Aust. N. Z. J. Stat. 2014 doi: 10.1111/j.1467-842X.XXX

Joint modeling of survival and longitudinal ordered data using a semiparametric approach

KEMMAWADEE PREEDALIKIT[1], IVY LIU[2,*], YUICHI HIROSE[2], NOKUTHABA SIBANDA[2], AND DANIEL FERNÁNDEZ[3]

University of Phayao and Victoria University of Wellington, New Zealand

Summary

Medical research frequently focuses on the relationship between quality of life and survival time of subjects. Quality of life may be one of the most important factors that could be used to predict survival, making it worth identifying factors that jointly affect survival and quality of life. We propose a semiparametric joint model that consists of item response and survival components, where these two components are linked through latent variables. Several popular ordinal models are considered and compared in the item response component, while the Cox proportional hazards model is used in the survival component. We estimate the baseline hazard function and model parameters simultaneously, through a profile likelihood approach. We illustrate the method using an example from a clinical study.

Key words: Joint model; Longitudinal study; Ordinal responses; Profile likelihood; Semiparametric model; Stereotype model; Survival

1. Introduction

In clinical studies, quality of life (QOL) and the length of survival of patients are often the main points of interest. Two interventions can have very similar survival outcomes but substantially different effects on a patient's quality of life. One can analyze the quality of life and survival data using two separate models. Alternatively, one can jointly analyse both of them if the two data components are likely to be related. Although methods for separate analyses of the two data components are well established, joint analysis of the two endpoints is of more recent provenance although it is being increasingly developed (Henderson 2005). There are potential gains in power to be made by considering the two endpoints simultaneously.

Survival time can refer to time to death and can also include other common events of interest, such as time to recurrence of symptoms or time to infection with a disease. Staccato is a multicenter clinical trial carried out by a research collaboration undertaken by Switzerland, Australia and Thailand from January 2002 until November 2005 (Ananworanich & the Staccato study group 2006). It is a randomized trial comparing continuous anti-retroviral treatment to CD4-guided-interruption treatment for patients with Human Immunodeficiency Virus (HIV) infection. Patients' CD4 count (a measure of immune recovery) and viral load were assessed every four weeks for an initial screening period of 12 weeks. Recruitment

* Author to whom correspondence should be addressed.
[1] School of Science, University of Phayao, Thailand.
[2] School of Mathematics and Statistics, Victoria University of Wellington, Wellington, 6140, New Zealand
e-mail: Ivy.Liu@vuw.ac.nz

Acknowledgment. We are grateful to J. Ananworanich, the HIV Netherlands Australia Thailand Research Collaboration (HIV-NAT) and the Staccato Study Group for providing the data set analysed in this paper. We also thank reviewers for their useful comments.

Prepared using anzsauth.cls *[Version: 2014/01/06 v1.01]*

into the trial occurred only if over the screening period, CD4 count remained above 350 cells per μL, HIV viral load remained below 50 copies per mL, and there was no evidence of pre-existing drug resistance. Patients were then randomised to one of the two treatment groups. One third of patients received anti-retroviral treatment continuously, and the rest received treatment whenever their CD4 count dropped to 350 cells per μL or less. Patients were classified as being symptomatic of HIV if they manifested any symptoms listed in the classification system for HIV infection as specified by the U.S. Center for Disease Control (CDC) (Castro *et al.* 1993). Time taken for symptoms to progress to the different levels of the classification system, CDC-B and CDC-C, was recorded. CDC-B consists of symptomatic conditions in an HIV-infected adolescent or adult, including for example bacillary angiomatosis, thrush, herpes zoster (shingles), and listeriosis, that are not included among conditions listed in clinical Category C. On the other hand CDC-C includes the clinical conditions indicative of AIDS, such as cervical cancer (invasive), cytomegalovirus retinitis (with loss of vision), and lymphoma. Patients are placed in only one of the CDC classes and remain in Category C when a condition in the CDC-C class occurs (Castro *et al.* 1993) regardless of whether the CDC-B symptoms occur.

For this study, we consider two approaches. Firstly, we consider progression to CDC-B as the only survival endpoint, with patients in category CDC-C (five out of 354) right-censored. This is reasonable for a single survival endpoint approach since patients did not necessarily experience CDC-B conditions before a condition in CDC-C occurred. Secondly, we use a competing risk approach with two mutually exclusive survival endpoints: progression to CDC-B or CDC-C. In both approaches, patients with no progression time were right-censored at the latest follow-up time recorded, and the impact of treatment on time to symptom progression was the main focus. Patients were also asked to complete the Medical Outcome Study HIV Health Survey (MOS-HIV) questionnaire (Wu *et al.* 1991, 1997) to evaluate their quality of life (QOL) every 24 weeks beginning at the baseline visit and at weeks 24, 48, 72, 96, 120 and 144 of follow-up. There are 32 questions on the MOS-HIV questionnaire and each question has five response categories, corresponding to the following ordinal rank: poor(1), fair(2), good(3), very good(4) and excellent(5). Instead of transforming the ordinal categorical scale into a continuous measure, we use an ordinal model directly to model the QOL. This paper proposes a method to jointly model longitudinal quality of life and time to progress to CDC-B or CDC-C.

The joint model not only allows investigation into both the quality of life and survival components, but also incorporates all information simultaneously with valid and efficient inference. Tsiatis & Davidian (2004) reviewed various methods for analysing joint models. Hogan & Laird (1997), Xu & Zeger (2001) and Hsu *et al.* (2006) among others used the joint model to make more efficient inference on the survival model by incorporating the longitudinal data as auxiliary information. Recently, Wu *et al.* (2012) conducted a simulation study on joint modeling of longitudinal and survival data, and demonstrated that Bayesian methods have a similar performance to likelihood methods. Qiu, Stein & Elston (2013) developed a discrete survival model incorporating longitudinal measurements. Furthermore, Tsiatis & Davidian (2001), Song, Davidian & Tsiatis (2002) and Song & Wang (2008) proposed analysis of the joint model from a semi-parametric perspective, in the sense that no parametric density function is assumed for random effects. Wulfsohn & Tsiatis (1997) and Hsieh, Tseng & Wang (2006) developed joint models with no parametric assumptions on the baseline hazard function in the Cox model (Cox 1972) and used the method of nonparametric

MLE to estimate the baseline hazard function. All of the foregoing techniques model the survival component jointly with continuous longitudinal responses. This paper focuses on modelling the survival component jointly with ordered categorical responses that should not be treated as continuous.

In the past there has been little research into joint models for categorical longitudinal outcomes (Rizopoulos 2012, pp.142). In the last few years, such joint models have gradually received more attention in the literature. For ordered categorical responses, Wang, Douglas & Anderson (2002) used a proportional odds model (McCullagh 1980) and the Cox model to jointly model ordinal and survival data, using a partial likelihood approach. Li *et al.* (2010) proposed a likelihood approach to model a partial proportional odds model (Peterson & Harrell 1990) and a cause-specific hazards model. He & Luo (2013) developed a joint model combining a multilevel item response model with a Cox model using a Bayesian approach. Njagi *et al.* (2013) studied the joint analysis of time-to-event data with various types of responses, including continuous, binary and count. However there is lack of literature on the comparison of various ordinal response models in this area.

Liu & Agresti (2005) reviewed many ordinal response models. Currently, a proportional odds model is the most popular. Such a model is parsimonious, with no need to assign scores to the response categories. However the model entails making a strong assumption about the odds ratios and consequently may be inadequate for some data. Alternatively, a partial proportional odds model does not involve the strong assumption, but contains more parameters than the proportional odds model, especially when there are many response categories. In this paper, we propose the use of an ordered stereotype model (Anderson 1984), which is more general than the proportional odds model. Unlike the partial proportional odds model, the ordered stereotype model still achieves the same level of model parsimony as does the proportional odds model. The survival component is jointly modelled with the ordered stereotype model. We also discuss and compare use of the proportional odds and partial proportional odds models in place of the ordered stereotype model for the item response component. In this paper the survival component is based on a Cox proportional hazards model, which treats the baseline hazard, $h_0(t)$, as an unspecified discrete function. The Cox model is a product of this unspecified baseline hazard and an exponential term involving covariates.

The ordinal responses from the questionnaire are sets of repeated observations over time and over different questions for each subject. There are several methods for dealing with repeated ordered categorical data (Liu & Agresti 2005). One approach adds subject-level terms to the model. Because they are unobserved, they are often considered to be random effects. An alternative approach, explored in this paper, allows the subject level terms to come from a finite mixture with R components or groups. However, group memberships are latent variables and unknown. Similar ideas of grouping have been explored by Titterington, Smith & Makov (1985), Woodruff & Hanson (1996), and Pickles & Croudace (2010). Note that the latent variables are a link between the ordinal response model and the Cox proportional hazards model. We use the EM algorithm to estimate the parameters of the finite mixture. In order to estimate all parameters and the baseline hazards simultaneously, we propose a method combining use of the EM algorithm and profile likelihood.

The outline of the remainder of this paper is as follows. Section 2 describes the structure of the joint model. Section 3 presents an EM algorithm for estimating the model parameters. The baseline hazard is iteratively estimated together with other parameters in the EM steps.

In section 4 we apply the method to the data from the Staccato study. A simulation study in section 5 is used to evaluate the performance of parameter estimators in the survival model when a wrong ordinal response model is fitted. The paper concludes with a discussion in section 6.

2. Model Structure

This section presents the full likelihood function for the three options which we consider for an ordinal response model linked by a latent variable to the Cox proportional hazards model.

2.1. Ordinal Response Models

Let Y_{ijm} be the ordered categorical response from 1 (poor) to L (excellent) on item (or question) j for subject i at the m^{th} protocol-specified time point, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, J$ and $m = 1, 2, \ldots, M$. In total, there are J items in the questionnaire, collected at times $t_1, t_2, \ldots, t_M$.

A stereotype model

Given that subject i belongs to group r, a stereotype model can be written as

$$\log\left[\frac{\Pr(Y_{ijm} = \ell \mid \theta_r)}{\Pr(Y_{ijm} = 1 \mid \theta_r)}\right] = a_\ell + \phi_\ell(b_j + \theta_r), \quad r = 1, \ldots, R, \tag{1}$$

where a_ℓ is a response level intercept parameter with $\ell = 2, \ldots, L$, b_j is an item effect, and θ_r is associated with the discrete latent variable. To make the model identifiable we impose the constraints $a_1 = 0$, $b_1 = 0$, $\phi_1 = 0$ and $\theta_1 = 0$. The parameter θ_r can be referred to as a group effect of the quality of life for patients in group r. However, the group memberships are unknown. The $\{\phi_\ell\}$ parameters can be regarded as unknown scores for the outcome categories. Because $\phi_\ell(b_j + \theta_r) = (A\phi_\ell((b_j + \theta_r)/A))$ for any constant $A \neq 0$, for identifiability, we need to impose a constraint on ϕ_ℓ. To retain an ordinal structure of Y_{ijm}, ϕ_ℓ should increase monotonically as ℓ increases. Consequently we impose the constraint $0 = \phi_1 \leq \phi_2 \leq \ldots \leq \phi_L = 1$ on the model. This stereotype model gives a way of estimating how close adjacent response categories ℓ and $\ell + 1$ are, based on how close ϕ_ℓ and $\phi_{\ell+1}$ are. The model implies that when b_j or θ_r increases, the probability of a response being in category ℓ relative to that of being in category $\ell = 1$ increases. Given r, the odds of response ℓ rather than 1 for item j are $\exp(\phi_\ell b_j)$ times the odds of response ℓ rather than 1 for item 1.

A proportional odds model

Given that subject i belongs to group r, a proportional odds model has the form

$$\log\left[\frac{\Pr(Y_{ijm} \leq \ell \mid \theta_r)}{1 - \Pr(Y_{ijm} \leq \ell \mid \theta_r)}\right] = a_\ell - b_j - \theta_r, \tag{2}$$

where $\ell = 1, \ldots, L-1$, $j = 1, \ldots, J$, and $r = 1, \ldots, R$ with $a_1 < a_2 < \ldots < a_{L-1}$, $b_1 = 0$, and $\theta_1 = 0$. The minus signs make the sign of parameters b_j or θ_r have the usual interpretation, that is, for a larger value of b_j or θ_r, the probability of responding in a lower

category, relative to a higher category is reduced. Therefore a larger value of b_j or θ_r implies that response in a higher category is more likely. One can motivate the model by thinking in terms of an underlying continuous response (Anderson & Philips 1981). The ordinal response is obtained by dividing the continuous response into L categories using cutpoints $\{a_\ell\}$. If the model holds, it does not matter how many ordered categories the response variable has. The effects $\{b_j\}$ and $\{\theta_r\}$ are unchanged.

The descriptions of b_j and θ_r are different from those that are appropriate under stereotype model in the sense that the odds ratios are given with respect to a cumulative probability. For each binary collapsing response ($\leq \ell, > \ell$), Model (2) is simply a logit model. This fact implies that the values of b_j and θ_r are the same for all $L-1$ logit models for all possible collapsings. This is called *the proportional odds assumption*. Thus, given r, the odds of obtaining a response $\geq \ell$ for item j are $\exp(b_j)$ times the odds of obtaining a response $\geq \ell$ for item 1.

A partial proportional odds model

When the proportional odds structure is not appropriate, a partial proportional odds model allows separate effects for each logit model for some predictors. For example, the following model allows item effects to be different for different response levels:

$$\log\left[\frac{\Pr(Y_{ijm} \leq \ell \mid \theta_r)}{1-\Pr(Y_{ijm} \leq \ell \mid \theta_r)}\right] = a_\ell - b_{j\ell} - \theta_r, \quad \ell = 1, \ldots, L-1. \tag{3}$$

Model (3) uses $(L-1)\times(J-1)$ parameters to describe item effects, $\{b_{j\ell}\}$, and is therefore not as parsimonious as Model (2), which has $J-1$ parameters in $\{b_j\}$. In comparison, there are $(L-2)+(J-1)$ parameters $\{\phi_\ell,\ b_j\}$ in the stereotype model (1), which is more parsimonious than Model (3) but less parsimonious than Model (2). Moreover the stereotype model is more general than the proportional odds model, because it does not assume the proportional odds structure and uses the data to estimate the distances between response categories through ϕ_ℓ.

For each of Models, if we replace $\{\theta_r, r = 1, \ldots, R\}$ by random effects $\{\theta_i, i = 1, \ldots, n\}$, the model becomes a random effects model for repeated measurements. The random effects in the model have subject-specific interpretations (Agresti 2013, Chapter 13).

Let $y_{ijm\ell}$ be the binary outcome ("Yes" and "No") on the response category ℓ for the i^{th} subject, j^{th} item and m^{th} time point. If the response is on level ℓ, then $y_{ijm\ell} = 1$, otherwise $y_{ijm\ell} = 0$. Under *the local independence assumption*, given that the latent variable θ_r and the responses $\{Y_{ijm}\}$ are treated as independent, the response probability of the i^{th} individual

becomes $\Pr(\boldsymbol{Y}_i \mid \theta_r, \boldsymbol{\alpha}) = \prod_{m=1}^{M_i} \prod_{j=1}^{J} \Pr(Y_{ijm} \mid \theta_r, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\phi})$ and

$$\Pr(Y_{ijm} \mid \theta_r, \boldsymbol{\alpha}) =$$

$$\prod_{\ell=1}^{L} \left(\frac{\exp(a_\ell + \phi_\ell(b_j + \theta_r))}{1 + \sum_{k=2}^{L} \exp(a_k + \phi_k(b_j + \theta_r))} \right)^{y_{ijm\ell}} \text{for Model (1)} \tag{4}$$

$$\prod_{\ell=1}^{L} \left(\frac{\exp(a_\ell - b_j - \theta_r)}{1 + \exp(a_\ell - b_j - \theta_r)} - \frac{\exp(a_{\ell-1} - b_j - \theta_r)}{1 + \exp(a_{\ell-1} - b_j - \theta_r)} \right)^{y_{ijm\ell}} \text{for Model (2)} \tag{5}$$

$$\prod_{\ell=1}^{L} \left(\frac{\exp(a_\ell - b_{j\ell} - \theta_r)}{1 + \exp(a_\ell - b_{j\ell} - \theta_r)} - \frac{\exp(a_{\ell-1} - b_{j(\ell-1)} - \theta_r)}{1 + \exp(a_{\ell-1} - b_{j(\ell-1)} - \theta_r)} \right)^{y_{ijm\ell}} \text{for Model (3).} \tag{6}$$

Each follow-up time point may have a different number of observations due to missing patient responses.

2.2. The Cox Proportional Hazards Model

For the survival part of the joint model, we first consider the Cox proportional hazards model that treats progression to CDC-B stage as the only survival endpoint. Let X be a time-independent covariate, such as the treatment in this study (Continuous anti-retroviral treatment: $X = 1$, or CD4-guided interruption treatment: $X = 0$). The hazard function for the failure time of the i^{th} subject is of the form

$$\begin{aligned} \lim_{\Delta t \to 0} \left[\frac{\Pr\left(t \leqslant T_i < (t + \Delta t) \middle| T_i \geqslant t, X_i, \theta_r\right)}{\Delta t} \right] &= h(t|X_i, \theta_r) \\ &= h_0(t) \exp(\theta_r \delta_0 + X_i \delta_1) \end{aligned} \tag{7}$$

where $h_0(t)$ is a positive-valued baseline hazard function and T_i is the time to the CDC-B status for subject i. The latent variable θ_r is linked with the ordered stereotype model and δ_0 and δ_1 are the parameters for the exponential term. The model can be extended by linking the grouping information only, whereby the exponential term becomes $\exp(\eta_r + X_i\delta_1)$, where η_r is the effect associated with the rth group. Furthermore, the model can have additional covariates besides X_i.

When the competing risk approach is considered, the hazard function has the form $h_{0k}(t) \exp(\theta_r \delta_{k0} + X_i \delta_{k1})$ for event type k ($k = 1, \ldots, K$). For example, the Staccato study has two types of events – CDC-B and CDC-C. Throughout the paper, we illustrate the proposed method using the hazard function given by (7) for convenience, but we give results for both approaches in the example section.

For the estimation of the baseline hazard function $h_0(t)$, we use the method of nonparametric maximum likelihood described in Kalbfleisch & Prentice (2002, section 4.3) and in van der Vaart (2000, p. 403). Let λ_i be the hazard at time t_i, where $t_1 < t_2 < \ldots < t_n$. Assume that the hazard is zero between adjacent times so that the survival time is discrete. The baseline hazard function is $h_0(t) = \lambda_i$ if $t = t_i$ and 0 otherwise. The corresponding cumulative hazard function $H_0(t_i) = \sum_{p \leq i} \lambda_p$ is a step function with a jump at each failure time t_i. Given this structure, semiparametric problem in (7) could be considered as a parametric problem. The contribution of the ith subject to the likelihood of the survival part

becomes

$$\Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) = \left(\lambda_i \exp(\theta_r \delta_0 + X_i \delta_1)\right)^{d_i} \times \exp\Big(- \sum_{p \le i} \lambda_p \exp(\theta_r \delta_0 + X_i \delta_1)\Big). \tag{8}$$

This contribution is composed of parametric and nonparametric components whence the model is called a *semi-parametric* model. The parametric part is the exponential function of the unknown coefficients (δ_0 and δ_1) and the nonparametric component involves the baseline hazard function. The variable d_i is an indicator of an observed event for individual i. If we observe an event time for individual i then $d_i = 1$, otherwise $d_i = 0$.

2.3. The Full Likelihood Function

The joint model in this paper is a combination of a finite mixture model and a semi-parametric model. The joint likelihood function is obtained by combining the probability function from one of the three ordinal response models (4), (5), (6), and the semi-parametric proportional hazards model (8), by assuming the two models to be independent given the latent discrete random variables.

Let π_r be the unknown probability ($r = 1, \ldots, R$) that a subject lies in group r, and $\boldsymbol{\Theta}$ be all the unknown parameters of the joint model. The incomplete data likelihood function is the product of the likelihoods for the individuals

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}) = \prod_{i=1}^{n} \left(\sum_{r=1}^{R} \Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right) \Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) \pi_r \right). \tag{9}$$

Let Z_{ir} be a group membership indicator for individuals that Z_{ir} = 1 if the i^{th} individual is from the r^{th} group and 0 otherwise. The complete data likelihood can then be written as

$$L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{Z}) = \prod_{i=1}^{n} \prod_{r=1}^{R} \Big(\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right) \Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) \pi_r \Big)^{Z_{ir}}. \tag{10}$$

The complete data log likelihood becomes

$$\begin{aligned} \log\big(L(\boldsymbol{\Theta}|\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{Z})\big) &= \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \, \log\Big(\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right) \Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) \pi_r \Big) \\ &= \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \, \log(\pi_r) + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log\Big(\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right) \Big) \\ &\quad + \sum_{i=1}^{n} \sum_{r=1}^{R} Z_{ir} \log\Big(\Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right) \Big), \end{aligned} \tag{11}$$

where $\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right)$ and $\Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right)$ are defined in (4), (5) or (6) and (8) respectively.

3. Parameter Estimation

To estimate all parameters and the baseline hazards simultaneously, we combine the EM algorithm and the method of nonparametric maximum likelihood. We first write the

expectation of the complete data log likelihood given the observed data $\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}$ and the current estimates $\boldsymbol{\Theta}^{(t)}$ as follows:

$$
\begin{aligned}
Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(t)}) &= E_Z\Big[\log \Pr(\boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{Z}|\boldsymbol{\Theta}) \,\Big|\, \boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{\Theta}^{(t)}\Big] \\
&= E_Z\Big[\sum_{i=1}^{n}\sum_{r=1}^{R} Z_{ir}\, \log(\pi_r) + \sum_{i=1}^{n}\sum_{r=1}^{R} Z_{ir} \log\Big(\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right)\Big) \\
&\quad + \sum_{i=1}^{n}\sum_{r=1}^{R} Z_{ir} \log\Big(\Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right)\Big) \,\Big|\, \boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{\Theta}^{(t)}\Big] \\
&= \sum_{i=1}^{n}\sum_{r=1}^{R} \Big\{ \left[\log(\pi_r) + \log\big(\Pr(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha})\big) + \log\big(\Pr(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta})\big)\right] \\
&\quad \times \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)})\Big\} \\
&= \sum_{i=1}^{n}\sum_{r=1}^{R} \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}) \log(\pi_r) \\
&\quad + \sum_{i=1}^{n}\sum_{r=1}^{R} \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}) \log\big(\Pr(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha})\big) \\
&\quad + \sum_{i=1}^{n}\sum_{r=1}^{R} \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}) \log\big(\Pr(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta})\big). \qquad (12)
\end{aligned}
$$

In the following sections (3.1, 3.2, and 3.3), we show the steps of implementing the EM algorithm and the method of nonparametric maximum likelihood.

3.1. Baseline Hazard Estimation

Before starting the EM-step, we profile out the baseline hazard function $h_0(t)$. The third part of equation (12) is composed of the baseline hazard function and can be separately written by substituting from (8) as

$$
\begin{aligned}
\ell(\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^{n}\sum_{r=1}^{R} \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}) \Big\{ d_i \Big(\log\lambda_i + \theta_r\delta_0 + X_i\delta_1\Big) \\
- \sum_{p \le i} \lambda_p \exp(\theta_r\delta_0 + X_i\delta_1)\Big\}. \qquad (13)
\end{aligned}
$$

To apply profile likelihood function whereby the log likelihood (13) is maximized with respect to λ_i holding $(\boldsymbol{\theta}, \boldsymbol{\delta})$ fixed, we set

$$
\begin{aligned}
\frac{\partial}{\partial\lambda_i}\ell(\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\delta}) \;=\; & \sum_{r=1}^{R} \Pr(Z_{ir} = 1 | \boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}) \frac{d_i}{\lambda_i} \\
& - \sum_{p \ge i}\sum_{r=1}^{R} \Pr(Z_{pr} = 1 | \boldsymbol{Y}_p, T_p, D_p, \boldsymbol{\Theta}^{(t)}) \exp(\theta_r\delta_0 + X_p\delta_1) = 0.
\end{aligned}
$$

This implies

$$\begin{aligned}&\sum_{r=1}^{R}\Pr(Z_{ir}=1|\boldsymbol{Y}_i,T_i,D_i,\boldsymbol{\Theta}^{(t)})\frac{d_i}{\lambda_i}\\ =\;&\sum_{p\geq i}\sum_{r=1}^{R}\Pr(Z_{pr}=1|\boldsymbol{Y}_p,T_p,D_p,\boldsymbol{\Theta}^{(t)})\exp(\theta_r\delta_0+X_p\delta_1).\end{aligned}$$

Using $\sum_{r=1}^{R}\Pr(Z_{ir}=1|\,\boldsymbol{Y}_i,T_i,D_i,\boldsymbol{\Theta}^{(t)})=1$ we obtain

$$\begin{aligned}\frac{d_i}{\lambda_i}&=\sum_{p\geq i}\sum_{r=1}^{R}\Pr(Z_{pr}=1|\boldsymbol{Y}_p,T_p,D_p,\boldsymbol{\Theta}^{(t)})\exp(\theta_r\delta_0+X_p\delta_1)\\ \widehat{\lambda}_i&=\frac{d_i}{\sum_{p\geq i}\sum_{r=1}^{R}\Pr(Z_{pr}=1|\boldsymbol{Y}_p,T_p,D_p,\boldsymbol{\Theta}^{(t)})\exp(\theta_r\delta_0+X_p\delta_1)}.\end{aligned}\tag{14}$$

The E- and M-steps are carried out iteratively with the baseline hazard replaced by its nonparametric maximum likelihood estimate (14).

Note: The parametrization used for the baseline hazard function is consistent with the standard one. If the observed time t_i is a censoring time, then $d_i=0$ and as a result the estimator $\hat{\lambda}_i$ in (14) is 0.

3.2. The E-step

In the E-step, we use the current parameter estimates $\boldsymbol{\Theta}^{(t)}$ to find the expected values of Z_{ir} of the complete data log likelihood. The expected values of a Bernoulli distribution are determined by the probability of success for individual i being in group r given the observed data. Thus, using Bayes' rule, we can compute

$$\begin{aligned}&E\big[Z_{ir}|\,\boldsymbol{Y}_i,T_i,D_i,\,\boldsymbol{\Theta}^{(t)}\big]=\Pr(Z_{ir}=1|\,\boldsymbol{Y}_i,T_i,D_i,\,\boldsymbol{\Theta}^{(t)})\\ =\;&\frac{\Pr(\boldsymbol{Y}_i|\,Z_{ir}=1,\,\boldsymbol{\Theta}^{(t)})\Pr(T_i,D_i|\,Z_{ir}=1,\,\boldsymbol{\Theta}^{(t)})\Pr(Z_{ir}=1|\,\boldsymbol{\Theta}^{(t)})}{\sum_{g=1}^{R}\Pr(\boldsymbol{Y}_i|\,Z_{ig}=1,\,\boldsymbol{\Theta}^{(t)})\Pr(T_i,D_i|\,Z_{ig}=1,\,\boldsymbol{\Theta}^{(t)})\Pr(Z_{ig}=1|\,\boldsymbol{\Theta}^{(t)})}\\ =\;&\frac{\pi_r^{(t)}\;\Pr\left(\boldsymbol{Y}_i\,|\,\theta_r^{(t)},\boldsymbol{\alpha}^{(t)}\right)\Pr\left(T_i,D_i\,|\,\boldsymbol{\lambda},\theta_r^{(t)},\boldsymbol{\delta}^{(t)}\right)}{\sum_{g=1}^{R}\pi_g^{(t)}\;\Pr\left(\boldsymbol{Y}_i\,|\,\theta_g^{(t)},\boldsymbol{\alpha}^{(t)}\right)\Pr\left(T_i,D_i\,|\,\boldsymbol{\lambda},\theta_g^{(t)},\boldsymbol{\delta}^{(t)}\right)}.\end{aligned}$$

Consequently the posterior class membership probabilities for the i^{th} individual are given by

$$\Pr\left(Z_{ir}=1|\,\boldsymbol{Y}_i,T_i,D_i,\,\boldsymbol{\Theta}^{(t)}\right)=\frac{\pi_r^{(t)}\;\Pr\left(\boldsymbol{Y}_i\,|\,\theta_r^{(t)},\boldsymbol{\alpha}^{(t)}\right)\Pr\left(T_i,D_i\,|\,\boldsymbol{\lambda},\theta_r^{(t)},\boldsymbol{\delta}^{(t)}\right)}{\sum_{g=1}^{R}\pi_g^{(t)}\;\Pr\left(\boldsymbol{Y}_i\,|\,\theta_g^{(t)},\boldsymbol{\alpha}^{(t)}\right)\Pr\left(T_i,D_i\,|\,\boldsymbol{\lambda},\theta_g^{(t)},\boldsymbol{\delta}^{(t)}\right)}.\tag{15}$$

3.3. The M-step

In the M-step, we maximize equation (12) with respect to π_r and $\boldsymbol{\Theta}=(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\delta})$. Due to the fact that there is no relationship between π_r and $\boldsymbol{\Theta}$, they can be estimated separately.

1. It is straight forward to calculate the estimates of π_r. These are:

$$\widehat{\pi_r} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \Pr(Z_{ir}=1|\boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}^{(t)}).$$

2. To estimate $\boldsymbol{\Theta}$, we maximize the second and third parts of equation (12) numerically with respect to $\boldsymbol{\Theta}$. Note that $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta})$ consists of parameters from one of the three ordinal response models and the Cox model.
   Throughout this process, the stereotype likelihood function is optimized iteratively. We use an alternating algorithm (Greenland 1994) to estimate the parameters in this model because of the complicating factor of the likelihood function being multiplicative in its parameters. That is, $\{\phi_\ell\}$ and $(b_j + \theta_r)$ are alternately held fixed while the other is estimated. We propose a two-step iterative method to estimate all parameters in both models simultaneously, as follows:
   (a) Update all parameters in $\boldsymbol{\Theta}$ simultaneously with fixed ϕ_ℓ and replacing λ_i with $\hat{\lambda}_i$ as given in (14). In this step, ϕ_ℓ is treated as a known predictor.
   (b) Treat the estimated $\boldsymbol{\Theta}$ from the previous step as fixed. That is, we treat $(b_j + \theta_r)$ as a known predictor and estimate ϕ_ℓ using maximum likelihood.
   (c) Calculate λ_i by substituting the estimated parameters from the previous steps (2a and 2b) into equation (14).
   (d) Repeat the previous steps until a convergence criterion is met. We use the criterion that there is a small change ($< 10^{-5}$) in each of the estimates in subsequent iterations.

The estimated parameters from each M-step are substituted into each corresponding E-step.

We use the method described in McLachlan & Peel (2000, Section 2.15.3) to find a variance estimate for the estimator $\widehat{\boldsymbol{\Theta}}$, based on the observed information matrix from the incomplete likelihood (9) with λ_i replaced by $\hat{\lambda}_i(\hat{\boldsymbol{\Theta}})$ given by (14). Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{R-1})$ be the vector of group probabilities, where π_R is dropped due to the constraint $\pi_1 + \cdots + \pi_R = 1$. Let $\boldsymbol{\Omega} = (\boldsymbol{\Theta}, \boldsymbol{\pi})$ be the combined vector of parameters. The information matrix $I(\boldsymbol{\Omega})$ is estimated by

$$\hat{I}(\hat{\boldsymbol{\Omega}}) = \sum_{i=1}^{n} s(\boldsymbol{Y}_i, T_i, D_i; \hat{\boldsymbol{\Omega}}) s(\boldsymbol{Y}_i, T_i, D_i; \hat{\boldsymbol{\Omega}})^T$$

where $\hat{\boldsymbol{\Omega}} = (\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})$ is the MLE of $\boldsymbol{\Omega} = (\boldsymbol{\Theta}, \boldsymbol{\pi})$, and

$$\begin{aligned} & s(\boldsymbol{Y}_i, T_i, D_i; \boldsymbol{\Omega}) \\ = & \sum_{r=1}^{R} \Pr\left(Z_{ir}=1|\boldsymbol{Y}_i, T_i, D_i, \boldsymbol{\Theta}\right) \frac{\partial}{\partial \boldsymbol{\Omega}} \log\left\{\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right)\ \Pr\left(T_i, D_i \,|\, \hat{\boldsymbol{\lambda}}(\boldsymbol{\Theta}), \theta_r, \boldsymbol{\delta}\right)\ \pi_r\right\}, \end{aligned}$$

where $\hat{\boldsymbol{\lambda}}(\boldsymbol{\Theta}) = (\hat{\lambda}_1(\boldsymbol{\Theta}), \ldots, \hat{\lambda}_n(\boldsymbol{\Theta}))$. The terms $\Pr\left(\boldsymbol{Y}_i \,|\, \theta_r, \boldsymbol{\alpha}\right)$ and $\Pr\left(T_i, D_i \,|\, \boldsymbol{\lambda}, \theta_r, \boldsymbol{\delta}\right)$ are defined in (4), (5) or (6) and (8) respectively. The variance estimate for $\hat{\boldsymbol{\Theta}}$ is the $(1,1)$ component of the partitioned matrix

$$\hat{I}(\hat{\boldsymbol{\Omega}})^{-1} = \widehat{\mathrm{Var}}(\hat{\boldsymbol{\Omega}}) = \begin{pmatrix} \widehat{\mathrm{Var}}(\hat{\boldsymbol{\Theta}}) & \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) \\ \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})^T & \widehat{\mathrm{Var}}(\hat{\boldsymbol{\pi}}) \end{pmatrix}$$

TABLE 1

Summary of profile Akaike information criterion (pAIC).

Covariates in the survival part	N_{gr}	Stereotype N_{par}	Stereotype pAIC	PO N_{par}	PO pAIC
(a) QOL, trt	2	41	126764.2	38	119106.8
(b) QOL, trt, CD4	2	42	119005.5	39	119106.0
(c) QOL, trt, CD4, age, weight	3	45	116841.1	42	116537.2
(d) QOL, trt, CD4, age, sex	3	45	116611.0	42	116538.0
(e) QOL, trt, CD4, age	3	44	116594.8	41	116538.1
(f) QOL, trt, CD4, sex	3	44	116483.8	41	116537.6
(g) QOL, trt, age, sex	3	44	116480.8	41	116537.7
(h) QOL, trt, CD4	3	43	116468.3	40	116539.1
(i) QOL, trt	3	42	116466.3	39	116539.8
(j) QOL, trt, age	3	43	116465.3	40	116538.9
(k) QOL, trt, sex	3	43	116464.0	40	116537.2
(l) QOL, trt, CD4	4	44	115605.4	41	115539.2
(m) QOL, trt	4	43	115603.4	40	115535.4
(n) QOL, trt	5	44	115205.2	41	115252.4
(o) QOL, trt	6	45	115217.8	42	115264.2

4. Application

In the analysis of data from the Staccato study, we fit various joint models where the survival component contains different sets of covariates and a single survival endpoint. We only consider proportional odds and stereotype models for the item response component. Because the questionnaire has many questions, the partial proportional odds model (3) would have 124 $(= (L-1) \times (J-1) = 4 \times 31)$ item effect parameters. It is not feasible to estimate this large a number of parameters from our data. However the partial proportional odds model was included in our simulation study.

The covariates in the survival model involve the QOL score from an ordinal response model, the treatment (trt), initial CD4 count (CD4), age, sex and weight. The number of groups (R) associated with the latent variables varies as well. We use the profile Akaike information criterion (pAIC) for model selection since the models contain nuisance parameters $\{\lambda_i\}$. Table 1 displays a summary of the profile Akaike information criterion for each of the stereotype and proportional odds (PO) models. The number of groups for the discrete latent variable in the model is represented by N_{gr} and the number of parameters is N_{par}. The results are first ordered by N_{gr}, and then by decreasing stereotype model pAIC within each N_{gr} level. The model with the smallest pAIC is preferred.

From Table 1 we see that the model with a five-group QOL effect and treatment (Model (n)) has the lowest pAIC and is therefore the preferred model for both the stereotype and proportional odds options. The stereotype model is preferable to the proportional odds model as it has a lower pAIC. For models with $N_{gr} = 3$, the stereotype model pAIC generally increases as more covariates are added to a model containing QOL and treatment. For $N_{gr} = 2$, inclusion of CD4 reduced pAIC and the opposite was true for $N_{gr} = 4$. The influence of N_{gr} on the significance of additional covariates suggests that QOL grouping may in part be related to and may be a proxy for the covariate effects on QOL scores. Therefore, once the optimum number of QOL groups is found ($N_{gr} = 5$ in this case), additional covariates cease to provide any additional information. If our reasoning holds,

TABLE 2

EM estimates from the selected model under the first approach.

Parameter	Estimated value	Standard error
a_2	0.845	0.072
a_3	0.052	0.069
a_4	-1.080	0.063
a_5	-4.174	0.047
b_2	2.606	0.121
b_3	4.759	0.110
b_4	5.653	0.122
b_5	3.152	0.124
b_6	0.186	0.114
b_7	2.951	0.115
b_8	0.648	0.113
b_9	5.085	0.112
b_{10}	0.456	0.122
b_{11}	3.934	0.113
b_{12}	1.786	0.115
b_{13}	1.473	0.109
b_{14}	1.793	0.112
b_{15}	2.997	0.111
b_{16}	3.861	0.112
b_{17}	2.217	0.109
b_{18}	2.416	0.110
b_{19}	1.938	0.112
b_{20}	2.727	0.112
b_{21}	2.570	0.111
b_{22}	3.372	0.110
b_{23}	2.329	0.109
b_{24}	1.496	0.109
b_{25}	3.379	0.112
b_{26}	2.650	0.108
b_{27}	2.536	0.118
b_{28}	4.880	0.114
b_{29}	6.289	0.120
b_{30}	8.496	0.137
b_{31}	4.696	0.181
b_{32}	3.006	0.121
ϕ_2	0.000	0.113
ϕ_3	0.364	0.007
ϕ_4	0.569	0.015
δ_0	-0.172	0.012
δ_1	-0.943	0.008
θ_2	3.790	0.051
θ_3	7.251	0.439
θ_4	5.427	0.015
θ_5	2.233	0.092

then use of latent variables provides an efficient way of incorporating important covariate information, some of which may not be available or possible to measure. For the proportional odds option, there was generally no material difference between models with equal N_{gr}, suggesting that model choice is largely driven by QOL grouping.

Table 2 displays the estimated parameters and their standard errors for the selected model. It includes effects of the treatment, and the quality of life where patients belong to five different groups. Both treatment and quality of life are significant factors for describing time of symptom progression to CDC-B. Since b_j expresses the strength of the j^{th} question,

the higher the value of b_j, the higher the probability that patients' responses fall in a high category in such questions. For example, if $b_1 > b_2$, the probability that patients will respond in a high category for question one is higher than for question two. Similarly, θ_r serves as a quality of life effect for group r; the greater the value of θ_r, the higher the probability that a patient in that group will respond in a high category. This also indicates that patients who are in a group with a high value of θ have a better quality of life. In the survival time component, the regression coefficient δ_0 is the effect of quality of life and δ_1 corresponds to the effect of the treatment on the hazard of symptom progression to CDC-B. A negative value of $\hat{\delta}_0$ suggests that better quality of life is associated with a reduced hazard of symptom progression to CDC-B. The hazard of progression to CDC-B reduced multiplicatively by a factor of $(e^{\hat{\delta}_0})^{\theta_r} = 0.842^{\theta_r}$ with a 95% confidence interval $(0.822^{\theta_r}, 0.862^{\theta_r})$ for patients in QOL group r compared to QOL group 1 after controlling for the patients' treatment. Likewise, δ_1 corresponds to the effect of treatment. The hazard of progression to CDC-B reduced multiplicatively by a factor of 0.389 $(= e^{\hat{\delta}_1})$ with a 95% confidence interval $(0.383,\ 0.396)$ for patients undergoing continuous anti-retroviral treatment compared to those who underwent CD4-guided interruption treatment, after controlling for patients' general quality of life through θ_r. Therefore patients on continuous treatment have a lower hazard of CDC-B progression than patients on CD4-guided treatment. By incorporating and controlling for the effect of QOL in the Cox proportional hazards model, we have shown that treatment has an effect on time to symptom progression over and above the improvement in QOL that may be associated with it.

The cumulative hazard function is obtained from the calculation of (14) using the EM algorithm. However, in most applications, we typically describe how long the study subjects live rather than how quickly they die. Thus, the survival function has received more attention than the baseline cumulative hazard function in terms of interpretation. The survival function, $S(t)$, can be derived from the cumulative hazard function $H(t)$, by $S(t) = \exp(-H(t))$.

The estimated baseline survival function is illustrated in Figure 1. It decreases sharply for approximately the first 24 weeks. The graph significantly declines again from week 40 to week 72. There are larger drops in the steps in the graph from week 72 to week 144 due to the small number of patients still at risk, but many of these are censored during the follow-up time. The initial steep decrease is because there is a high rate of progression to CDC-B in the first 24 weeks. The progression rate then decreases and remains at about the same level for the remainder of the follow-up period.

Parameter estimates for the selected model under the competing risk approach are given in Table 3. Both treatment and quality of life are significant factors for describing time of progression to CDC-B or to CDC-C. The regression coefficient δ_{10} is the coefficient of quality of life and δ_{11} corresponds to the effect of the treatment on the hazard of symptom progression to CDC-B. The results are similar to those found in Table 2. Similarly, the coefficients δ_{20} and δ_{21} correspond to the effects of quality of life and treatment on the hazard of progression to CDC-C. The hazard of progression to CDC-C reduced multiplicatively by a factor of $(e^{\hat{\delta}_{20}})^{\theta_r} = 0.803^{\theta_r}$ with a 95% confidence interval $(0.696^{\theta_r}, 0.926^{\theta_r})$ for patients in QOL group r compared to QOL group 1 after controlling for the patients' treatment. Likewise, the hazard of progression to CDC-C reduced multiplicatively by a factor of 0.749 $(= e^{\hat{\delta}_{21}})$

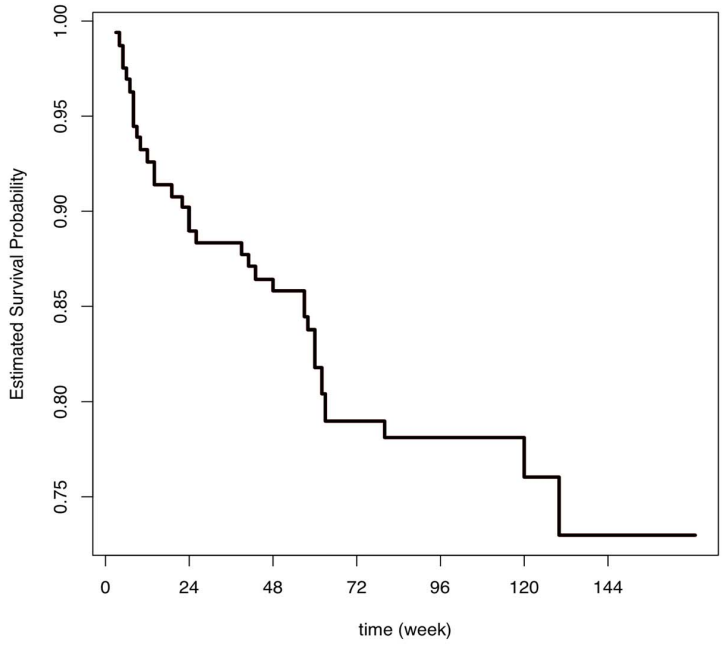

Figure 1. Estimated survival function

with a 95% confidence interval $(0.663, 0.846)$ for patients undergoing continuous antiretroviral treatment compared to those who underwent CD4-guided interruption treatment, after controlling for patients' general quality of life through θ_r.

Notice that when the item response model (1) and the survival model (7) were fitted separately, we obtained similar estimates of δ_0 and δ_1. For example, cluster information based on patients' quality of life (QOL) outcomes was obtained from the item response model (1). These clusters were then used in the survival model (7) to estimate δ_0 and δ_1. The estimates were $\hat{\delta}_0 = -0.126$ and $\hat{\delta}_1 = -0.958$, compared to the joint modeling results $\hat{\delta}_0 = -0.172$ and $\hat{\delta}_1 = -0.943$. Nonetheless separate modeling has many disadvantages including invalid inferences and lack of model selection methods for joint models.

5. Simulations

We conducted a simulation study to investigate whether a wrong choice of ordinal response model affected the performance of parameter estimators in the survival model. Because the meaning of θ_r was different for different models, we focused on the treatment effect δ_1 in Model (7). Often a clinical researcher might be interested in the patient grouping information. The simulation study also evaluated the proportion of times that patients were allocated into correct groups for different ordinal models. In particular, when the fitted ordinal

TABLE 3

EM estimates from the competing risks model.

Parameter	Estimated value	Standard error
a_2	0.845	0.027
a_3	0.053	0.045
a_4	-1.077	0.065
a_5	-4.170	0.115
b_2	2.604	0.133
b_3	4.756	0.139
b_4	5.650	0.144
b_5	3.150	0.135
b_6	0.184	0.137
b_7	2.949	0.135
b_8	0.645	0.136
b_9	5.082	0.142
b_{10}	0.453	0.136
b_{11}	3.931	0.137
b_{12}	1.784	0.134
b_{13}	1.470	0.134
b_{14}	1.791	0.134
b_{15}	2.995	0.135
b_{16}	3.858	0.137
b_{17}	2.214	0.134
b_{18}	2.413	0.134
b_{19}	1.936	0.134
b_{20}	2.724	0.134
b_{21}	2.568	0.134
b_{22}	3.369	0.135
b_{23}	2.326	0.134
b_{24}	1.494	0.134
b_{25}	3.376	0.135
b_{26}	2.647	0.133
b_{27}	2.533	0.133
b_{28}	4.878	0.141
b_{29}	6.286	0.151
b_{30}	8.493	0.182
b_{31}	4.693	0.141
b_{32}	3.003	0.136
ϕ_2	0.000	0.112
ϕ_3	0.364	0.012
ϕ_4	0.569	0.035
δ_{10}	-0.173	0.064
δ_{11}	-0.928	0.088
δ_{20}	-0.220	0.073
δ_{21}	-0.289	0.062
θ_2	3.789	0.050
θ_3	7.249	0.434
θ_4	5.425	0.155
θ_5	2.232	0.832

response model was incorrect, we evaluated the effect of this miss specification on the proportion of times that patients were allocated into correct groups.

We took the true item response model to be the stereotype model (1), since, in terms of model parsimony, the stereotype model was in the middle among the three models (1), (2), and (3). We fixed $n = 100$, $J = 6$, $M = 6$, $\delta_0 = -0.2$, and $a_\ell = 0$ for all ℓ and let R, L, b_j and δ_1 vary. Let b_j be 0 or follow $N(0, 3^2)$ for all j and let δ_1 be 0, -1, or -2. Table 4 gives

TABLE 4

The setting of parameters in simulations

R	L	parameters
3	3	$\theta_1 = 0, \theta_2 = 1$, and $\theta_3 = -1$
		$\phi_1 = 0, \phi_2 = 0.5$, and $\phi_3 = 1$
3	5	$\theta_1 = 0, \theta_2 = 1$, and $\theta_3 = -1$
		$\phi_1 = 0, \phi_2 = 0.25, \phi_3 = 0.5, \phi_4 = 0.75$, and $\phi_5 = 1$
5	3	$\theta_1 = 0, \theta_2 = 1, \theta_3 = -1, \theta_4 = 2$, and $\theta_5 = -2$
		$\phi_1 = 0, \phi_2 = 0.5$, and $\phi_3 = 1$
5	5	$\theta_1 = 0, \theta_2 = 1, \theta_3 = -1, \theta_4 = 2$, and $\theta_5 = -2$
		$\phi_1 = 0, \phi_2 = 0.25, \phi_3 = 0.5, \phi_4 = 0.75$, and $\phi_5 = 1$

the setting of parameters θ_r and ϕ_ℓ. Because the main purpose of our simulation study was to compare three ordinal models, we did not cover a wide range of scenarios on the censoring and survival functions.

Samples were generated as follows:

1. Randomly assigned each individual to the continuous treatment using a Bernoulli(p = 0.3) random variable, otherwise assigned to the CD-4 guided treatment.
2. Allocated each patient to a cluster with $\pi_r = 1/R$ for all $r = 1, \ldots, R$.
3. Generated survival times from an exponential distribution with the rate $\exp(\theta_r \delta_0 + X_i \delta_1)$.
4. Independently, generated a censoring time from an exponential distribution with the rate 0.2. Assigned to a subject a survival time or censoring time depending on which of the generated time is smaller. For example, if the survival time was smaller than the censoring time, then $d_i = 1$ for the ith patient. Otherwise, $d_i = 0$, i. e., the ith patient was censored.
5. Generated item responses from the stereotype model.

For each case, we generated 100 data sets. To each dataset we fitted three different ordinal response models jointly with the survival model.

Table 5 summarises the sample mean and the mean squared error for $\hat{\delta}_1$ over 100 simulations. All three ordinal response models gave similar estimates close to the true δ_1, but the mean squared errors showed that the partial proportional odds model was the worst among these three models. Therefore, if the true model is a stereotype model, but one uses a (partial) proportional odd model, the treatment parameter estimator in the survival model is still reliable, but suffers a loss of efficiency.

The posterior class membership probabilities (15) provide fuzzy clustering memberships for subjects. That is, an individual might belong to the first cluster with probability 0.4 and to the second cluster with probability 0.6. To compare the three ordinal response models, we simply allocated subjects to the cluster with the highest probability and used the pairwise grouping information between subjects to evaluate clustering performance. We calculated the widely used Jaccard index (Ben-Hur, Elisseeff & Guyon 2002), originally proposed by Jaccard (1901). Consider two sets A and B. The Jaccard index is $(A \cap B)/(A \cup B)$ that measures the similarity between two sets. Applying to the concept of measuring the similarity of two clusterings C and C', let N_{11} be the number of shared pairs in clusters C and C'. Let

TABLE 5

Sample mean $\hat{\delta_1}$ and mean squared error (in brackets) over 100 simulations.

R	L	δ_1	$\{b_j\}$	Stereotype	PO	Partial PO
3	3	0	$\sim N(0, 3^2)$	-0.005 (0.062)	-0.005 (0.063)	-0.003 (0.062)
3	3	0	0	0.026 (0.046)	0.027 (0.047)	0.029 (0.047)
3	3	-1	$\sim N(0, 3^2)$	-0.969 (0.064)	-0.969 (0.064)	-0.978 (0.070)
3	3	-1	0	-1.010 (0.056)	-1.014 (0.055)	-1.027 (0.062)
3	3	-2	$\sim N(0, 3^2)$	-2.019 (0.110)	-2.017 (0.111)	-2.085 (0.148)
3	3	-2	0	-1.984 (0.082)	-1.986 (0.082)	-2.036 (0.103)
3	5	0	$\sim N(0, 3^2)$	-0.006 (0.047)	-0.011 (0.048)	-0.007 (0.048)
3	5	0	0	-0.010 (0.050)	-0.011 (0.048)	0.004 (0.057)
3	5	-1	$\sim N(0, 3^2)$	-0.965 (0.072)	-0.958 (0.070)	-0.985 (0.074)
3	5	-1	0	-0.995 (0.064)	-0.990 (0.061)	-1.027 (0.067)
3	5	-2	$\sim N(0, 3^2)$	-2.010 (0.131)	-2.004 (0.128)	-2.084 (0.150)
3	5	-2	0	-1.979 (0.094)	-1.967 (0.093)	-2.041 (0.139)
5	3	0	$\sim N(0, 3^2)$	0.009 (0.051)	0.008 (0.050)	0.011 (0.051)
5	3	0	0	0.037 (0.056)	0.041 (0.060)	0.048 (0.064)
5	3	-1	$\sim N(0, 3^2)$	-0.948 (0.090)	-0.944 (0.087)	-0.972 (0.086)
5	3	-1	0	-0.959 (0.088)	-0.958 (0.090)	-0.994 (0.090)
5	3	-2	$\sim N(0, 3^2)$	-1.894 (0.112)	-1.893 (0.107)	-1.984 (0.138)
5	3	-2	0	-1.918 (0.125)	-1.923 (0.127)	-1.999 (0.157)
5	5	0	$\sim N(0, 3^2)$	0.001 (0.063)	0.004 (0.063)	0.005 (0.074)
5	5	0	0	0.036 (0.069)	0.039 (0.069)	0.054 (0.079)
5	5	-1	$\sim N(0, 3^2)$	-0.996 (0.057)	-0.986 (0.059)	-1.033 (0.072)
5	5	-1	0	-0.900 (0.071)	-0.887 (0.073)	-0.957 (0.095)
5	5	-2	$\sim N(0, 3^2)$	-1.977 (0.113)	-1.961 (0.111)	-2.104 (0.142)
5	5	-2	0	-1.925 (0.133)	-1.903 (0.130)	-2.136 (0.145)

N_{10} be the number of pairs in cluster C but not in C', and likewise let N_{01} be the number of pairs in cluster C' but not in C. Therefore the form of the Jaccard index can be written as

$$\mathcal{J}(C, C') = \frac{N_{11}}{N_{11} + N_{01} + N_{10}}$$

If the two clusters are similar, $\mathcal{J}(C, C')$ tends to 1.

Table 6 shows the average Jaccard index for 100 simulated datasets for each of the cases. It is a measure of the similarity between the true clustering and the clustering fitted by an ordinal model. For example, the clustering C represents the true grouping among the subjects and the clustering C' is the grouping result fitted by a stereotype model based on the posterior class membership probabilities (15). For the first case in Table 6, we calculated $\mathcal{J}(C, C')$ for each simulated data set. The value 0.520 was the average $\mathcal{J}(C, C')$ over 100 datasets. It was not surprising that the stereotype model was the best for all cases, because the true grouping information was based on the stereotype model. The grouping information given by the partial proportional odds model was the worst. It confirmed that if the true model was a stereotype model, a (partial) proportional odd model could result in different clusters of patients. However the information was similar between the stereotype and proportional odds models when the number of response categories was large (e. g., $L = 5$).

Notice that the similarity measure between the true clustering and the clustering fitted by a model depends on the value of $\{\theta_r\}$ and R. However the comparisons across different number of clusters are meaningless. That is, we cannot compare the Jaccard index between the one with $R = 3$ and the one with $R = 5$ (Meilă 2007). This simulation study only focused

TABLE 6

Jaccard Index Comparison.

R	L	δ_1	$\{b_j\}$	Stereotype	PO	Partial PO
3	3	0	$\sim N(0, 3^2)$	0.520	0.368	0.327
3	3	0	0	0.493	0.358	0.328
3	3	-1	$\sim N(0, 3^2)$	0.512	0.364	0.325
3	3	-1	0	0.493	0.355	0.325
3	3	-2	$\sim N(0, 3^2)$	0.509	0.372	0.325
3	3	-2	0	0.502	0.370	0.324
3	5	0	$\sim N(0, 3^2)$	0.459	0.394	0.326
3	5	0	0	0.453	0.383	0.326
3	5	-1	$\sim N(0, 3^2)$	0.463	0.420	0.325
3	5	-1	0	0.450	0.391	0.325
3	5	-2	$\sim N(0, 3^2)$	0.457	0.406	0.323
3	5	-2	0	0.452	0.392	0.322
5	3	0	$\sim N(0, 3^2)$	0.416	0.262	0.191
5	3	0	0	0.412	0.263	0.193
5	3	-1	$\sim N(0, 3^2)$	0.409	0.265	0.192
5	3	-1	0	0.426	0.269	0.191
5	3	-2	$\sim N(0, 3^2)$	0.429	0.270	0.190
5	3	-2	0	0.415	0.260	0.190
5	5	0	$\sim N(0, 3^2)$	0.389	0.315	0.191
5	5	0	0	0.389	0.304	0.192
5	5	-1	$\sim N(0, 3^2)$	0.394	0.311	0.191
5	5	-1	0	0.391	0.301	0.189
5	5	-2	$\sim N(0, 3^2)$	0.389	0.316	0.190
5	5	-2	0	0.385	0.299	0.188

on comparing the performance across three ordinal models given scenarios, it did not focus on investigating the performance for the stereotype model alone.

6. Discussion

Joint model approaches to the analysis of quality of life and survival data are appropriate when these two endpoints are likely to be related. We consider a joint model using latent variables to link the two components. We use a likelihood approach to estimate all parameters in an ordinal response model and a Cox proportional hazards model simultaneously. Our approach includes estimation of the baseline hazards. Simultaneous estimation has the advantage of strong consistency (Gao, Manatunga & Chen 2007). We treat the baseline hazard as a step function with no parametric form. Our model is thus semiparametric. We fit it by the means of the EM algorithm.

An advantage of using discrete latent variables, instead of subject-level random effects, is that this procedure allows us to classify patients into their most likely group. Each group is composed of a set of individuals who are homogeneous with respect to their characteristics or their attitudes as indicated by the quality of life survey. The similar idea of grouping a set of individuals using discrete latent variables has also been proposed by Pledger & Arnold (2014) for binary and count data. Since the number of groups is unknown, the profile Akaike information criterion is used to provide information on the best choice of the number of groups. For the Staccato study, the results showed that we could classify the patients by their

quality of life into five groups. The patients who were in a group with a high value of θ_r had a better quality of life.

We discuss the differences among stereotype, proportional odds, and partial proportional odds models. The parameters in each model have their own meaning. On the basis of a simulation study, we found that when the true model is a stereotype model, a wrong choice of models does not severely affect parameter estimates in the survival model, although a wrong choice might result in a loss of efficiency. In practice, researchers can fit different ordinal models and choose the best one using an information criterion. In our example, we found the stereotype model to have a better fit.

The data from Staccato study contain some missing responses on the QOL questionnaire. For example, some patients might not answer all 32 questions at a given time point. In our analysis, we excluded these missing responses, assuming missing at random. The joint model approach has the capacity to gain information from both components effectively under the assumption. In practice, the missing at random assumption might not hold. Future research will investigate the impact of non–random missingness on our joint model. In addition, further work will be carried out to evaluate model fit for a joint model involving finite mixtures.

References

AGRESTI, A. (2010). *Analysis of Ordinal Categorical Data, 2nd edition*, John Wiley & Sons, New Jersey.

AGRESTI, A. (2013). *Categorical Data Analysis, 3rd edition*, John Wiley & Sons, New Jersey.

ANANWORANICH, J. & THE STACCATO STUDY GROUP (2006). CD4-guided scheduled treatment interruptions compared with continuous therapy for patients infected with HIV-1: results of the Staccato randomised trial. *Lancet* **368**, 459–465.

ANDERSON, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B* **46**, 1–30.

ANDERSON, J. A. & PHILIPS, P. R. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied Statistics* **30**, 22–31.

BEN-HUR, A. & ELISSEEFF, A. & GUYON, I. (2002). A stability based method for discovering structure in clustered data. *Pacific Symposium on Biocomputing* **7**, 6–17.

CASTRO, K. G. & WARD, J. W. & SLUTSKER, L. & BUEHLER, J. W. & JAFFE, H. W. & BERKELMAN, R. L. & CURRAN, J. W. (1993). 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *Clinical Infectious Diseases* **17**, 802–810.

COX, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **37**, 187–220.

GAO, F. & MANATUNGA, A. K. & CHEN, S. (2007). Non-parametric estimation for baseline hazards function and covariate effects with time-dependent covariates. *Statistics in Medicine* **26**, 857–868.

GREENLAND, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine* **13**, 1665–1677.

HE, B. & LUO,S. (2013). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research* DOI:10.1177/0962280213480877.

HENDERSON, R.(2005). Joint modeling of longitudinal and event time data. *Encyclopedia of Biostatistics* Armitage, P. & Colton, T. (eds), John Wiley & Sons.

HOGAN, J. W. & LAIRD, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259–272.

HSIEH, F. & TSENG, Y.K. & WANG, J.L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* **62**, 1037–1043.

HSU, C. H. & TAYLOR, J. M. & MURRAY, S. & COMMENGES, D. (2006). Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* **25**, 3503–3517.

JACCARD, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**, 547–579.

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data, 2nd edition*, John Wiley & Sons, New York.

LI, N. & ELASHOFF, R. & LI, G. & SAVER, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine* **29**, 546–557.

LIU, I. & AGRESTI, A. (2005). The analysis of ordered categorical data: an overview and a survey of recent developments. *Test* **14**, 1–73.

MCCULLAGH, P. (1980). Regression-models for ordinal data. *Journal of the Royal Statistical Society, Series B* **2**, 109–142.

MCLACHLAN, G. & PEEL, D. (2000). *Finite Mixture Models*, Wiley, New York.

MEILĂ, M. (2007). Comparing clusterings – an information based distance. *Journal of Multivariate Analysis* **98**, 873–895.

NJAGI, E. N. & MOLENBERGHS, G. & RIZOPOULOS, D. & VERBEKE, G. & KENWARD, M. G. & DENDALE, P. & WILLEKENS, K. (2013). A flexible joint modeling framework for longitudinal and time-to-event data with overdispersion. *Statistical Methods in Medical Research* DOI:10.1177/0962280213495994.

PETERSON, B. & HARRELL, F. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society, Series C* **39**, 205–217.

PICKLES, A. & CROUDACE, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research* **19**, 271–289.

PLEDGER, S. & ARNOLD, R. (2014). Multivariate methods using mixtures: correspondence analysis, scaling and pattern detection. *Computational Statistics and Data Analysis* **71**, 241–261.

QIU, F. & STEIN, C. M. & ELSTON, R. C. (2013). Joint modeling of longitudinal data and discrete-time survival outcome. *Statistical Methods in Medical Research*, DOI:10.1177/0962280213490342.

RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data With Applications in R*, CRC Press.

SONG, X. & DAVIDIAN, M. & TSIATIS, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.

SONG, X. & WANG, C. Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics* **64**, 557–566.

TITTERINGTON, D. M. & SMITH, A. F. M. & MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York.

TSIATIS, A. A. & DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.

TSIATIS, A. A. & DAVIDIAN, M. (2004). Joint modeling of longitudinal and time to event data: an overview. *Statistica Sinica* **14**, 809–834.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*, Cambridge university press, Cambridge.

WANG, C. & DOUGLAS, J. & ANDERSON, S. (2002). Item response model for joint analysis of quality of life and survival. *Statistics in Medicine* **21**, 129–142.

WOODRUFF, D. J. & HANSON, B. A. (1996). Estimation of item response models using the em algorithm for finite mixtures. *ACT Research Report Series, American Coll. Testing Program, Iowa City, IA*

WU, A. W. & RUBIN, H. R. & MATHEWS, W. C. & WARE JR, J. E. & BRYSK, L. T. & HARDY, W. D. & BOZZETTE, S. A. & SPECTOR, S. A. & RICHMAN, D. D. (1991). A health status questionnaire using 30 items from the Medical Outcomes Study: preliminary validation in persons with early HIV infection. *Medical Care* **29**, 786–798.

WU, A. W. & REVICKI, D. A. & JACOBSON, D. & MALITZ, F. E. (1997). Evidence for reliability, validity and usefulness of the medical outcomes study HIV health survey (MOS-HIV). *Quality of Life Research* **6**, 481–493.

WU, L. & LIU, W. & YI, G.Y. & HUANG, Y. (2012). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics.*

WULFSOHN, M. S. & TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

XU, J. & ZEGER, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society, Series C* **50**, 375–387.