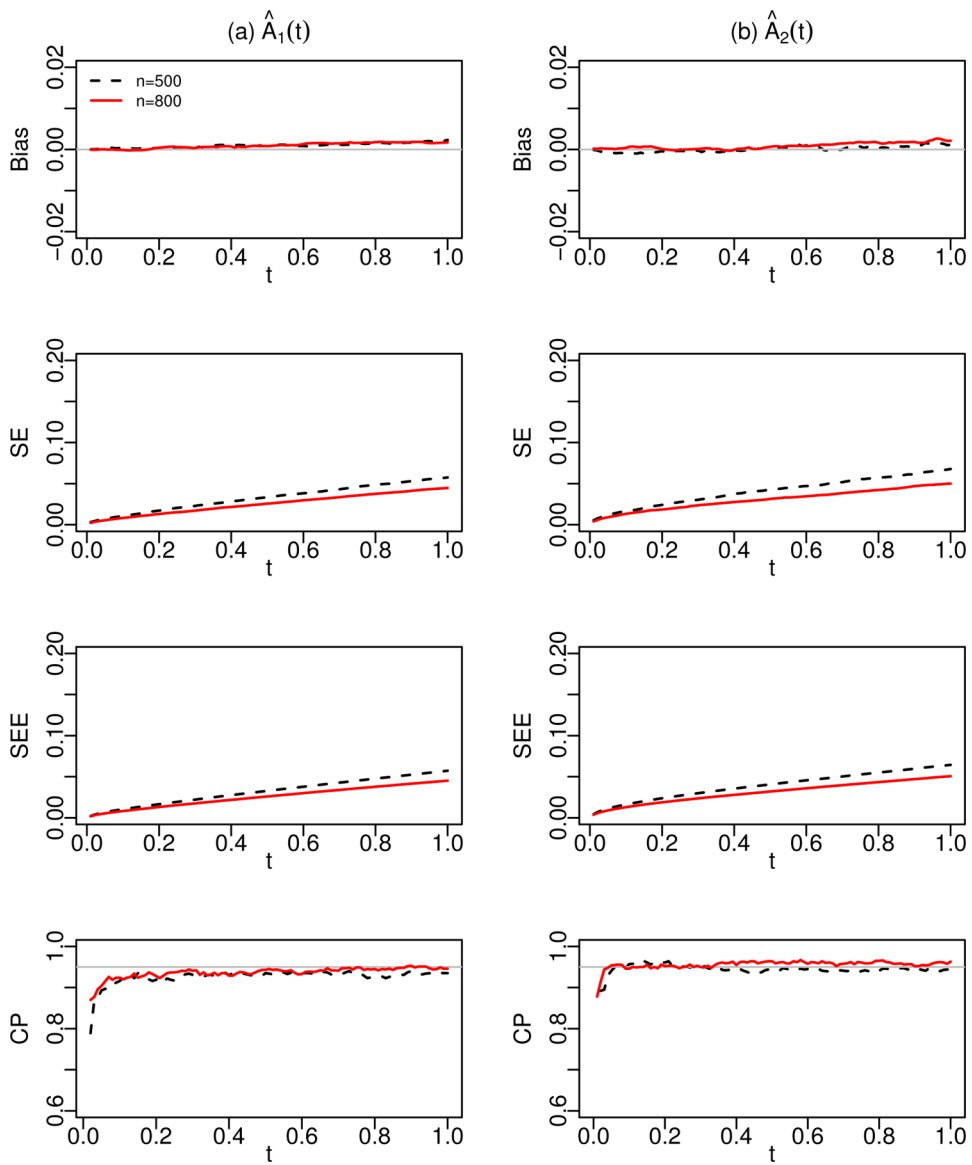


Figure 1. Estimation results for (a) $A_1(t) = \log(1 + t/4)$ and (b) $A_2(t) = 0.1t$ in Scenario 1, under the logarithmic transformation $G(x) = r^{-1} \log(1 + rx)$ with $r = 0$. The dashed and solid lines are for data sets with $n = 500$ and $n = 800$, respectively. Bias, SE, SEE, and CP stand, respectively, for the bias, empirical standard error, mean of the standard error estimator, and empirical coverage probability of the 95% confidence interval.

Kaplan–Meier plot

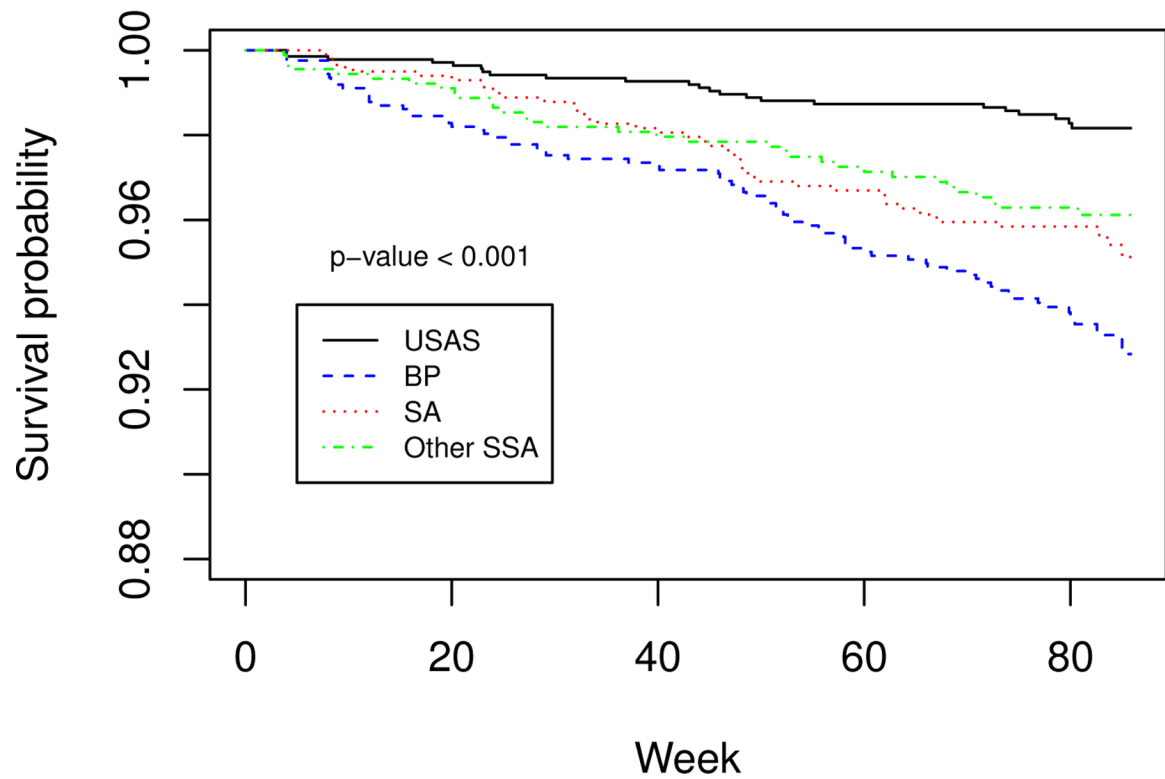


Figure 2. Kaplan–Meier plot for four regions in the full cohort. Here, “USAS”, “BP”, “SA” and “Other SSA” represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

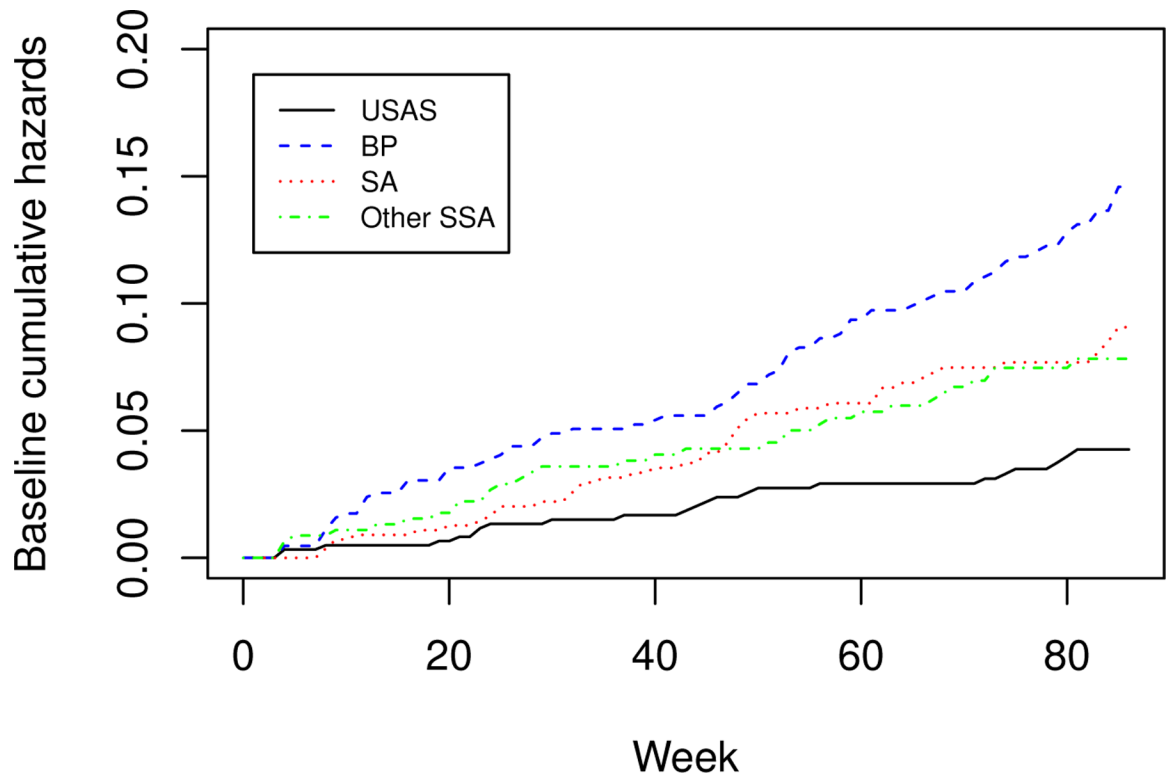


Figure 3. Estimated baseline cumulative hazard function for four regions under the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 0$. Here, “USAS”, “BP”, “SA” and “Other SSA” represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

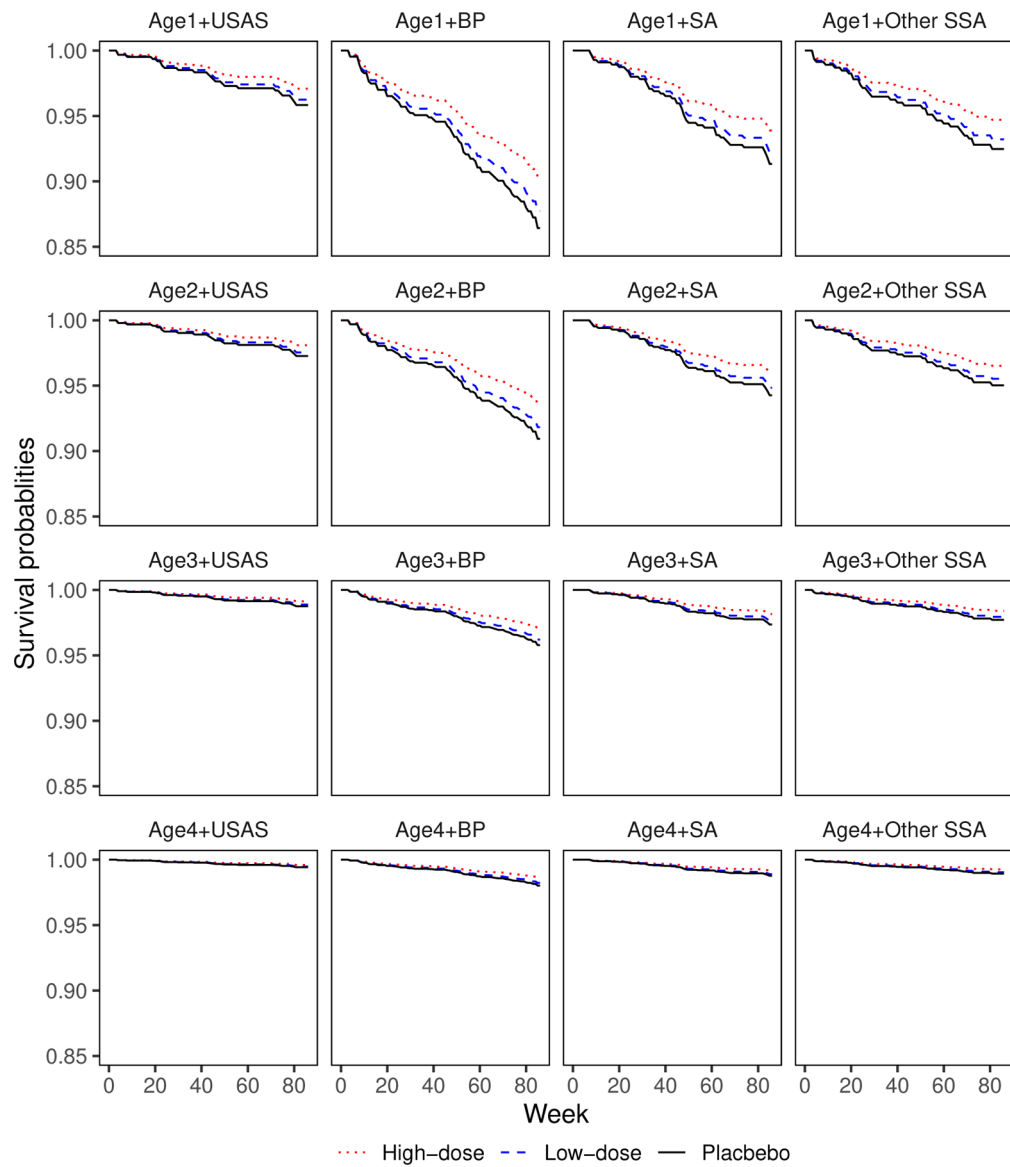


Figure 4. Estimated survival functions by considering different combinations of covariates under the proposed model with the transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 0$. Here, “Age1”, “Age2”, “Age3” and “Age4” stand for the age groups, [17, 20], [21, 30], [31, 40] and [41, 52], respectively. “USAS”, “BP”, “SA” and “Other SSA” represent USA and Switzerland, Brazil and Peru, South Africa and other sub-Saharan African countries, respectively.

Table 1

Simulation results for estimation of the regression parameters under Scenarios 1 to 4

<i>r</i>	<i>n</i>	$\beta_1 = 0.5$				$\beta_2 = -0.5$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
<i>Scenario 1</i>									
0	200	0.003	0.350	0.340	0.950	-0.014	0.587	0.574	0.947
	500	0.007	0.212	0.212	0.952	-0.005	0.354	0.359	0.956
	800	-0.003	0.172	0.167	0.948	-0.001	0.279	0.285	0.951
0.5	200	-0.001	0.380	0.369	0.949	-0.014	0.639	0.630	0.950
	500	0.007	0.227	0.230	0.961	-0.002	0.388	0.395	0.957
	800	-0.004	0.185	0.181	0.951	-0.001	0.302	0.313	0.951
1	200	-0.002	0.402	0.402	0.956	-0.020	0.690	0.699	0.957
	500	0.008	0.246	0.252	0.957	0.002	0.415	0.437	0.968
	800	-0.005	0.198	0.198	0.946	-0.003	0.323	0.347	0.957
<i>Scenario 2</i>									
0	200	0.004	0.337	0.334	0.954	-0.002	0.574	0.563	0.942
	500	0.006	0.208	0.209	0.962	-0.004	0.348	0.353	0.960
	800	-0.002	0.169	0.164	0.945	-0.006	0.275	0.280	0.949
0.5	200	-0.002	0.363	0.362	0.951	-0.014	0.624	0.620	0.948
	500	0.008	0.223	0.227	0.954	-0.000	0.384	0.389	0.963
	800	-0.003	0.184	0.179	0.946	-0.003	0.300	0.308	0.948
1	200	-0.005	0.389	0.396	0.957	-0.019	0.672	0.690	0.958
	500	0.007	0.240	0.248	0.957	0.002	0.412	0.432	0.965
	800	-0.004	0.197	0.195	0.944	-0.005	0.325	0.342	0.961
<i>Scenario 3</i>									
0	200	0.005	0.288	0.287	0.958	-0.009	0.494	0.486	0.947
	500	0.002	0.180	0.180	0.949	-0.004	0.305	0.305	0.953
	800	-0.001	0.146	0.141	0.949	-0.009	0.245	0.241	0.950
0.5	200	0.001	0.319	0.321	0.954	-0.009	0.552	0.555	0.953
	500	0.001	0.198	0.202	0.956	-0.004	0.348	0.348	0.953
	800	-0.001	0.160	0.159	0.947	-0.009	0.276	0.276	0.942
1	200	0.010	0.343	0.365	0.964	-0.002	0.624	0.647	0.962
	500	-0.002	0.216	0.229	0.959	-0.002	0.380	0.403	0.963
	800	-0.006	0.175	0.180	0.963	-0.003	0.305	0.318	0.960
<i>Scenario 4</i>									
0	200	0.010	0.339	0.334	0.951	-0.011	0.572	0.562	0.938
	500	0.005	0.211	0.209	0.948	-0.003	0.348	0.353	0.958
	800	-0.000	0.169	0.164	0.948	-0.000	0.280	0.279	0.948
0.5	200	0.007	0.366	0.362	0.946	-0.014	0.615	0.619	0.947
	500	0.003	0.228	0.227	0.950	-0.005	0.378	0.389	0.968
	800	0.001	0.183	0.179	0.950	-0.001	0.308	0.307	0.949
1	200	0.002	0.389	0.396	0.951	-0.008	0.657	0.690	0.960

<i>r</i>	<i>n</i>	$\beta_1 = 0.5$				$\beta_2 = -0.5$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
	500	0.002	0.244	0.248	0.957	0.002	0.412	0.433	0.967
	800	0.001	0.195	0.196	0.956	0.000	0.330	0.341	0.954

Note: Bias, bias of the parameter estimator; SE, empirical standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, empirical coverage percentage of the 95% confidence interval.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Regression analysis results for the HIV trials from Zeng and Lin’s model fit to each of the four geographic regions separately and the proposed Cox-Aalen transformation model based on the full cohort data with the logarithmic transformation $G(x) = r^{-1}\log(1 + rx)$ with $r = 0$

Covariates	USA/Switzerland			Brazil/Peru		
	Est	SE	<i>p</i> -value	Est	SE	<i>p</i> -value
Low-dose	-0.107	0.484	0.825	-0.167	0.276	0.545
High-dose	-0.437	0.524	0.404	-0.279	0.283	0.325
21 – 30	-0.454	0.630	0.472	-0.525	0.262	0.045
31 – 40	-2.709	1.141	0.018	-1.283	0.396	0.001
41 – 52	-1.152	0.903	0.202	-16.859	1.668	<0.001
Covariates	South Africa			Other SSA		
	Est	SE	<i>p</i> -value	Est	SE	<i>p</i> -value
Low-dose	-0.025	0.354	0.943	-0.080	0.400	0.842
High-dose	-0.392	0.392	0.317	-0.509	0.449	0.258
21 – 30	-0.187	0.380	0.623	-0.480	0.498	0.335
31 – 40	-0.954	0.601	0.112	-0.719	0.586	0.220
41 – 52	-13.867	2.366	<0.001	-13.871	2.626	<0.001
Covariates	The Proposed Model					
	Est	SE	<i>p</i> -value			
Low-dose	-0.108	0.178	0.542			
High-dose	-0.363	0.190	0.056			
21 – 30	-0.429	0.187	0.022			
31 – 40	-1.219	0.274	<0.001			
41 – 52	-1.989	0.721	0.006			

Note: Est and SE stand for the estimates of the regression parameters and the estimated standard errors, respectively. “Other SSA” is for other sub-Saharan African countries. “USA/Switzerland”, “Brazil/Peru”, “South Africa”, and “Other SSA” correspond to the estimation results by fitting Zeng and Lin’s model to each geographic region. “The Proposed Model” corresponds to the estimation results when fitting the proposed Cox-Aalen transformation models to the full cohort data.