Published in final edited form as:

Stat Med. 2017 March 30; 36(7): 1157–1171. doi:10.1002/sim.7204.

An Expectation Maximization algorithm for fitting the generalized odds-rate model to interval censored data

Jie Zhou^a, Jiajia Zhang^{a,*†}, and Wenbin Lu^b

^aDepartment of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, U.S.A

^bDepartment of Statistics, North Carolina State University, Raleigh, NC, U.S.A

Abstract

The generalized odds-rate model is a class of semiparametric regression models, which includes the proportional hazards and proportional odds models as special cases. There are few works on estimation of the generalized odds-rate model with interval censored data because of the challenges in maximizing the complex likelihood function. In this paper, we propose a gamma-Poisson data augmentation approach to develop an Expectation Maximization algorithm, which can be used to fit the generalized odds-rate model to interval censored data. The proposed Expectation Maximization algorithm is easy to implement and is computationally efficient. The performance of the proposed method is evaluated by comprehensive simulation studies and illustrated through applications to datasets from breast cancer and hemophilia studies. In order to make the proposed method easy to use in practice, an R package ‘ICGOR’ was developed.

Keywords

data augmentation; EM algorithm; interval censoring; generalized odds-rate models

1. Introduction

Although the proportional hazards (PH) model [1] is the popularly used semiparametric regression model in survival analysis, the PH assumption may not be satisfied in practice. Alternative models, such as the proportional odds (PO) model [2], may provide better fit than the PH model. More generally, the generalized odds-rate (GOR) model [3] is a class of semiparametric regression models, which includes the PH and PO models as special cases.

Let T denote the event time of interest and X be the associated covariate vector. The GOR model is defined as

*Correspondence to: Jiajia Zhang, Department of Epidemiology and Biostatistics, University of South Carolina 915 Greene Street, Suite 453, Columbia, SC 29208, U.S.A.

†jzhang@mailbox.sc.edu

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.

$$g_r\{S_T(t|X)\} = H(t) + \beta'X, \quad (1)$$

where $S_T(\cdot|X)$ is the survival function of T given X , β is a vector of unknown regression parameters, $H(\cdot)$ is an unknown strictly increasing transformation function with $H(0) = -\infty$, and $g_r(\cdot)$ is a known decreasing link function indexed by a parameter r . Specifically, $g_0(s) = \log(-\log s)$ and $g_r(s) = \log(r^{-1}(s^{-r} - 1))$ when $r > 0$. Therefore, $r = 0$ corresponds to the PH model while $r = 1$ corresponds to the PO model.

The discussions on the GOR model had mainly focused on right censored data [3–6]. To our knowledge, there is few discussion on the estimation of the GOR model when data are interval censored. Interval censored data commonly occurred in medical research, animal carcinogenicity, and epidemiological studies [7, 8], where each subject has a sequence of follow-up visits $0 = V_0 < V_1 < \dots < V_v < V_{v+1} = \infty$ during the study. The true event time of interest T is never observed exactly, but is covered by two consecutive visits, say $V_j < T < V_{j+1}$, for some $j = 0, \dots, v$. The observed time interval is denoted as $(L, R]$. If $L = 0$, the event time is left censored; if $R = \infty$, the event time is right censored. When $v > 1$, it is the general case of type II interval censoring [9]. A special case is type I interval censoring, or sometimes called current status data, where $v = 1$ and the event time T is either left censored or right censored.

Traditional approaches dealing with interval censored data are mainly based on the sieve maximum likelihood [9], which have been applied to both the PH model [10] and the PO model [11–13]. Other estimation methods include the iterative convex minorant algorithm [14] and multiple imputation [15] for the PH model, and the conditional logistic regression approach [16] for the PO model. Sun [17] and Zhang [18] proposed general inference procedures based on estimating equations for the linear transformation model [19]. However, the estimation of the GOR model with interval censored data is less studied because of the difficulties in maximizing the complex likelihood function directly. Existing methods usually require that covariates are discrete in deriving the nonparametric maximum likelihood estimation [20].

Recently, McMahan et al. [21] developed a Poisson data augmentation method for efficient estimation of the PH and PO model with current status data, and Wang et al. [22] extended the approach to the PH model with interval censored data. The data augmentation simplifies the likelihood structure and leads to the fast convergence of the Expectation Maximization (EM) algorithm. Furthermore, the variance estimates can be obtained in closed forms by the Louis method [23]. In this article, we propose a gamma-Poisson data augmentation for the GOR model with interval censored data, and include the Poisson augmentation as a special case. The generalization to the GOR model provides more flexibilities in practice when the PH assumption is not satisfied.

The rest of the article is organized as follows. The proposed gamma-Poisson data augmentation approach is introduced in Section 2. We focus on the estimation with a prespecified r in this paper. The EM algorithm and variance estimation are discussed in

Section 3. Simulation studies and data applications are presented in Sections 5 and 6, respectively. The contributed R package 'ICGOR' is briefly illustrated in Section 7. Some conclusions and discussions are outlined in Section 8. An extended EM algorithm, which includes the estimation of r , is discussed in Section 4, and more details are provided in the supplementary materials.

2. Gamma-Poisson data augmentation

2.1. Notation and likelihood

Based on the GOR model in (1), the survival function of T is given by

$$S_{T(t|X)} = \begin{cases} \exp\{-e^{H(t) + \beta'X}\}, & r = 0, \\ [1 + re^{H(t) + \beta'X}]^{-\frac{1}{r}}, & r > 0. \end{cases}$$

Here, we focus on the discussion of the GOR model with $r > 0$, because the PH model ($r = 0$) is the special case as $r \rightarrow 0$.

Let $\mathbf{O} = (L_i, R_i, \delta_{L,i}, \delta_{R,i}, \delta_{I,i}, X_i, i = 1, 2, \dots, n)$ denote the observed data for n subjects, where $(L_i, R_i]$ is the observed time interval including the exact event time T_i for the i th subject, and $\delta_{L,i}$, $\delta_{R,i}$ and $\delta_{I,i}$ are the censoring indicators, where $\delta_{L,i} = 1$ when T_i is left censored ($L_i = 0$), $\delta_{R,i} = 1$ when T_i is right censored ($R_i = \infty$), and $\delta_{I,i} = 1$ when T_i is interval censored ($0 < L_i < R_i < \infty$). Note that $\delta_{L,i} + \delta_{I,i} + \delta_{R,i} = 1$ for all i .

Assuming non-informative censoring, the observation times L_i and R_i and the event time T_i are independent given the covariates X_i . The observed likelihood function of the parameters β and H_e under the GOR model is given by

$$\mathcal{L}(\beta, H_e) = \quad (2)$$

$$\begin{aligned} & \partial \prod_{i=1}^n \left[1 - \left\{ 1 + r H_e(R_i) e^{\beta' X_i} \right\}^{-\frac{1}{r}} \right]^{\delta_{L,i}} \left[\left\{ 1 + r H_e(L_i) e^{\beta' X_i} \right\}^{-\frac{1}{r}} - \left\{ 1 + r H_e(R_i) e^{\beta' X_i} \right\}^{-\frac{1}{r}} \right]^{\delta_{I,i}} \\ & \times \left\{ 1 + r H_e(L_i) e^{\beta' X_i} \right\}^{-\frac{1}{r} \delta_{R,i}}, \end{aligned}$$

where $H_e(t) = \exp\{H(t)\}$. Direct maximization of (2) with respect to parameters is not reliable because of the complex structure of the likelihood function. To facilitate the maximum likelihood estimation, we propose a gamma-Poisson data augmentation method. The complete data likelihood function is defined in Section 2.2, based on which the EM algorithm is developed.

2.2. Complete data likelihood

Let $\phi = (\phi_1, \dots, \phi_n)$ denote a vector of independent latent variables from $\text{Gamma}(1/r, r)$ distribution. Then, the conditional survival function of the GOR model can be written as

$$S(t|X_i) = \int_0^\infty \exp\left\{-\phi_i H_e(t) e^{\beta' X_i}\right\} \frac{\phi_i^{\frac{1}{r}-1} e^{-\frac{1}{r}\phi_i}}{\Gamma\left(\frac{1}{r}\right) r^{\frac{1}{r}}} d\phi_i$$

which is the marginal survival function of the gamma PH frailty model. Treating ϕ as missing data, the complete data likelihood function is

$$\mathcal{L}(\beta, H_e | \phi) \quad (3)$$

$$\begin{aligned} &= \prod_{i=1}^n \left[1 - \exp\left\{-\phi_i H_e(R_i) e^{\beta' X_i}\right\} \right]^{\delta_{L,i}} \left[\exp\left\{-\phi_i H_e(L_i) e^{\beta' X_i}\right\} - \exp\left\{-\phi_i H_e(R_i) e^{\beta' X_i}\right\} \right]^{\delta_{I,i}} \\ &\times \exp\left\{-\delta_{R,i} \phi_i H_e(L_i) e^{\beta' X_i}\right\} \frac{\phi_i^{\frac{1}{r}-1} e^{-\frac{1}{r}\phi_i}}{\Gamma\left(\frac{1}{r}\right) r^{\frac{1}{r}}}. \end{aligned}$$

To handle the interval censoring in the aforementioned complete data likelihood, we introduce Poisson latent variables conditional on the gamma frailty term ϕ . To be specific, for subject i , two latent variables are introduced: $Y_i | \phi_i \sim \text{Pois}(\lambda_i \phi_i)$ and $W_i | \phi_i \sim \text{Pois}(\omega_i \phi_i)$, where $\text{Pois}(v)$ denotes the Poisson distribution with mean v ,

$\lambda_i = e^{\beta' X_i} \left\{ \delta_{L,i} H_e(R_i) + \delta_{I,i} H_e(L_i) \right\}$ and $\omega_i = e^{\beta' X_i} \left\{ H_e(R_i) - H_e(L_i) \right\} + \delta_{R,i} H_e(L_i)$. Moreover,

we set $Y_i \equiv 0$ if $Y_i \sim \text{Pois}(0)$. Note that Y_i and W_i are independent of each other conditional on ϕ_i .

If T_i is left censored, we have $W_i \equiv 0$ since $\omega_i = 0$. Therefore,

$$P(Y_i > 0, W_i = 0 | \phi_i, \delta_{L,i} = 1) = P(Y_i > 0 | \phi_i, \delta_{L,i} = 1) = 1 - e^{-\lambda_i \phi_i} = 1 - \exp\left\{-\phi_i H_e(R_i) e^{\beta' X_i}\right\}.$$

Similarly, if T_i is interval censored, because of the conditional independence of Y_i and W_i , we have

$$\begin{aligned}
 P(Y_i = 0, W_i > 0 | \phi_i, \delta_{L,i} = 1) &= P(Y_i = 0 | \phi_i, \delta_{L,i} = 1) \times P(W_i > 0 | \phi_i, \delta_{L,i} = 1) \\
 &= e^{-\lambda_i \phi_i} (1 - e^{-\omega_i \phi_i}) = \exp \left\{ -\phi_i H_e(L_i) e^{\beta' X_i} \right\} - \exp \left\{ -\phi_i H_e(R_i) e^{\beta' X_i} \right\}.
 \end{aligned}$$

Finally, if T_i is right censored, we have $Y_i \equiv 0$ since $\lambda_i = 0$. Then,

$$P(Y_i = 0, W_i = 0 | \phi_i, \delta_{R,i} = 1) = P(W_i = 0 | \phi_i, \delta_{R,i} = 1) = e^{-\omega_i \phi_i} = \exp \left\{ -\phi_i H_e(L_i) e^{\beta' X_i} \right\}.$$

Therefore, treating $\mathbf{Y} \equiv (Y_1, \dots, Y_n)$ and $\mathbf{W} \equiv (W_1, \dots, W_n)$ as missing data, the complete data likelihood function (3) can be further augmented as

$$\begin{aligned}
 \mathcal{L}_c(\boldsymbol{\beta}, H_e | \boldsymbol{\phi}, \mathbf{Y}, \mathbf{W}) &= \prod_{i=1}^n \left\{ \frac{1}{Y_i!} (\lambda_i \phi_i)^{Y_i} e^{-\lambda_i \phi_i} \right\}^{\delta_{L,i}} \times \left\{ \frac{1}{W_i!} (\omega_i \phi_i)^{W_i} e^{-(\lambda_i + \omega_i) \phi_i} \right\}^{\delta_{L,i}} \quad (4) \\
 &\times e^{-\omega_i \phi_i \delta_{R,i}} \times \frac{\phi_i^{\frac{1}{r}-1} e^{\frac{1}{r} \phi_i}}{\Gamma(\frac{1}{r}) r^{\frac{1}{r}}}
 \end{aligned}$$

with the constraints: $Y_i > 0$ if $\delta_{L,i} = 1$ and $W_i > 0$ if $\delta_{L,i} = 1$, for $i = 1, \dots, n$. It can be shown that after integrating out the latent variables $\boldsymbol{\phi}$, $\mathbf{Y}$, and $\mathbf{W}$, the complete data likelihood function (4) reduces to the observed likelihood function (2).

To estimate the unknown monotone transformation function H_e , we use monotone splines:

$$H_e(t) = \sum_{l=1}^k \gamma_l b_l(t),$$

where b_l 's are nondecreasing integrated spline basis functions ranging from 0 to 1, and the coefficients $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ are nonnegative to ensure the monotonicity. As suggested by Ramsay [24], the choice of knots for the spline functions can be either equally spaced or at the quartiles. In practice, information criteria like AIC or BIC can be used to select the number of knots. In our simulation setting, it shows that the monotone splines have stable performance and different choices of knots have very limited impact on the coefficient estimation. To be consistent with [21] and [22], we use five equally spaced knots at percentiles in the simulation studies. The AIC criteria is used in number of knots selection in real data analysis.

To further simplify the calculation, the second layer of Poisson latent variables are introduced as follows: $Y_{il} | \phi_i \sim \text{Pois}(\lambda_{il} \phi_i)$ and $W_{il} | \phi_i \sim \text{Pois}(\omega_{il} \phi_i)$, $l = 1, \dots, k$, where

$\lambda_{il} = \gamma_l e^{\beta' X_i} \left\{ \delta_{L,i} b_l(R_i) + \delta_{L,i} b_l(L_i) \right\}$ and $\omega_{il} = \gamma_l e^{\beta' X_i} \left\{ \delta_{L,i} [b_l(R_i) - b_l(L_i)] + \delta_{R,i} b_l(L_i) \right\}$. Note that $\sum_{l=1}^k \lambda_{il} = \lambda_i$ and $\sum_{l=1}^k \omega_{il} = \omega_i$. Moreover, we have $Y_i \stackrel{d}{=} \sum_{l=1}^k Y_{il}$ and $W_i \stackrel{d}{=} \sum_{l=1}^k W_{il}$, where “ $\stackrel{d}{=}$ ” means equal in distribution. Define $\mathbf{Y}_I = (Y_{1I}, \dots, Y_{nI})$ and $\mathbf{W}_I = (W_{1I}, \dots, W_{nI})$, $I = 1, \dots, k$. Treating ϕ , $\mathbf{Y}_I$'s and $\mathbf{W}_I$'s as missing data, the complete data likelihood function can be written as

$$\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\gamma}, r) \equiv \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, r | \phi, \mathbf{Y}_I, \mathbf{W}_I, l = 1 \dots k) = \prod_{i=1}^n \frac{\phi_i^{\frac{1}{r} - 1} e^{-\frac{1}{r} \phi_i}}{\Gamma\left(\frac{1}{r}\right) r^{\frac{1}{r}}} \quad (5)$$

$$\left\{ \prod_{l=1}^k \left[\frac{1}{Y_{il}!} (\lambda_{il} \phi_i)^{Y_{il}} e^{-\lambda_{il} \phi_i} \right]^{\delta_{L,i}} \left[\frac{1}{W_{il}!} (\omega_{il} \phi_i)^{W_{il}} e^{-(\lambda_{il} + \omega_{il}) \phi_i} \right]^{\delta_{L,i}} e^{-\delta_{R,i} \omega_{il} \phi_i} \right\}$$

with the constraints: $\sum_{l=1}^k Y_{il} > 0$ if $\delta_{L,i} = 1$ and $\sum_{l=1}^k W_{il} > 0$ if $\delta_{L,i} = 1$, for $i = 1, \dots, n$.

Similarly, it can be shown that after integrating out ϕ , $\mathbf{Y}_I$'s and $\mathbf{W}_I$'s, the complete data likelihood function (5) reduces to the observed likelihood function (2).

In order to maximize the observed likelihood function with respect to the unknown parameters, an EM algorithm can be derived based on the complete data likelihood function (5). The transformation parameter r is treated as a prespecified nonnegative number in most of the literature studies [25–28]. We develop the EM algorithm with a prespecified r in Section 3, and more general case with unknown r is discussed in Section 4.

3. Expectation Maximization algorithm with prespecified r

Let us assume that r is a prespecified nonnegative value and define $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$. The expectation of the logarithm of complete data likelihood function (5) with respect to the latent variables ϕ , $\mathbf{Y}_I$'s and $\mathbf{W}_I$'s and conditional on the observed data $\mathbf{O}$ and current parameter estimate $\boldsymbol{\theta}^{(d)}$ can be expressed as

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(d)}) &= E[\log \left\{ \mathcal{L}_c(\boldsymbol{\theta}) \right\} | \mathbf{O}, \boldsymbol{\theta}^{(d)}] \\ &= \sum_{i=1}^n \sum_{l=1}^k (\log \gamma_l + \beta' X_i) \left\{ \delta_{L,i} E(Y_{il} | \mathbf{O}, \boldsymbol{\theta}^{(d)}) + \delta_{L,i} E(W_{il} | \mathbf{O}, \boldsymbol{\theta}^{(d)}) \right\} \\ &\quad - \gamma_l e^{\beta' X_i} E(\phi_i | \mathbf{O}, \boldsymbol{\theta}^{(d)}) \left\{ (1 - \delta_{R,i}) b_l(R_i) + \delta_{R,i} b_l(L_i) \right\} + L(\boldsymbol{\theta}^{(d)}), \end{aligned} \quad (6)$$

where $L(\boldsymbol{\theta}^{(d)})$ is a function of $\boldsymbol{\theta}^{(d)}$ but free of $\boldsymbol{\theta}$. Once the conditional expectations are calculated, $\boldsymbol{\theta}$ can be updated by maximizing $\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(d)})$. The maximization can be achieved by finding the root of its first derivative. Notice that there is a closed-form solution to

$$\frac{\partial \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})}{\partial \gamma_l} = 0,$$

which is the first partial derivative of $\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(d)})$ with respect to γ_l , $l = 1, \dots, k$. The solutions of the γ_l 's can be written as a function of $\boldsymbol{\beta}$:

$$\gamma_l(\boldsymbol{\beta}) = \frac{\sum_{i=1}^n \delta_{L,i} E(Y_{il} | \boldsymbol{O}, \boldsymbol{\theta}^{(d)}) + \delta_{L,i} E(W_{il} | \boldsymbol{O}, \boldsymbol{\theta}^{(d)})}{\sum_{i=1}^n e^{\boldsymbol{\beta}' X_i} E(\phi_i | \boldsymbol{O}, \boldsymbol{\theta}^{(d)}) \left\{ (1 - \delta_{R,i}) b_l(R_i) + \delta_{R,i} b_l(L_i) \right\}}, \quad l = 1, \dots, k. \quad (7)$$

Plugging-in the $\gamma_l(\boldsymbol{\beta})$'s back into (6), the function $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ becomes a function of $\boldsymbol{\beta}$, which can be maximized directly based on Newton-type algorithms. Once $\boldsymbol{\beta}$ is updated as $\hat{\boldsymbol{\beta}}^{(d+1)}$, γ_l can be updated by plugging in $\hat{\boldsymbol{\beta}}^{(d+1)}$ into the function $\gamma_l(\boldsymbol{\beta})$ for $l = 1, \dots, k$.

3.1. Conditional expectations

When subject i is left censored, Y_i conditionally follows the zero-truncated negative binomial distribution (ZTNB), which is denoted as $Y_i \sim \text{ZTNB}(1/r, \delta_{L,i} r \lambda_i^{(d)} / (1 + r \lambda_i^{(d)}))$. The probability mass function can be written as

$$Pr(Y_i = y | \boldsymbol{O}, \boldsymbol{\theta}^{(d)}) = \frac{\delta_{L,i}}{c_i^{(d)}} \binom{y + \frac{1}{r} - 1}{y} \left(\frac{r \lambda_i^{(d)}}{1 + r \lambda_i^{(d)}} \right)^y \left(\frac{1}{1 + r \lambda_i^{(d)}} \right)^{\frac{1}{r}} I(y > 0),$$

where

$$c_i^{(d)} = 1 - \left(\frac{\delta_{L,i}}{1 + r \lambda_i^{(d)}} \right)^{\frac{1}{r}},$$

and $\lambda_i^{(d)}$ is the value of μ_i evaluated at $\boldsymbol{\theta}^{(d)}$. Similarly, when the subject is interval censored, we have $W_i \sim \text{ZTNB}(1/r, \delta_{L,i} r \omega_i^{(d)} / (1 + r \lambda_i^{(d)} + r \omega_i^{(d)}))$. The probability mass function of W_i is

$$Pr(W_i = w | \boldsymbol{O}, \boldsymbol{\theta}^{(d)}) = \frac{\delta_{L,i}}{d_i^{(d)}} \binom{w + \frac{1}{r} - 1}{w} \left(\frac{r \omega_i^{(d)}}{1 + r \lambda_i^{(d)} + r \omega_i^{(d)}} \right)^w \left(\frac{1 + r \lambda_i^{(d)}}{1 + r \lambda_i^{(d)} + r \omega_i^{(d)}} \right)^{\frac{1}{r}} I(w > 0),$$

where

$$d_i^{(d)} = 1 - \left(\frac{\delta_{L,i}(1 + r\lambda_i^{(d)})}{1 + r\lambda_i^{(d)} + r\omega_i^{(d)}} \right)^{\frac{1}{r}},$$

and $\omega_i^{(d)}$ is the value of ω_i evaluated at $\theta^{(d)}$.

The conditional expectations of Y_i and W_i are

$$E(Y_i | \mathbf{O}, \theta^{(d)}) = \frac{\delta_{L,i}\lambda_i^{(d)}}{c_i^{(d)}}, \text{ and } E(W_i | \mathbf{O}, \theta^{(d)}) = \frac{\delta_{L,i}\omega_i^{(d)}}{(1 + r\lambda_i^{(d)})d_i^{(d)}}.$$

Based on the fact that $Y_{il} | Y_i \sim \text{Binomial}(Y_i, \lambda_{il}^{(d)} / \lambda_i^{(d)})$, and $W_{il} | W_i \sim \text{Binomial}(W_i, \omega_{il}^{(d)} / \omega_i^{(d)})$, and applying the iterated rule of expectations, the conditional expectations of Y_{il} and W_{il} are

$$E(Y_{il} | \mathbf{O}, \theta^{(d)}) = \frac{\lambda_{il}^{(d)}}{\lambda_i^{(d)}} E(Y_i | \mathbf{O}, \theta^{(d)}) = \frac{\delta_{L,i}\lambda_{il}^{(d)}}{c_i^{(d)}}, \text{ and } \quad (8)$$

$$E(W_{il} | \mathbf{O}, \theta^{(d)}) = \frac{\omega_{il}^{(d)}}{\omega_i^{(d)}} E(W_i | \mathbf{O}, \theta^{(d)}) = \frac{\delta_{L,i}\omega_{il}^{(d)}}{(1 + r\lambda_i^{(d)})d_i^{(d)}}. \quad (9)$$

Similarly, $\lambda_{il}^{(d)}$ ($\omega_{il}^{(d)}$) is the value of λ_{il} (ω_{il}) evaluated at $\theta^{(d)}$. The conditional expectation of ϕ_i can be derived from the joint distributions of Y_i , W_i and ϕ_i , which is

$$E(\phi_i | \mathbf{O}, \theta^{(d)}) = \frac{\delta_{L,i} \{1 - (1 - c_i^{(d)})^{r+1}\}}{c_i^{(d)}} + \frac{\delta_{L,i} \{1 - (1 - d_i^{(d)})^{r+1}\}}{(1 - \lambda_i^{(d)})d_i^{(d)}} + \frac{\delta_{R,i}}{1 + r\omega_i^{(d)}}. \quad (10)$$

Remark 1—When $r \rightarrow 0$, ϕ_i converges in probability to the value of 1, and the limiting conditional distributions of Y_i and W_i become zero-truncated Poisson, which lead to the same results as in Wang et al. [22] for the PH model with interval censored data.

3.2. Algorithm

The EM algorithm is described as follows:

Step 0: For a pre-specified value of r , give initial values $\theta^{(0)}$ as $\beta^{(0)} = \mathbf{0}$ and $\gamma^{(0)} = \mathbf{1}$.

Step 1: In the $(d+1)$ th iteration, calculate conditional expectations in equations from (8) to (10) based on $\theta^{(d)}$.

Step 2: Update $\beta^{(d+1)}$ by maximizing $Q(\theta; \theta^{(d)})$ with γ being replaced by $\gamma(\beta) = (\gamma_1(\beta), \dots, \gamma_k(\beta))$, where $\gamma_l(\beta)$'s, $l = 1, \dots, k$ are defined in Equation (7).

Step 3: Calculate $\gamma_l^{(d+1)} = \gamma_l(\beta^{(d+1)})$ for $l = 1, \dots, k$.

Step 4: Repeat Step 1 to Step 3 until the absolute difference of log-likelihood values between two consecutive iterations is less than 0.001. The estimates $\hat{\theta} = (\hat{\beta}, \hat{\gamma})$ are obtained in the last iteration.

Remark 2—Benefit from the data augmentation procedure, there exists a closed-form expression for γ in Equation (7), and it makes the object function $Q(\theta; \theta^{(d)})$ depend on β only. Thus, it greatly simplifies the maximization task in Step 2. The maximization can be easily realized using Newton-type algorithms.

Remark 3—The convergence criterion based on log-likelihood values is commonly applied for models involving monotone splines [21, 22]. Some other parameter-based convergence criteria [29] are not as efficient because of the slow convergence in some of the γ estimates.

3.3. Variance estimation

The variance of $\hat{\theta}$ can be obtained by the Louis method [23]. That is,

$$\text{Var}(\hat{\theta}) = (-\hat{\mathcal{H}})^{-1}.$$

where

$$\hat{\mathcal{H}} = \frac{\partial^2 Q(\theta; \hat{\theta})}{\partial \theta \partial \theta'} + \text{Var} \left(\frac{\partial \log L_c(\theta)}{\partial \theta} \right) \Big|_{\theta = \hat{\theta}}.$$

All quantities in the aforementioned hessian matrix have closed forms and are calculated in the Appendix. The Wald-type confidence intervals for θ can be obtained in the regular way.

4. Expectation Maximization algorithm with unknown r

So far, we have treated the parameter r in the link function $g_r(\cdot)$ as known in deriving the EM algorithm. In fact, r is identifiable in the model [25] and can be estimated the same way as other parameters. The extension is straightforward, and the details of the algorithm can be found in the supplementary materials, where we refer to as the ‘extended EM algorithm’.

However, the extended EM algorithm has several disadvantages for practical use. First, it converges to the maximum likelihood estimators (MLEs) very slowly. This is because the likelihood function is very flat around the true value of r , and we need to change the convergence tolerance to 1e-6 to make the estimated r close to the truth. This problem is also mentioned in [25], where they estimate the promotion time cure model with survival probability being specified similarly.

Second, correct estimation of r requires large sample size. Zeng *et al.* [25] claimed a sample size of more than 1500 to correctly estimate the transformation parameter. We conduct a simulation study for this extended EM algorithm with both 200 and 1500 sample sizes. The results are provided in Table S1 in the supplementary materials. We estimate $\log(r)$ instead to avoid the restriction of $r > 0$ and apply the delta method to obtain estimates of r . The bias of $\log(r)$ is large under 200 sample size and coverage probabilities of r and $\log(r)$ are below the nominal level. The performance improves when the sample size increases to 1500. The histograms of estimated r are provided in Figure S1 in the supplementary materials. The empirical distribution of $\hat{r}$ is not symmetric with small sample size, which may explain the low coverage probability.

Third, the closed-form expressions for the Hessian matrix are no longer available because the distribution of $\log(\phi)$ does not have a closed form. We apply the numerical approximation of Hessian matrix in the simulation to obtain the standard error of the parameters. This can be computationally expensive when the sample size is large.

Considering practical needs, we suggest not to estimate r directly but to choose r based on the log-likelihood values, which is commonly applied in the literature for the transformation model [25–27]. The details of this approach and a sensitivity analysis are discussed in Section 5.2.

5. Simulation studies

5.1. Parameter estimation with fixed value of r

We generate data from the GOR model

$$S(t|X) = \left\{ 1 + r H_e(t) e^{\beta' X} \right\}^{-\frac{1}{r}}.$$

For $H_e(t)$, we consider both a concave function $H_e(t) = \log(1 + t) + t^{3/2}$ and a convex function $H_e(t) = 0.5 \times t^{5/6}$. A two-dimensional covariate $X = (X_1, X_2)'$ is considered, where $X_1 \sim \text{Uniform}(0, 2)$ and $X_2 \sim \text{Bernoulli}(0.5)$. The coefficients are set to be $\beta = (1, -1)$. Different models with $r = 0.5, 1, 2$ are considered.

For the i th subject, the ordered visiting times are $0 = V_{i0} < V_{i1} < \dots < V_{iv_i}$, where v is the total number of visits. The lengths between two consecutive visits are $\tau_i^{(1)}, \dots, \tau_i^{(v_i)}$, where $\tau_i^{(j)} = V_{ij} - V_{i(j-1)}$ for $j = 1, \dots, v_i$. Here, $\tau_i^{(j)}$'s are generated independently from an exponential distribution with rate of 2. We consider two censoring scenarios. In the first scenario, we have 10% left censoring and 20% right censoring. In the second scenario, we increase the left and right censoring to 20% and 50%, respectively. The baseline visit time V_{i0} is generated from a Uniform distribution $U(0, \kappa)$. The number of visits of each subject v_i is generated from the Poisson distribution with mean ζ . The constants κ and ζ are chosen to have the designed censoring rates accordingly.

In general, the number of knots can be selected using the Akaike information criterion (AIC). Based on our experience in the simulation, the performance of the monotone splines is quite stable and the resulted parameter estimates for β and H_e are very close with different choices of k . To stay in consistent with the choice in Wang et al.[22], we use cubic splines with $k = 5$ equally spaced knots at percentiles to represent the monotone transformation function H_e in the simulation studies. Sample size of 200 and the true values of r are used for each setting. The proposed gamma-Poisson augmentation-based EM estimator is compared with the numerical MLE of the likelihood function (3) with H_e being specified by the monotone splines with the same set of knots and order. The optimization methods proposed by Nelder and Mead [30] is used to obtain the numerical MLE based on the likelihood function (3).

We conduct 1000 replications for each setting, and report the bias, average estimated standard deviation (StErr), empirical standard deviation (StDev) and empirical coverage probability (CP) of 95% Wald-type confidence intervals. From the results in Table I, the Numerical MLE method produces biased estimates in some settings and the CPs are below the nominal level of 0.95. The bias increases as the censoring proportions increase from Scenario 1 to Scenario 2. This could be because the numerical method, which directly maximizes the likelihood function, gets stuck at local extrema. The gamma-Poisson data augmentation simplifies the likelihood function and the resulted EM algorithm gives unbiased estimates, comparable StErrs and StDevs, and reasonable CPs that are close to the nominal level of 0.95.

We further investigate the estimated baseline survival functions of the proposed EM algorithm. For illustration, we present the estimated baseline survival curves along with 2.5% and 97.5% quantiles under Scenario 2 in Figure 1. The mean of the estimated survival curves are close to the truth in all the cases. The proposed EM algorithm greatly reduces the computational time. The convergence time under each setting is summarized in Table II. On the average, the proposed EM algorithm reduces more than half of the convergence time compared with the numerical MLE method.

5.2. Selecting r based on log-likelihood values

In practice, it is also of interest to select a model with the best fit from a class of candidate models. A commonly adopted approach is to compare the log-likelihood values of these candidate models and choose the one with the largest log-likelihood value [25–27]. Specifically, consider a set of grids $0 < r_1 < \dots < r_M$. For each r_m , $m = 1, \dots, M$, we run the EM algorithm to obtain the maximum likelihood estimates of θ and calculate the associated observed likelihood values at convergence. We choose r_m with the largest likelihood value.

Here, we study the model selection performance of the proposed method by considering three models: PH model ($r = 0$), PO model ($r = 1$), and a variant of PO model ($r = 2$). We use the same simulation setting in previous Section 5.1. The transformation function of $H_e(t) = \log(1 + t) + t^{3/2}$ is assumed and cubic splines with five equally spaced knots are used for approximating $H_e(t)$. The true parameter r takes values of 0, 1, or 2. Left censoring proportion is set at 10%, and two right censoring scenarios of 10% and 50% are considered.

We study the performance under $n = 200, 500$, and 1000 , corresponding to small, moderate, and large sample sizes, respectively.

One thousand replications are conducted for each setting. For each simulated data, we fit the GOR model with $r = 0, 1$, or 2 . The value of r that gives the largest estimated observed likelihood value is chosen and the frequencies of choosing each r value over 1000 runs are reported in Table III. From the table, we see that the frequencies of selecting the true r are much larger than those of selecting the wrong model for most of cases. When the true model is the PH model, we have high frequencies of selecting the true model. Even when the sample size is small and right censoring proportion is large, the accuracy is above 80%. With moderate to large sample sizes, the frequencies of selecting the true model are above 90%. It is difficult to detect a PO model when sample size is small and right censoring proportion is large. But the accuracy increases as the sample size increases.

6. Real data analysis

6.1. Breast cancer study

Beadle *et al.* [31] reported a retrospective study to compare the cosmetic effects of radiotherapy alone versus radiotherapy and adjuvant chemotherapy on 94 women with early breast cancer. The event of interest was the time to first appearance of moderate or severe breast retraction. There are 46 patients in radiation only group and 48 patients in radiation plus chemotherapy group. Patients were observed initially every 46 months, but, as their recovery progressed, the interval between visits lengthened. Due to the fact that patients were observed only at these random times, the exact time of breast retraction is known to fall in the interval between visits. As a result, 5.3% of the women were left censored, 55.8% were interval censored and 38.9% were right censored. The dataset is available in the R package `KMsurv`.

We fit the data using the proposed method. The AIC criterion suggests that $k = 1$ knots for the monotone splines is enough for estimating the transformation function H_c . To find the best parameter r in the link function, we search for r ranges from 0 to 2 with step 0.1 by comparing the resulted log-likelihood values in Figure 2a. It is observed that the log-likelihood values decrease as r increases. Therefore, a PH model is appropriate for this data. The estimated coefficient comparing the radiotherapy and adjuvant chemotherapy group with the radiotherapy alone group is 0.897 with standard error 0.287. In comparison, our estimates are consistent with the results 0.890 (0.297) provided in [32] and 0.878 (0.294) in [33]. Based on the estimates, the hazard ratio comparing the radiotherapy+chemotherapy group and the radiotherapy alone group is about 2.452 with 95% confidence interval (1.397, 4.304). The estimated survival curves for the two study groups are plotted in Figure 2b, which are comparable with the Turnbull nonparametric estimates.

6.2. Hemophilia study

The hemophilia study data [34] were collected as a part of a multi-center prospective study aimed at assessing the HIV-1 infection rate among people with hemophilia. In particular, the individuals in the study were at risk from contracting HIV-1 from blood products made from

donor plasma. In this study, 544 patients were categorized into one of four groups based on their average annual dose of blood products. Specifically, there are 236 patients have none dose of blood products, 132 patients in the low dose group, 102 patients in the medium dose group, and 74 patients in the high dose group. The goal of the study was to compare the HIV-1 infection rate between these dose groups. The time at which patients contracted HIV-1 is not known exactly, but were known to occur before, between, or after certain observation times and the left, interval and right censoring proportions are 11.6%, 37.5%, and 50.9%, respectively. The dataset is available in the `ICSurv` package.

The AIC criterion suggests to use five equally spaced knots at percentiles. The parameter r is chosen among a sequence values from 0 to 2 with step 0.1 based on the log-likelihood values. The estimated observed log-likelihood values are plotted in Figure 3a. Based on the plot, we suggest to fit a PO model to the data. The estimates of the parameters and their standard errors are reported in Table IV, which indicates the higher the dose level, the higher odds to being contracted HIV-1. The estimated survival curves for each group are plotted in Figure 3b, which are comparable with the Turnbull nonparametric estimates.

7. Software

We develop the R package named ‘`ICGOR`’ based on the proposed method and contribute to CRAN for public use. The main function is `GORH()`, where the formula is specified based on the `Surv()` function with the left (right) interval end points of the left (right) censoring observations being set as NA. The transformation parameter r can be specified as any positive numbers. For example, the following code fit the Hemophilia data with $r = 1$.

```
data(Hemophilia)

Hemophilia$L[Hemophilia$d1==1]<-Hemophilia$R[Hemophilia
$d3==1]<-NA

fit <- GORH (Surv (L, R) ~ Low + Medium + High, data =
Hemophilia, r = 1)
```

A summary table for the coefficients, including the estimates, standard errors, and test results, as well as the log-likelihood value can be obtained by `summary(fit)`. The `predict` function can be used to obtain the estimated survival curve of a new individual, and the covariates are specified by `new.x`.

```
pred<-predict(fit, new.x = c(0, 0, 0))

plot(pred)
```

8. Discussion

We proposed a gamma-Poisson data augmentation-based EM algorithm for deriving the maximum likelihood estimate of the GOR model to interval censored data. The proposed gamma-Poisson data augmentation facilitates the convergence of the EM algorithm. Moreover, the variance of the EM estimates have closed forms and can be obtained using the Louis method. The selection of the parameter r based on the loglikelihood values are studies

in simulation and applied in real data analysis. The comprehensive simulation demonstrates that the proposed method greatly improves the accuracy of estimates and shortens the computational time.

The total number of knots does not need to be large in practice, we suggest to use 5–8 equally spaced knots at percentiles for the splines. The order of 3 (cubic splines) is flexible enough in practice. The AIC and BIC criteria may be used in knots selection. The proposed method works pretty well with moderate to large sample size and moderate right censoring rate as shown in Section 5. When the sample size is small, say $N < 100$, or the left or right censoring is extremely large, say $> 80\%$, the algorithm sometimes does not converge, and the estimated hessian matrix might be singular.

When the sample size is large, we can include r as a parameter in the model. The asymptotic distribution of r is normal, and inference of r can be made based on the Wald-type confidence interval. However, to develop a hypothesis testing procedure for $H_0 : r = 0$ versus $H_1 : r > 0$ is challenging. Because $r = 0$ is a boundary value of the range, and the asymptotic distribution of the likelihood ratio test statistic is no longer the χ_1^2 distribution. Chernoff [35] proved the asymptotic distribution of the likelihood ratio test statistic at the boundary value is a mixture of chi-squared distributions $0.5\chi_0^2 + 0.5\chi_1^2$ under parametric models. Numerical studies are conducted and the distribution of the test statistics is found to be close to the $0.5\chi_0^2 + 0.5\chi_1^2$ distribution. However, more theoretical work is needed to investigate the distribution of the likelihood ratio test statistic under the GOR model in the future.

The GOR model belongs to a broad family of linear transformation models, which include different types of transformations such as the logarithmic transformation and the Box-Cox transformation [26]. The GOR model corresponds to the logarithmic transformation, which is the Laplace transform of a gamma frailty. The proposed gamma-Poisson data augmentation can hardly be extended to the general linear transformation model that includes other types of transformations such as the Box-Cox transformation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix: Calculation of H

The first part of the hessian matrix $\mathcal{H}$ involves the second derivatives of $Q(\theta; \hat{\theta})$ with respect to θ , which include the following quantities. All the expectations of latent variables are conditioned on the observed data $\mathbf{O}$ and the final estimate $\hat{\theta}$, the conditions are omitted here.

$$\begin{aligned}\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})}{\partial \gamma_l \partial \gamma_{l'}} &= - \sum_{i=1}^n \frac{1}{\hat{\gamma}_l^2} [\delta_{L,i} E(Y_{il}) + \delta_{L,i} E(W_{il})] I(l' = l) \\ \frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})}{\partial \beta_j \partial \beta_{j'}} &= - \sum_{i=1}^n X_{ij} X_{ij'} e^{\hat{\beta}' X_i E(\phi_i)} [(1 - \delta_{R,i}) H_e(R_i) + \delta_{R,i} H_e(L_i)] \\ \frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})}{\partial \gamma_l \partial \beta_{j'}} &= - \sum_{i=1}^n X_{ij'} e^{\hat{\beta}' X_i E(\phi_i)} [(1 - \delta_{R,i}) b_l(R_i) + \delta_{R,i} b_l(L_i)]\end{aligned}$$

The second part include covariance between the first derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$, all of the quantities are listed next, where all the variance and covariance of latent variables are conditioned on the observed data $\boldsymbol{O}$ and the final estimate $\hat{\boldsymbol{\theta}}$.

$$\begin{aligned}\text{Cov}\left[\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \gamma_{l'}}\right] &= \sum_{i=1}^n \frac{1}{\hat{\gamma}_l \hat{\gamma}_{l'}} [\delta_{L,i} \text{Cov}(Y_{il}, Y_{il'}) + \delta_{L,i} \text{Cov}(W_{il}, W_{il'})] - e^{\hat{\beta}' X_i} \left\{ \frac{1}{\hat{\gamma}_l} b_{l'}(R_i) [\delta_{L,i} \text{Cov}(Y_{il}, \phi_i) + \delta_{L,i} \text{Cov}(W_{il}, \phi_i)] + \hat{\gamma}_{l'} b_l(R_i) [\delta_{L,i} \text{Cov}(Y_{il'}, \phi_i) + \delta_{L,i} \text{Cov}(W_{il'}, \phi_i)] \right\} + e^{2\hat{\beta}' X_i} \text{Var}(\phi_i) [(1 - \delta_{R,i}) b_l(R_i) b_{l'}(R_i) + \delta_{R,i} b_l(L_i) b_{l'}(L_i)] \\ \text{Cov}\left[\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \beta_j}, \frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \beta_{j'}}\right] &= \sum_{i=1}^n X_{ij} X_{ij'} \left\{ [\delta_{L,i} \text{Var}(Y_i) + \delta_{L,i} \text{Var}(W_i)] - 2e^{\hat{\beta}' X_i} H_e(R_i) [\delta_{L,i} \text{Cov}(Y_{il}, \phi_i) + \delta_{L,i} \text{Cov}(W_{il}, \phi_i)] + e^{2\hat{\beta}' X_i} \text{Var}(\phi_i) [(1 - \delta_{R,i}) H_e^2(R_i) + \delta_{R,i} H_e^2(L_i)] \right\} \\ \text{Cov}\left[\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \beta_j}\right] &= \sum_{i=1}^n \frac{X_{ij}}{\hat{\gamma}_l} \left\{ \delta_{L,i} \text{Cov}(Y_{il}, Y_i) + \delta_{L,i} \text{Cov}(W_{il}, W_i) - e^{\hat{\beta}' X_i} H_e(R_i) [\delta_{L,i} \text{Cov}(Y_{il}, \phi_i) + \delta_{L,i} \text{Cov}(W_{il}, \phi_i)] - X_{ij'} e^{\hat{\beta}' X_i} b_l(R_i) [\delta_{L,i} \text{Cov}(Y_i, \phi_i) + \delta_{L,i} \text{Cov}(W_i, \phi_i)] + X_{ij} e^{2\hat{\beta}' X_i} \text{Var}(\phi_i) [(1 - \delta_{R,i}) b_l(R_i) H_e(R_i) + \delta_{R,i} b_l(L_i) H_e(L_i)] \right\}\end{aligned}$$

Next, we calculate the conditional variance and covariance of the latent variables in the previous equations. Since we know $Y_i \sim \text{ZTNB}(1/r, \delta_{L,i} \hat{r} \hat{\lambda}_i / (1 + r \hat{\lambda}_i))$ and

$W_i \sim \text{ZTNB}(1/r, \hat{r} \hat{\omega}_i / (1 + r \hat{\lambda}_i + r \hat{\omega}_i))$, the conditional variances can be calculated directly as

$$\begin{aligned}\text{Var}(Y_i) &= \frac{\delta_{L,i} \hat{\lambda}_i}{\hat{c}_i^2} [(1 + r) \hat{c}_i \hat{\lambda}_i + \hat{c}_i - \hat{\lambda}_i], \\ \text{Var}(W_i) &= \frac{\delta_{L,i} \hat{\omega}_i}{(1 + r \hat{\lambda}_i)^2 \hat{d}_i^2} [(1 + r) \hat{d}_i \hat{\omega}_i + \hat{d}_i (1 + r \hat{\lambda}_i) - \hat{\omega}_i].\end{aligned}$$

The conditional variance of ϕ_i is

$$\begin{aligned} \text{Var}(\phi_i) = & \frac{\delta_{L,i}}{\hat{c}_i^2} \left\{ \hat{c}_i(r+1)[1 - (1 - \hat{c}_i)^{2r+1}] - [1 - (1 - \hat{c}_i)^{r+1}]^2 \right\} + \frac{\delta_{L,i}}{\hat{d}_i^2(1 + r\hat{\lambda}_i)^2} \left\{ \hat{d}_i(r+1)[1 - (1 - \hat{d}_i)^{2r+1}] \right. \\ & \left. - [1 - (1 - \hat{d}_i)^{r+1}]^2 \right\} + \frac{r\delta_{R,i}}{(1 + r\hat{\omega}_i)^2}. \end{aligned}$$

The joint pdf of $(Y_{i1}, \dots, Y_{ik}, \phi_i)$ is

$$f(Y_{i1}, \dots, Y_{ik}, \phi_i) = \frac{\delta_{L,i} I(\sum_{l=1}^k Y_{il} > 0)}{c_i \Gamma\left(\frac{1}{r}\right) r^{\frac{1}{r}}} \phi_i^{\sum_{l=1}^k Y_{il} + \frac{1}{r} - 1} e^{-\phi_i(\frac{1}{r} + \lambda_i)} \prod_{l=1}^k \frac{1}{Y_{il}!} \lambda_i^{Y_{il}},$$

based on which, we have the conditional covariance between Y_{il} and ϕ_i is

$$\text{Cov}(Y_{il}, \phi_i) = \frac{\delta_{L,i} \hat{\lambda}_{il}}{\hat{c}_i^2} [\hat{c}_i(r+1) + (1 - \hat{c}_i)^{r+1} - 1].$$

Furthermore, using the fact that $Y_i \stackrel{d}{=} \sum_{l=1}^k Y_{il}$, we have

$$\text{Cov}(Y_i, \phi_i) = \sum_{l=1}^k \text{Cov}(Y_{il}, \phi_i) = \frac{\delta_{L,i} \hat{\lambda}_i}{\hat{c}_i^2} [\hat{c}_i(r+1) + (1 - \hat{c}_i)^{r+1} - 1].$$

Similarly, the joint pdf of $(W_{i1}, \dots, W_{ik}, \phi_i)$ is

$$f(W_{i1}, \dots, W_{ik}, \phi_i) = \frac{\delta_{L,i} (1 + r\hat{\lambda}_i)^{\frac{1}{r}} I(\sum_{l=1}^k W_{il} > 0)}{d \Gamma\left(\frac{1}{r}\right) r^{\frac{1}{r}}} \phi_i^{\sum_{l=1}^k W_{il} + \frac{1}{r} - 1} e^{-\phi_i(\frac{1}{r} + \lambda_i + \omega_i)} \prod_{l=1}^k \frac{1}{W_{il}!} \omega_i^{W_{il}},$$

and the conditional covariance between W_{il} and ϕ_i is

$$\text{Cov}(W_{il}, \phi_i) = \frac{\delta_{L,i} \hat{\omega}_{il}}{\hat{d}_i^2(1 + r\hat{\lambda}_i)^2} [\hat{d}_i(r+1) + (1 - \hat{d}_i)^{r+1} - 1].$$

We also have

$$\text{Cov}(W_i, \phi_i) = \sum_{k=1}^k \text{Cov}(W_{il}, \phi_i) = \frac{\delta_{L,i} \hat{\omega}_i}{\hat{d}_i^2(1 + r\hat{\lambda}_i)^2} [\hat{d}_i(r+1) + (1 - \hat{d}_i)^{r+1} - 1].$$

The conditional variance and covariance of $Y_{jl}(W_{jl})$ can be obtained by using the iterated rules for the variance/covariance and recognizing that the conditional distribution of $Y_{jl}(W_{jl})$ given $Y_j(W_j)$ is binomial.

References

1. Cox, DR. Breakthroughs in Statistics. Springer; New York: 1992. Regression models and life-tables; p. 527-541.
2. Bennett S. Analysis of survival data by the proportional odds model. *Statistics in Medicine*. 1983; 2(2):273–277. [PubMed: 6648142]
3. Dabrowska DM, Doksum KA. Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*. 1988; 83(403):744–749.
4. Bickel P. Efficient testing in a class of transformation models. *Proceedings of the 45th Session of the International Statistical Institute*. 1986; 23:63–23.
5. Scharfstein DO, Tsiatis AA, Gilbert PB. Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime data analysis*. 1998; 4(4):355–391. [PubMed: 9880995]
6. Banerjee T, Chen MH, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Analysis*. 2007; 13(2):241–260. [PubMed: 17401683]
7. Finkelstein DM, Wolfe RA. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*. 1985; 41(4):933–945. [PubMed: 3830259]
8. Hoel DG, Walburg H. Statistical analysis of survival experiments. *Journal of the National Cancer Institute*. 1972; 49(2):361–372. [PubMed: 5076821]
9. Sun, J. *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer Science & Business Media; New York: 2007.
10. Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics*. 1986; 42(4):845–854. [PubMed: 3814726]
11. Murphy S, Rossini A, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*. 1997; 92(439):968–976.
12. Huang J, Rossini A. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*. 1997; 92(439):960–967.
13. Shen X. Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*. 1998; 85(1):165–177.
14. Pan W. Extending the iterative convex minorant algorithm to the cox model for interval-censored data. *Journal of Computational and Graphical Statistics*. 1999; 8(1):109–120.
15. Pan W. A multiple imputation approach to cox regression with interval-censored data. *Biometrics*. 2000; 56(1):199–203. [PubMed: 10783796]
16. Rabinowitz D, Betensky RA, Tsiatis AA. Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*. 2000; 56(2):511–518. [PubMed: 10877311]
17. Sun J, Sun L. Semiparametric linear transformation models for current status data. *Canadian Journal of Statistics*. 2005; 33(1):85–96.
18. Zhang Z, Sun L, Zhao X, Sun J. Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*. 2005; 33(1):61–70.
19. Cheng S, Wei L, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82(4):835–845.
20. Zhang B, Tong X, Zhang J, Wang C, Sun J. Efficient estimation for linear transformation models with current status data. *Communications in Statistics-Theory and Methods*. 2013; 42(17):3191–3203.
21. McMahan CS, Wang L, Tebbs JM. Regression analysis for current status data using the em algorithm. *Statistics in Medicine*. 2013; 32(25):4452–4466. [PubMed: 23761135]

22. Wang L, McMahan CS, Hudgens MG, Qureshi ZP. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*. 2016; 72(1):222–231. [PubMed: 26393917]
23. Louis TA. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*. 1982; 44(2):226–233.
24. Ramsay JO. Monotone regression splines in action. *Statistical Science*. 1988; 3(4):425–441.
25. Zeng D, Yin G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*. 2006; 101(474):670–684.
26. Zeng D, Lin D. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006; 93(3):627–640.
27. Zeng D, Lin D. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(4):507–564.
28. Mao M, Wang JL. Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association*. 2010; 105(489):302–311. [PubMed: 22865944]
29. McLachlan, G., Krishnan, T. *The EM Algorithm and Extensions*. John Wiley & Sons; New Jersey: 2007. New Jersey
30. Nelder JA, Mead R. A simplex method for function minimization. *The Computer Journal*. 1965; 7(4):308–313.
31. Beadle GF, Come S, Henderson IC, Silver B, Hellman S, Harris JR. The effect of adjuvant chemotherapy on the cosmetic results after primary radiation treatment for early stage breast cancer. *International Journal of Radiation Oncology* Biology* Physics*. 1984; 10(11):2131–2137.
32. Satten GA. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*. 1996; 83(2):355–370.
33. Satten GA, Datta S, Williamson JM. Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the American Statistical Association*. 1998; 93(441):318–327.
34. McMahan, CS., Wang, L. *ICsurv: A Package for Semiparametric Regression Analysis of Interval-Censored Data*. 2014. <http://CRAN.R-project.org/package=ICsurv>, r package version 1.0. Release date: 6/9/2014
35. Chernoff H. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*. 1954; 25(3):573–578.

Figure 1.

Estimated baseline survival curves using the proposed method. Solid lines are the truth, dashed lines are the mean of the estimated survival curves, and dotted lines are the 2.5% and 97.5% quantiles of the estimated survival curves. (a) $r = 0.5, H_e(t) = \log(1+t) + t^{3/2}$; (b) $r = 1, H_e(t) = \log(1+t) + t^{3/2}$; (c) $r = 2, H_e(t) = \log(1+t) + t^{3/2}$; (d) $r = 0.5, H_e(t) = 0.5 \times t^{5/6}$; (e) $r = 1, H_e(t) = 0.5 \times t^{5/6}$; and (f) $r = 2, H_e(t) = 0.5 \times t^{5/6}$.

Figure 2. Breast cancer study. (a) Choose r based on grid search and (b) estimated survival curves.

Figure 3. Hemophilia study. (a) Choose r based on grid search and (b) estimated survival curves.

Table 1

Simulation results.

Numerical MLE										EM	
	r	par.	Bias	StDev	StErr	CP	Bias	StDev	StErr	CP	
$H_0(t) = 0.5 \times t^{0.6}$											
Scenario 1	0.5	β_1	-0.11	0.22	0.19	0.86	0.02	0.20	0.20	0.95	
		β_2	-0.15	0.25	0.22	0.84	-0.01	0.22	0.22	0.96	
	1	β_1	-0.09	0.25	0.23	0.87	0.02	0.23	0.23	0.95	
		β_2	-0.14	0.28	0.27	0.89	-0.02	0.27	0.26	0.95	
	2	β_1	0.15	0.30	0.29	0.89	0.03	0.30	0.30	0.96	
		β_2	-0.03	0.34	0.33	0.94	-0.02	0.33	0.34	0.95	
Scenario 2	0.5	β_1	-0.20	0.27	0.23	0.79	0.02	0.23	0.25	0.96	
		β_2	-0.20	0.31	0.26	0.82	-0.02	0.26	0.27	0.96	
	1	β_1	-0.21	0.29	0.25	0.80	0.03	0.25	0.28	0.96	
		β_2	-0.20	0.34	0.29	0.84	-0.02	0.30	0.30	0.95	
	2	β_1	-0.25	0.34	0.29	0.79	0.02	0.30	0.33	0.96	
		β_2	-0.23	0.40	0.34	0.85	-0.03	0.37	0.36	0.95	
$H_0(t) = \log(1+t) + t^{0.2}$											
Scenario 1	0.5	β_1	0.04	0.24	0.22	0.93	0.02	0.23	0.23	0.94	
		β_2	0.01	0.25	0.25	0.95	-0.02	0.25	0.25	0.95	
	1	β_1	0.01	0.29	0.28	0.94	0.01	0.25	0.26	0.96	
		β_2	-0.02	0.32	0.31	0.94	-0.01	0.28	0.29	0.96	
	2	β_1	0.03	0.32	0.32	0.94	0.00	0.31	0.32	0.94	
		β_2	-0.01	0.39	0.37	0.94	-0.01	0.36	0.36	0.95	
Scenario 2	0.5	β_1	0.02	0.26	0.26	0.94	0.04	0.25	0.27	0.95	
		β_2	-0.04	0.30	0.29	0.93	-0.03	0.29	0.29	0.95	
	1	β_1	0.07	0.25	0.24	0.93	0.02	0.28	0.30	0.95	
		β_2	0.06	0.28	0.28	0.94	-0.03	0.32	0.32	0.95	
	2	β_1	0.19	0.31	0.30	0.90	0.02	0.33	0.35	0.93	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

r	par.	Numerical MLE					EM		
		Bias	StDev	StErr	CP	Bias	StDev	StErr	CP
	β_2	0.14	0.35	0.34	0.91	-0.02	0.38	0.38	0.94

Table II

Summary of convergence time (in seconds).

		Numerical MLE						EM			
r		Mean	StDev	Median	Q1	Q3	Mean	StDev	Median	Q1	Q3
$H_{\hat{\theta}}(\theta) = 0.5 \times \hat{\theta}^6$											
Scenario 1	0.5	17.60	0.61	17.52	17.16	18.01	3.10	1.08	2.99	2.37	3.80
	1	17.56	0.66	17.48	17.07	18.04	2.85	1.05	2.72	2.06	3.55
	2	17.26	0.63	17.25	16.78	17.71	2.43	0.95	2.26	1.73	2.90
Scenario 2	0.5	13.67	0.85	13.52	12.98	14.27	2.89	0.98	2.83	2.25	3.50
	1	13.65	0.93	13.39	12.87	14.47	2.89	0.92	2.85	2.25	3.48
	2	13.48	0.81	13.40	12.85	14.08	3.01	1.01	2.95	2.25	3.63
$H_{\hat{\theta}}(\theta) = \log(1 + \theta) + \hat{\theta}^{1/2}$											
Scenario 1	0.5	15.73	1.42	15.08	14.48	16.93	6.37	3.07	5.83	4.32	7.56
	1	15.79	1.20	15.50	14.78	16.57	5.72	2.12	5.45	4.07	7.09
	2	16.79	0.66	16.75	16.32	17.23	5.38	1.86	5.10	4.06	6.52
Scenario 2	0.5	14.45	0.85	14.19	13.76	15.24	5.68	3.09	4.98	3.90	6.60
	1	13.87	0.68	13.77	13.39	14.20	5.12	2.41	4.64	3.67	5.86
	2	14.27	0.92	13.97	13.56	15.11	5.05	1.82	4.81	3.83	5.99

Table III

Model selection frequency.

n	True <i>r</i>	10% Right censor			50% Right censor		
		0	1	2	0	1	2
200	0	94.9%	5.1%	0.0%	83.2%	14.6%	2.2%
	1	13.3%	59.0%	27.7%	30.6%	33.3%	36.1%
	2	1.6%	30.8%	67.6%	14.5%	21.9%	63.6%
500	0	99.6%	0.4%	0.0%	91.8%	8.1%	0.1%
	1	2.8%	84.5%	12.7%	20.5%	54.9%	24.6%
	2	0.0%	23.2%	76.8%	4.4%	26.5%	69.1%
1000	0	99.9%	0.1%	0.0%	98.0%	2.0%	0.0%
	1	0.4%	91.6%	8.0%	9.8%	74.8%	15.4%
	2	0.0%	16.2%	83.8%	0.6%	21.4%	78%

Table IV

Proportional odds model for hemophilia study.

Variable	Estimate	Standard error
Low versus None	2.268	0.160
Medium versus None	4.102	0.237
High versus None	5.034	0.305