South African Statistical Journal
Vol. 56, No. 1, pp. 21–36
https://doi.org/10.37920/sasj.2022.56.1.2

A regression analysis of discrete time competing risks data using a vertical model approach

Bonginkosi D. Ndlovu[1], *Sileshi F. Melesse*[2] *and Temesgen Zewotir*[3]

[1]Department of Statistics, Durban University of Technology, Durban, South Africa
[2]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
[3]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Over the years, the standard regression analysis method for discrete time competing risks data has been to model the data with discrete time cause-specific hazards. While a few continuous time competing risks models have been proposed in the literature, it is a well documented fact that these models are not appropriate for application in discrete time as is. The vertical regression model of Nicolaie et al. (2010) is the latest of these continuous time competing risks models. We reformulate this regression model for the purpose of application in discrete time. We demonstrate that the proposed model can easily be implemented by using existing software for discrete time models. We apply the proposed model together with some of the existing discrete time models to real discrete time competing risks data and find that the proposed model and these models compare favourably.

Keywords: Discrete time competing risks, Relative hazards, Total hazards, Vertical model.

1. Introduction

Competing risks models continue to be popular with researchers in applied sciences where time is measured in discrete units; see Hwang and Chu (2013), Cleric et al. (2014), Bertoni and Groh (2014), and Vallejos and Steel (2017) for examples of recent applications. The data that arise from such applications are often referred to as discrete time competing risks data, i.e., the subjects are exposed to multiple risks of failure and the time to failure is observed in discrete units. With J failure causes, let $\tilde{T}$ and D denote time to failure and failure type, respectively, such that $D \in \{1, 2, \ldots, J\}$. Suppose that C is time to censoring. In general, observed competing risks data can be represented by $\boldsymbol{y} = (t_i, \boldsymbol{x}_i^T, \Delta_i)^T$, $i = 1, \ldots n$, where $T_i = \tilde{T}_i \wedge C_i$ and $\Delta_i = I(\tilde{T} \leq C_i)D_i$. The vector $\boldsymbol{x}_i$ represents the set of covariates that describe subject i. To distinguish between continuous time and discrete time competing riks data, it is assumed that time is observed in discrete units, i.e., $\tilde{T}, C \in \{1, 2, \ldots, q\}$ for a positive integer q. The multinomial model has been the standard regression model for analysis of discrete time competing risks data (Ambrogi et al., 2009; Tutz and Schmid, 2016). This model

Corresponding author: Bonginkosi D. Ndlovu (bongi@dut.ac.za)
MSC2020 subject classifications: 62N02.

proposes discrete time cause-specific hazards for modelling competing risks data. The definition of discrete time cause-specific hazards is given by

$$\lambda_j(t) = P(T = t, D = j | T \geq t),$$

for $t = 1, \ldots, q$ and $j = 1, \ldots, J$. With covariates, the model for cause-specific hazards is given by

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\alpha}) = \frac{\exp\left(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j}\right)}{1 + \sum_{l=1}^{J} \exp\left(\alpha_{0lt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1l}\right)}, \tag{1}$$

for $j = 1, \ldots, J$ and $t = 1, \ldots, q$, where α_{0lt} is the duration coefficient at time t and $\boldsymbol{\alpha}_{1j}$ is a vector of regression coefficients. The cause-specific hazard parameters $\boldsymbol{\alpha}_j$, $j = 1, \ldots, J$, where $\boldsymbol{\alpha}_j = (\alpha_{0j1} \ldots \alpha_{0jq}, \boldsymbol{\alpha}_{1j}^T)^T$ are estimated simultaneously by fitting a multinomial distribution to observed data in a long format. The estimates for the cumulative incidence function can be obtained from

$$\hat{F}_j(t|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}) = \sum_{s:s \leq t} \hat{S}(s-1|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}) \hat{\lambda}_j(s|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}), \tag{2}$$

for $j = 1, \ldots, J$ and $s = 1, \ldots, q$, where $\hat{S}(t|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}) = \prod_{s=1}^{t} (1 - \hat{\lambda}(s|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}))$ and $\hat{S}(t = 0|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}) = 1$. Lee et al. (2018) proposed a regression model for estimating the cause-specific hazards individually by applying a binomial distribution. Henceforth, we shall refer to this model as the binomial model. The model for the cumulative incidence function in (2) complicates the evaluation of the covariate effects. Recently, Berger et al. (2020) advanced a method for modelling the cumulative incidence function directly on covariates as an extension of the model proposed by Fine and Gray (1999) for discrete time data. For further discussion of discrete time competing risks models; see, for example, Schmid and Berger (2020). There are other methods for modelling the cause-specific hazards such as *Classification and Regression Trees* (CART) which are also discussed in Schmid and Berger (2020).

The vertical model (Nicolaie et al., 2010) is the latest continuous time competing risks model. Our main objective in this article is to adapt this model for the purposes of application in discrete time. The vertical model postulates a decomposition of the joint distribution of $(\tilde{T}; D)$ into a distribution for failure type conditional on failure time and a marginal distribution for failure time:

$$P(\tilde{T}; D) = P(D|\tilde{T})P(\tilde{T}).$$

Amongst other things, this assumption gives rise to characterisation of competing risks data in terms of failure type probabilities conditional on failure time (relative hazards) and total hazards as an alternative to the more familiar and popular cause-specific hazards. In fact, the cause-specific hazards and all other estimands, including the cumulative incidence function estimates, are now obtained from the relative hazards and total hazard estimates. While modelling observed data with the cause-specific hazards has its advantages, such as allowing for a direct and simpler means for evaluation of covariate effects on the risks of failure, in some instances it is not possible to estimate these quantities directly from data. It is not possible, for example, to model data with cause-specific hazards when some of the subjects have failed with unknown failure causes. This means that, in the presence of missing failure causes, the existing discrete time models are immediately disqualified from application unless data are edited by, for example, excluding the affected subjects. Nicolaie et al. (2015) demonstrated that the vertical model of Nicolaie et al. (2010) is invariant to the presence or absence of missing failure causes, i.e., the model can also be applied to data with missing failure causes as is in the continuous time domain. Another complication that may arise with competing risk data is that it may come with a sizeable proportion of cured subjects. Cured subjects are assumed to be mixed with censored subjects and the various methods for handling these subjects advance different techniques for splitting cured subject from the subjects that will eventually fail. The multinomial model, on the other hand, retains the censored subjects intact and usually regards them as reference category when the cause-specific hazard parameters are estimated. In the estimation of target cause-specific hazards, the binomial model (Lee et al., 2018) is just as handicapped because it regards the competing failure times as censored and lumps them together with real censored subjects as a reference category. The same authors, that is, Nicolaie et al. (2018), have extended the vertical model for handling cured subjects in the continuous time setting. This is not the first time that another option for modelling data has been suggested as an alternative to the cause-specific hazards. The model that has come to be known as the mixture competing risks model (Larson and Dinse, 1985) proposes failure-type probabilities and component hazards for modelling competing risks data. This follows from yet another reformulation of the bivariate distribution of $(\tilde{T}, D)$ that is expressed in terms of a marginal distribution for failure type and a failure time distribution conditional on failure type:

$$P(\tilde{T}, D) = P(D)P(\tilde{T}|D).$$

This approach has proved to be more flexible than the cause-specific hazards — for example, the mixture model forms the bases for the well known mixture cure model; see Peng and Taylor (2014) for a review of this model. Furthermore, the model has been extended to handle cured subjects in competing risks settings (Choi, 2002; Zhiping, 2011). In presenting a vertical model in discrete time, it is envisaged that the proposed model can, in the least, be upscaled to address the limitations of the cause-specific hazard denominated discrete time competing risks model as discussed, i.e., the handling of missing failure causes and cured subjects.

The vertical model is a "mixture" of a failure type and a failure time component, and as such, it allows for separate models for total and relative hazards. When Nicolaie et al. (2010) introduced the vertical model in continuous time, they modelled the total hazards to follow the Cox (1972) proportional hazards assumption. To relocate the vertical model in the discrete time realm, we propose discrete time total hazards to characterise the marginal failure time distribution. Consider the following definition of discrete time total hazards:

$$\lambda(t) = P(T = t|T \geq t) = \sum_{j=1}^{J} \lambda_j(t),$$

for $t = 1, \ldots, q$. Consistent with discrete time survival analysis models (see, for example, Allison, 1982; Singer and Willet, 2003; Tutz and Schmid, 2016), the regression model for the total hazards can be expressed as

$$g(\lambda(t|\boldsymbol{x}, \boldsymbol{\beta})) = \beta_{0t} + \boldsymbol{x}^T \boldsymbol{\beta}_1,$$

for $t = 1, \ldots, q$, where $g(\cdot)$ is a link function within the GLM framework. The scalar β_{0t} is the baseline total hazard coefficient at time t and $\boldsymbol{\beta}_1$ is a vector of regression coefficients. The definition of relative hazards is given by

$$\pi_j(t) = P(D=j|T=t) = \frac{P(D=j;T=t)}{P(T=t)} = \frac{P(D=j;T=t)/P(T\geq t)}{P(T=t)/P(T\geq t)}$$
$$= \frac{P(D=j;T=t;T\geq t)/P(T\geq t)}{P(T=t;T\geq t)/P(T\geq t)} = \frac{\lambda_j(t)}{\lambda(t)}, \quad (3)$$

whence the term "relative hazards". Since $D \in \{1,\dots,J\}$, a multinomial distribution is the most natural model for the conditional failure type distribution where the model for relative hazards is given by

$$\pi_j(t|\boldsymbol{x},\gamma) = \frac{\exp\left(\gamma_{0jt} + \boldsymbol{x}^T\gamma_{1j}\right)}{\left\{1+\sum_{l=1}^{J-1}\exp\left(\gamma_{0lt}+\boldsymbol{x}^T\gamma_{1l}\right)\right\}},$$

for $j = 1,\dots,J-1$, and $t = 1,\dots,q$, with $\pi_J(t|\boldsymbol{x},\gamma) = 1 - \sum_{j=1}^{J-1}\pi_j(t|\boldsymbol{x},\gamma)$. The scalar γ_{0jt} is the duration coefficient at time t, and $\boldsymbol{\gamma}_{1j}$ is a vector of regression coefficients.

Collect all the unknown parameters of the proposed model in $\boldsymbol{\theta} = (\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T)^T$, where $\boldsymbol{\beta} = (\beta_{01},\dots\beta_{0q},\boldsymbol{\beta}_1^T)^T$ and $\boldsymbol{\gamma} = (\gamma_1^T,\dots\gamma_{(J-1)}^T)^T$, where $\gamma_j = (\gamma_{0j1},\dots\gamma_{0jq},\gamma_{1j}^T)^T$. As will be shown in the next section, the full likelihood function mimics the underlying assumption of the model by splitting into a failure time likelihood function and a conditional failure type likelihood function which are specified in terms of total hazards and relative hazards, respectively. The split allows for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to be estimated separately. The most notable result of the vertical model assumption, i.e, $P(\tilde{T};D) = P(D|\tilde{T})P(\tilde{T})$, is that the cause-specific hazard estimates are now derived from the total hazard and relative hazard estimates

$$\hat{\lambda}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \hat{\lambda}(t|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\pi}_j(t|\boldsymbol{x},\hat{\gamma}),$$

for $t = 1,\dots,q$ and $j = 1,\dots J$. This result follows from rearranging (3). The cumulative incidence function estimates can be obtained from

$$\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \sum_{s:s\leq t}\hat{S}(s-1|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\lambda}(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\pi}_j(s|\boldsymbol{x},\hat{\gamma}), \quad (4)$$

for $t = 1,\dots q$ and $j = 1,\dots J$, where $\hat{S}(t|\boldsymbol{x},\hat{\boldsymbol{\beta}}) = \sum_{s=1}^{t}(1-\hat{\lambda}(s|\boldsymbol{x},\hat{\boldsymbol{\beta}}))$ and $\hat{S}(t=0|\boldsymbol{x},\hat{\boldsymbol{\beta}}) = 1$. This expression for the cumulative incidence function is no simpler than the expression given in (2), because the cumulative incidence function is also modelled indirectly on data via the regression models for the total and relative hazards in comparison to the model advanced by Berger et al. (2020). This concludes the presentation of the vertical model as a discrete time model. The remainder of the article is organised as follows. In Section 2, we demonstrate that $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be estimated with standard statistical software. This is followed by an application of the proposed model to real discrete time competing risks data together with some of the existing discrete time competing risks models for comparison purposes in Section 3. We conclude with a discussion in Section 4. The standard errors for the cumulative incidence function estimates are derived in the Appendix.

2. Estimation

To determine the MLE of $\boldsymbol{\theta}$ we maximise the observed data likelihood function w.r.t. $\boldsymbol{\theta}$. In the construction of this likelihood function, the contribution of a subject i that failed at time t_i is now

$P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \gamma)P(T_i = t_i|\boldsymbol{x}_i, \boldsymbol{\beta})$. Let $d_{ij} = I(D_i = j)$ and $d_i = \sum_{j=1}^{J} d_{ij}$. The observed data log-likelihood can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \gamma)P(T_i = t_i|\boldsymbol{x}_i, \boldsymbol{\beta}) + (1-d_i)\log P(T_i > t_i|\boldsymbol{x}_i, \boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i; \boldsymbol{x}_i, \gamma) \\ &\quad + \sum_{i=1}^{n} d_i \log P(T_i = t_i|\boldsymbol{x}_i, \boldsymbol{\beta}) + (1-d_i)\log P(T_i > t_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = \mathcal{L}(\gamma) + \mathcal{L}(\boldsymbol{\beta}).\end{aligned}$$

The observed data log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$ therefore splits into $\mathcal{L}(\gamma)$, a conditional failure type log-likelihood function, and $\mathcal{L}(\boldsymbol{\beta})$, a standard univariate failure time log-likelihood function. The conditional failure type log-likelihood $\mathcal{L}(\gamma)$ can be written as

$$\mathcal{L}(\gamma) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_j(t_i|\boldsymbol{x}_i, \gamma).$$

The conditional failure type log-likelihood $\mathcal{L}(\gamma)$ is easily recognisable as a kernel of a multinomial log-likelihood function. Thus, γ can be estimated by fitting a multinomial distribution to the original data by including duration as factor because the failure type probabilities are determined at each observed failure time, i.e., at $t \in \{1, 2, \ldots, q\}$. By definition, censored subjects are excluded from the estimation of relative hazards. The failure time log-likelihood function $\mathcal{L}(\boldsymbol{\beta})$ is a standard failure time log-likelihood function. Note that since time to failure is discrete, the definition of the density function is given by

$$P(T = t|\boldsymbol{x}) = \frac{\lambda(t|\boldsymbol{x}, \boldsymbol{\beta})}{(1-\lambda(t|\boldsymbol{x}, \boldsymbol{\beta}))}\prod_{s=1}^{t}(1-\lambda(s|\boldsymbol{x}, \boldsymbol{\beta})).$$

Thus, $\mathcal{L}(\boldsymbol{\beta})$ can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} d_i \log\frac{\lambda(t_i|\boldsymbol{x}_i, \boldsymbol{\beta})}{(1-\lambda(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}))} + \sum_{s=1}^{t_i} d_i \log(1-\lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta})) + \sum_{s=1}^{t_i}(1-d_i)\log(1-\lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta})) \\ &= \sum_{i=1}^{n} d_i \log\frac{\lambda(t_i|\boldsymbol{x}_i, \boldsymbol{\beta})}{(1-\lambda(t_i|\boldsymbol{x}_i, \boldsymbol{\beta}))} + \sum_{s=1}^{t_i}\log(1-\lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta})).\end{aligned}$$

If we define d_{is} such that $d_{is} = 0$ for $s \le t_i - 1$ and $d_{it_i} = d_i$, then $\mathcal{L}(\boldsymbol{\beta})$ can be rewritten as

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{s=1}^{t_i} d_{is}\log\lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta}) + (1-d_{is})\log(1-\lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta})).$$

This means that $\boldsymbol{\beta}$ can be estimated within the GLM framework by fitting a binomial distribution, where $d_{is} \sim \mathcal{B}(1, \lambda(s|\boldsymbol{x}_i; \boldsymbol{\beta}))$. The data must be rearranged into a long format, where subject i contributes $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{it_i})^T$ as a response vector of successes out of $(1, \ldots, 1)^T$ trials with $\boldsymbol{x}_i$

repeated t_i times. This can easily be achieved by applying the R package `discSurv` (Welchowski and Schmid, 2019).

Thus, both the total hazards and the relative hazards can be estimated with standard software packages for the binomial and the multinomial distribution. Naturally, this ensures consistency and asymptotic normalcy of $\hat{\boldsymbol{\theta}}$.

3. Application

The main thrust of the proposed model is that it suggests an alternate method for the estimation of the cause-specific hazards, i.e., the estimates for these quantities are now derived indirectly from the total and relative hazard estimates. It should be instructive to compare the performance of the proposed model against other models that propose direct estimation of cause-specific hazards from data. The models that allow for direct estimation of the cause-specific hazards are the multinomial model (Ambrogi et al., 2009; Tutz and Schmid, 2016) and the binomial model (Lee et al., 2018). We have selected the multinomial model for comparison purposes because the model relies on maximisation of the full likelihood function for estimation of relevant parameters as the proposed model. It is straightforward to show that the relevant log-likelihood function for estimation of the cause-specific hazards simultaneously is given by

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{s=1}^{t_i} d_{ijs} \log \lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\alpha}) + (1 - \sum_{j=1}^{J} d_{ijs}) \log(1 - \lambda(s|\boldsymbol{x}_i, \boldsymbol{\alpha})),$$

where $d_{ij} = I(D_i = j)$ such that $d_{ijs} = 0$ for $j = 1, \dots, t_i - 1$ and $d_{ijt_i} = d_{ij}$. The model for the cause-specific hazards is given in (1). This log-likelihood function is a kernel of a multinomial log-likelihood function. This means that γ can be estimated by applying a multinomial distribution to data that have been arranged into a long format. Lee et al. (2018) proposed "collapsing" this log-likelihood by censoring the competing failure times as in Prentice et al. (1978) into J cause-specific log-likelihood functions

$$\mathcal{L}(\boldsymbol{\alpha}_j) = \sum_{i=1}^{n} \sum_{s=1}^{t_i} d_{ijs} \log \lambda_j(s|\boldsymbol{x}_i, \boldsymbol{\alpha}_j) + (1 - d_{ijs}) \log(1 - \lambda(s|\boldsymbol{x}_i, \boldsymbol{\alpha}_j)),$$

where

$$\lambda_j(t|\boldsymbol{x}, \boldsymbol{\alpha}_j) = \frac{\exp(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j})}{1 + \exp(\alpha_{0jt} + \boldsymbol{x}^T \boldsymbol{\alpha}_{1j})}.$$

Naturally, this is not proper because $\mathcal{L}(\boldsymbol{\alpha})$ does not factor into J cause-specific log-likelihood functions $\mathcal{L}(\boldsymbol{\alpha}_j)$, $j = 1, 2 \dots, J$, in discrete time due to the excessive number of ties even though similar estimates are obtained by either method. Furthermore, the $d_{i1s}, d_{i2s}, \dots (1 - \sum_{j=1}^{J} d_{ijs})$ are correlated. This fact must be taken into account when the cause-specific hazard parameters are estimated so as to obtain appropriate standard errors. Towards that end, Lee et al. (2018) used GEE methods which entail maximisation of J quasi log-likelihood functions to determine the parameter estimates and the corresponding standard errors. Note that identical parameter estimates are obtained by applying a straightforward binomial distribution or GEE methods — it therefore suffices to compare the proposed model to the multinomial model.

Table 1. Examples of data structures.

(a)
Original Data

ID	$(\Delta = 1)$	$(\Delta = 2)$	$(\Delta = 0)$	**T**	**X**
1	0	1	0	4	$\boldsymbol{x}_1$
2	0	0	1	3	$\boldsymbol{x}_2$
3	1	0	0	1	$\boldsymbol{x}_3$
4	1	0	0	2	$\boldsymbol{x}_4$
5	0	0	1	1	$\boldsymbol{x}_5$

(b)
Vertical Model: Failure Type Data

ID	$(\Delta = 1)$	**T**	**X**
1	0	4	$\boldsymbol{x}_1$
3	1	1	$\boldsymbol{x}_3$
4	1	2	$\boldsymbol{x}_4$

(c)
Multinomial Data

ID	$(\Delta = 1)$ d_{i1s}	$(\Delta = 2)$ d_{i2s}	$(\Delta = 0)$ $1 - d_{i1s} - d_{i2s}$	**T**	**X**
1	0	0	0	1	$\boldsymbol{x}_1$
1	0	0	0	2	$\boldsymbol{x}_1$
1	0	0	0	3	$\boldsymbol{x}_1$
1	0	1	0	4	$\boldsymbol{x}_1$
2	0	0	0	1	$\boldsymbol{x}_2$
2	0	0	0	2	$\boldsymbol{x}_2$
2	0	0	1	3	$\boldsymbol{x}_2$
3	1	0	0	1	$\boldsymbol{x}_3$
4	0	0	0	1	$\boldsymbol{x}_4$
4	1	0	0	2	$\boldsymbol{x}_4$
5	0	0	1	1	$\boldsymbol{x}_5$

(d)
Vertical Model: Failure Time Data

ID	$(\Delta = 1 \text{ or } 2)$ $d_{is} = d_{i1s} + d_{i2s}$	**T**	**X**
1	0	1	$\boldsymbol{x}_1$
1	0	2	$\boldsymbol{x}_1$
1	0	3	$\boldsymbol{x}_1$
1	1	4	$\boldsymbol{x}_1$
2	0	1	$\boldsymbol{x}_2$
2	0	2	$\boldsymbol{x}_2$
2	0	3	$\boldsymbol{x}_2$
3	1	1	$\boldsymbol{x}_3$
4	0	1	$\boldsymbol{x}_4$
4	1	2	$\boldsymbol{x}_4$
5	0	1	$\boldsymbol{x}_5$

In Table 1, we have created fictitious data to introduce the data structure that is used by the proposed model in contrast to the multinomial model. This table is based on $J = 2$ and 5 subjects where subject 1 failed due to cause 2 at $T = 4$, subject 2 censored at $T = 3$, subject 3 that failed at $T = 1$ due to cause 1, subject 4 that failed at $T = 2$ due to cause 1 and subject 5 that was censored at $T = 1$. Beginning with the familiar multinomial model, in the estimation of the cause-specific hazards simultaneously, a multinomial distribution is fitted to data set (c) with $d_{i1s}, d_{i2s}, (1 - d_{i1s} - d_{i2s})$ as the response vector, where $\boldsymbol{T}$, as a factor, and $\boldsymbol{X}$ are explanatory variables. Usually, $(1 - d_{i1s} - d_{i2s})$ is regarded as a reference category. For the vertical model, to estimate the relative hazards, a binomial distribution is fitted to data set (b) with, say $\Delta = 1$, as the response variable. The covariates for the proposed model are also $\boldsymbol{T}$, as a factor, and $\boldsymbol{X}$. The total hazards are estimated by fitting another binomial distribution to data set (d) with d_{is} as a response variable and again $\boldsymbol{T}$, as a factor, and $\boldsymbol{X}$ are explanatory variables.

The GLM methods are appropriate for application in the estimation of the total hazard parameters because $d_{i1}, \ldots, d_{it_i}$ are conditionally independent in the Markovian sense (Dinse and Larson, 1986). Under the binomial model (Lee et al., 2018), two binomial distributions that account for the correlation between d_{i1s} and d_{i2s} are fitted to data set (c) within the GEE framework, with d_{ijs}, $j = 1, 2$, as a repose variable where $\boldsymbol{T}$, as a factor, and $\boldsymbol{X}$ are explanatory variables.

Fitting a multinomial model is associated with a sizeable number of parameters that require estimation simultaneously. This may give rise to computational difficulties in relation to the stability of parameter estimates. Whilst the same number of parameters as in the multinomial model are estimated when the vertical model is applied, the estimation burden is reduced somewhat because the parameters to be estimated are shared as parameters for total hazards and the relative hazards. Suppose that $\boldsymbol{x}$ is a p-dimensional vector, then a total of $(J \times (q + p))$ parameters are estimated simultaneously when the multinomial model is applied. When the proposed model is applied, $((J - 1)(q + p))$ relative hazard parameters must be estimated simultaneously via the application of a multinomial distribution and the other $(q + p)$ parameters for the total hazards are estimated by applying a binomial distribution. In this respect, the binomial model (Lee et al., 2018) outperforms the proposed model and the multinomial model because the same number of parameters are estimated $(J \times (q + p))$, but $(p \times q)$ parameters for each failure cause are estimated separately.

We apply the proposed model to the freely available unemployment data (*UnempDur*) that come with `Ecdat` (Croissant and Graves, 2020) R package. This discrete time competing risks data set tracks time to failure by exit to full-time or part-time employment for 3343 unemployed subjects. There are 339 subjects that exit to part-time employment, 1073 that exit to full-time employment, 574 with missing failure causes, and 1255 subjects that are censored. There are also 102 subjects with incomplete information. These subjects are excluded from analysis together with the subjects that have missing failure causes, to leave a final sample of size 2667 for analysis. The time T is measured in two-week intervals, and $T \in \{1, 2, \ldots, 27, 28\}$. This has been adjusted because of sparse events beyond $T = 19$ by collapsing $19 < T \leq 28$ into a single interval so that $T \in \{1, 2 \ldots, 19\}$. The covariates are Unemployment Insurance, Disregard Rate, Replacement Rate, Logwage and Tenure in the lost job.

To illustrate the application of the proposed model, we proceed to demonstrate an established fact in econometrics that receipt of unemployment benefits tends to discourage unemployed individuals from seeking out employment opportunities. The two theories that support this view are *labour-supply* and *job search* theories. For illustrative purposes, we have set Unemployment Insurance (`ui`), Disregard Rate (`dr`) and Replacement Rate (`rr`) as explanatory variables. We have centred the continuous variables, i.e., (`dr`) and (`rr`), at their respective averages and set nonrecipients of insurance benefits as the reference category (`ui` = 0).

Accordingly, we have fitted two models, the proposed model (Model I) and the multinomial model (Model II) for comparison purposes. We have also considered the model proposed by Berger et al. (2020) for modelling the cumulative incidence function directly on data. We have referred to this model as Model III. For total hazards in Model I, we have assumed a logistic model

$$\text{Logit}(\lambda(t|\boldsymbol{x}, \boldsymbol{\beta})) = \beta_{0t} + \boldsymbol{x}^T \boldsymbol{\beta}_1,$$

for $t = 1, 2, \ldots, q$. Since $J = 2$, we have modelled the relative hazards via a binomial distribution, that is,

$$\text{Logit}(\pi_1(t|\boldsymbol{x}, \gamma_1)) = \gamma_{01t} + \boldsymbol{x}^T \gamma_1,$$

for $t = 1, \ldots, 19$, where cause 1 is assumed to be exit to full-time employment. We display the results of the analysis in Table 2. Returning to the main exercise of demonstrating the application of

Table 2. Maximum likelihood estimates (with standard errors) for the Discrete Time Vertical Model and the Multinomial Model (* denotes $p < 0.05$).

	Model I *Discrete Time Vertical Model*		**Model II** *Multinomial Model*	
Coefficient	γ	β	α_1	α_2
T1	1.082(0.119)*	1.203(0.060)*	-1.505(0.068)*	-2.572(0.111)*
T2	1.135(0.161)*	-1.512(0.075)*	-1.802(0.084)*	-2.917(0.144)*
T3	1.113(0.202)*	-1.713(0.089)*	-2.002(0.101)*	-3.121(0.175)*
T4	0.844(0.273)*	-2.244(0.122)*	-2.583(0.141)*	-3.506(0.227)*
T5	1.012(0.216)*	-1.441(0.096)*	-1.752(0.109)*	-2.783(0.182)*
T6	1.065(0.378)*	-2.483(0.161)*	-2.778(0.184)*	-3.877(0.324)*
T7	1.151(0.251)*	-1.307(0.109)*	-1.579(0.121)*	-2.776(0.217)*
T8	0.281(0.441)	-2.668(0.211)*	-3.163(0.264)*	-3.603(0.342)*
T9	1.424(0.435)*	-2.031(0.168)*	-2.243(0.183)*	-1.184(0.386)*
T10	-0.032(0.824)	-3.830(0.412)*	-4.548(0.581)*	-4.479(0.583)*
T11	1.524(0.504)*	-2.042(0.189)*	-2.243(0.206)*	-3.825(0.455)*
T12	0.661(0.705)	-3.049(0.323)*	-3.431(0.383)*	-4.215(0.583)*
T13	0.967(0.425)*	-1.702(0.187)*	-2.004(0.211)*	-4.215(0.363)*
T14	1.528(0.461)*	-1.386(0.182)*	-1.592(0.196)*	-3.131(0.418)*
T15	1.731(0.633)*	-1.634(0.228)*	-1.803(0.242)*	-3.586(0.586)*
T16	0.982(0.674)	-1.981(0.291)*	-2.266(0.329)*	-3.401(0.586)*
T17	2.015(1.066)	-2.138(0.347)*	-2.277(0.367)*	-4.298(1.006)*
T18	1.158(0.816)	-1.960(0.349)*	-2.232(0.392)*	-3.423(0.716)*
T19	0.775(0.416)	-0.337(0.214)	-0.702(0.244)*	-1.531(0.359)*
ui	0.165(0.145)	-1.185(0.058)*	-1.156(0.066)*	-1.273(0.114)*
dr	-1.848(0.996)	-1.467(0.455)*	-1.923(0.525)*	-0.529(0.811)
rr	-2.178(0.679)*	-0.344(0.279)	-0.759(0.310)*	0.879(0.522)*

the proposed model where we assess the impact of `ui` in terms of whether it improves the prospects of re-employment or not. Let us first consider the interpretation of $\hat{\gamma}_{01t}$, $\hat{\gamma}_{\mathtt{ui}}$ and $\hat{\beta}_{\mathtt{ui}}$. Beginning with $\hat{\gamma}_{01t}$, let $\mathbf{0} = (0, \bar{x}_{\mathtt{dr}}, \bar{x}_{\mathtt{rr}})$ and $\mathbf{1} = (1, \bar{x}_{\mathtt{dr}}, \bar{x}_{\mathtt{rr}})$ represent the reference and the treatment (`ui`) explanatory vectors, respectively. Then,

$$\frac{\hat{\pi}_1(t|\boldsymbol{x} = \mathbf{0}, \hat{\gamma})}{1 - \hat{\pi}_1(t|\boldsymbol{x} = \mathbf{0}, \hat{\gamma})} = \exp(\hat{\gamma}_{01t}).$$

Since $\hat{\gamma}_{01t} > 0$ for all t except for $t = 10$, it means that *given* that a job has been landed, that job is more likely to be a full-time job than a part-time job except for week 20 for nonrecipients of the unemployment benefits. These odds increase by about 18% ($\exp(\hat{\gamma}_{ui}) = 1.179$) in favour of a full-time job for recipients. This result does not answer our question, it merely compares the ratio of cause-specific hazards conditional on failure for benefit recipients and nonrecipients. Regarding $\hat{\boldsymbol{\beta}}$, since $\hat{\beta}_{\mathtt{ui}} < 0$, it means that holding `rr` and `dr` at their respective average values, the odds of re-employment to full-time or part-time employment are lower for `ui` recipients than

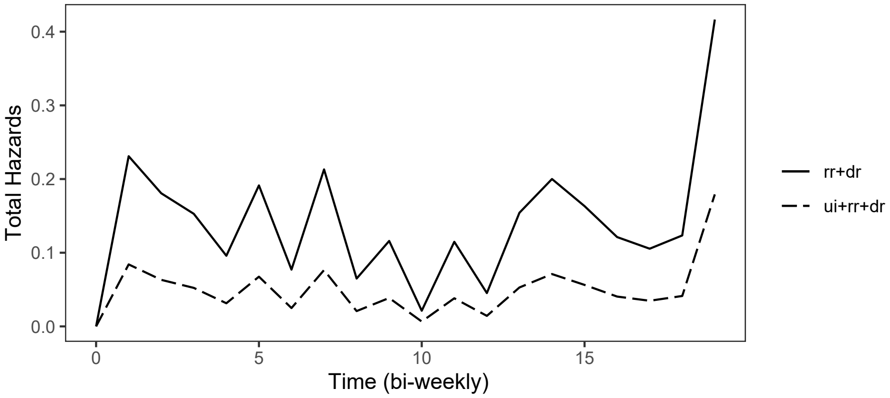

Figure 1. The total hazards of exit to employment with the effect of `ui`.

for non-recipients by about 69% ($1 - \exp(-1.185) = 0.69$). Overall, `ui` recipients tend to search for work less intensively than nonrecipients and as result, fewer individuals land jobs (full-time or part-time) amongst `ui` recipients than amongst nonrecipients. In Figure 1 we have plotted the total hazards and it can be seen that both total hazards decrease until around week 20 ($T = 10$) and then pick up somewhat thereafter. In the US, most of the states provide unemployment benefits for the first 26 weeks. That might be the explanation for the upward movement in the total hazard for the `ui` recipients. A possible explanation for similar movement in the total hazard curve for nonrecipients could be that the recently jobless individuals would be nearing exhausting their reserves and are, therefore, expected to double up their efforts to find employment also around that period. Thus, we were able to determine from $\hat{\beta}_{\mathtt{ui}} < 0$ that unemployment benefits tend to reduce the prospects of re-employment across failure causes (the type of employment). From Model II, the values $\hat{\alpha}_{1\mathtt{ui}} < 0$ and $\hat{\alpha}_{2\mathtt{ui}} < 0$ also indicate that unemployment benefits do not improve the chances of re-employment in agreement with $\hat{\beta}_{\mathtt{ui}} < 0$. Now, suppose that the cause-specific hazard parameters from Model II had opposite signs, say, $\hat{\alpha}_{1\mathtt{ui}} < 0$ and $\hat{\alpha}_{2\mathtt{ui}} > 0$, what then? In a study conducted by McCall (1996) using similar data, the author found that increasing the disregard rate had the effect of reducing full-time employment and increasing part-time employment for recipients of employment benefits. Now, is the gain in part-time employment large enough to offset the reduction in full-time employment such that the total effect is an increase in employment or otherwise? That is where the advantage of the proposed model lies perhaps because it looks at the total effect of a covariate across failure causes. The government might be interested in that information, that is, the total effect of unemployment benefits. Maybe that is also the disadvantage of the proposed model because the government could be interested in the effect of `ui` on full-time employment only as this type of employment is more meaningful and lasting. These are the differences between the proposed model and Model II insofar as the interpretation of the coefficients is concerned. We can supplement these preliminary findings by, for example, examining the effect of `ui` on the cumulative incidence function.

Towards that end, we computed the cumulative incidence function estimates from (2) and (4). As expected, we found that the cumulative incidence function estimates for Model I and Model II were almost identical, and therefore only show Model I estimates in Figure 2. The cumulative incidence

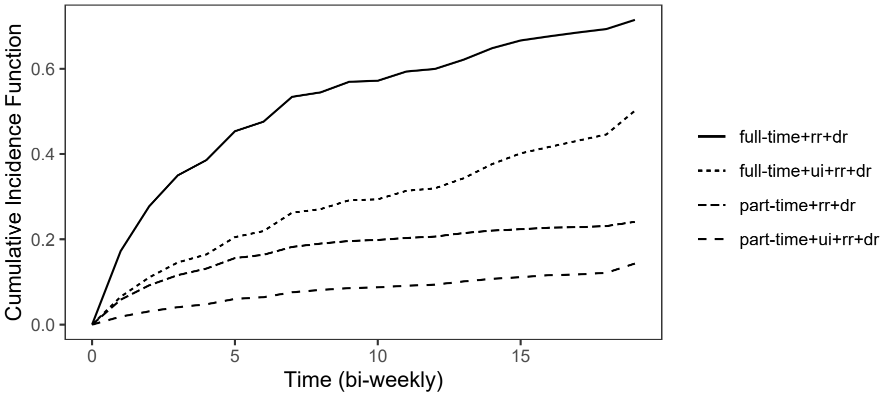

Figure 2. The cumulative incidence function of exit to full-time and part-time employment with the effect of `ui` via the Discrete Time Vertical Model and the Multinomial Model.

function plots from the proposed model and Model II compare favourably; in fact, the plots are almost indistinguishable from each other. Here, we can determine the effect of `ui` on any failure cause individually or the total effect of `ui` across failure causes with ease. Evidently, `ui` has the same effect on full-time and part-time employment, i.e., dampening the prospects of exiting the state of unemployment to full-time or part-time employment, but the effect is more pronounced for full-time employment — a drop from $F_1^{\texttt{rr+dr}}(19) = 71.47\%$ to $F_1^{\texttt{ui+rr+dr}}(19) = 50.18\%$ compared to a drop from $F_2^{\texttt{rr+dr}}(19) = 24.11\%$ to $F_2^{\texttt{ui+rr+dr}}(19) = 14.33\%$. The individuals who find full-time employment for nonrecipients is 74.47% compared to 50.28% for recipients. Likewise, the individuals who land part-time employment is 24.11% amongst nonrecipients and 14.33% amongst recipients. The total effect across the failure causes can be found from $F_1^{\texttt{rr+dr}} + F_2^{\texttt{rr+dr}}$ and $F_1^{\texttt{ui+rr+dr}} + F_2^{\texttt{ui+rr+dr}}$. Thus, 95.58% of unemployed individuals find employment amongst nonrecipients, whereas 64.51% of unemployed individuals find employment amongst the recipients. It is well known that the cumulative incidence function expression that is constructed from cause-specific hazards as given in (2) complicates the evaluation of covariate effects. The expression for the cumulative incidence function derived under the proposed model as given in (4) also does not allow for a one-to-one relationship between the cumulative incidence function and covariates. The expression in (4) is the sum of the terms $\hat{f}_j(s|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \hat{S}(s-1|\boldsymbol{x},\hat{\boldsymbol{\beta}})\lambda(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\pi}_j(s|\boldsymbol{x},\hat{\boldsymbol{\gamma}})$. If we were to consider the effect of $\hat{\beta}_{\texttt{ui}}$ and $\hat{\gamma}_{\texttt{ui}}$ on $\hat{f}_j(s|\boldsymbol{x},\hat{\boldsymbol{\theta}}) = \hat{S}(s-1|\boldsymbol{x},\hat{\boldsymbol{\beta}})\lambda(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})\hat{\pi}_j(s|\boldsymbol{x},\hat{\boldsymbol{\gamma}})$, the values of $\hat{\beta}_{\texttt{ui}}$ and $\hat{\gamma}_{\texttt{ui}}$ imply that moving from nonrecipients to recipients, $\hat{\pi}_j(s|\boldsymbol{x},\hat{\boldsymbol{\gamma}})$ increases, the total hazards $\lambda(s|\boldsymbol{x},\hat{\boldsymbol{\beta}})$ $(s = 1, 2, \ldots, s-1, s, \ldots, q)$ drop which leads to an increase in $\hat{S}(s-1|\boldsymbol{x},\hat{\boldsymbol{\beta}})$ and the total effect is an inexplicable reduction in $\hat{f}_j(s|\boldsymbol{x},\hat{\boldsymbol{\theta}})$. Likewise, the $\hat{\alpha}_{\texttt{jui}}$ in Model II lead to a decrease in $\hat{\lambda}_j(s|\boldsymbol{x},\boldsymbol{\gamma})$ $(s = 1, 2, \ldots, s-1, s)$ and an increase in $\hat{S}(s-1|\boldsymbol{x},\hat{\boldsymbol{\alpha}})$, but which eventually causes a reduction in $\hat{f}_j(s|\boldsymbol{x},\hat{\boldsymbol{\gamma}})$. In continuous time, the complication in relation to the evaluation of covariate effects on the expression given in (2) has lead to the development of other regression models for the cumulative incidence function, such as the *transformation models* (Fine and Gray, 1999; Klein and Anderson, 2005; Scheike and Gerds, 2008), where the cumulative incidence function is directly modelled on covariates. The model proposed by Berger et al. (2020) is an extension of the model advanced by Fine and Gray (1999) to a discrete time model. The model proposes the following expression for the cumulative incidence function

$$F_j(t|\boldsymbol{x}) = 1 - \prod_{s=1}^{t}(1 - h_j(t|\boldsymbol{x}) = 1 - S_j(t|\boldsymbol{x}),$$

where $h_j(t|\boldsymbol{x})$ is referred to as the subdistribution hazard. Note, here, that there is a direct relationship between $\boldsymbol{x}$ and $F_j(t|\boldsymbol{x}))$. For example, if $\boldsymbol{x}$ engenders an increase in $h_j(t|\boldsymbol{x})$, it leads to an increase in $1-S_j(t|\boldsymbol{x}) = F_j(t|\boldsymbol{x})$ as well. The difference between the cause-specific hazards and subdistribution hazards lies in their respective risk sets. The risk set for the cause-specific hazard at time t consists of all the subjects that have survived to time t. On the other hand, for the subdistribution hazard, $h_j(t|\boldsymbol{x})$, the risk set at time t consists of subjects that have not failed due to failure cause j and those that have failed due to other failure causes, i.e.,

$$h_j(t|\boldsymbol{x}) = P(T = t, D = j|(T \geq t) \cup (T \leq t-1, D \neq j), \boldsymbol{x})$$

The subdistribution hazard is also modelled on covariates within the GLM framework:

$$g(h_j(t|\boldsymbol{x}, \phi_j)) = \phi_{0jt} + \boldsymbol{x}^T\phi_{1j},$$

where ϕ_{0jt} is a duration coefficient and ϕ_{1j} is a vector of regression coefficients. The authors proceed to argue that $\phi_j = (\phi_{0jt}, \phi_{1j}^T)^T$, for $j = 1, 2 \ldots J$ can be estimated via an application of a binomial distribution with weights. That is, the solution of $\mathcal{L}(\phi_j)$ can be found by fitting a binomial distribution with δ_{is} as responses and w_{is} as weights such that

$$\mathcal{L}(\phi_j) = \sum_{i=1}^{n}\sum_{s=1}^{q-1} w_{is}\left\{\delta_{is}\log h_j(s|\boldsymbol{x}_i, \phi_j)) + (1-\delta_{is})\log(1 - h_j(s|\boldsymbol{x}_i, \phi_j))\right\},$$

where $\delta_{is} = 0$ for $s = 1, 2, \ldots, t_i - 1$, $\delta_{it_i} = 1$, $\delta_{is} = 0$ for $t_1 < s \leq q-1$ when subject i failed at time t_i at first due to cause j. If the subject was censored at t_i, then $\delta_{is} = 0$ for $s = 1, 2, \ldots, q-1$. For both subjects, $w_{is} = 1$ for $s \leq t_i$ and $w_{is} = 0$ for $s > t_i$. For a subject that experienced first failure by a cause other than cause j, $w_{is} = 1$ for $s \leq t_i$ and $w_{is} = G(s-1)/G(t_i - 1)$ for $t_i \leq s < q-1$, where $G(s)$ is an estimate of the censoring distribution. The preparation of data for the estimation of the subdistribution hazards can easily be conducted via the R package `discSurv` (Welchowski and Schmid, 2019). We have modelled the subdistribution hazards via the complementary log-log link function

$$h_j(t|\boldsymbol{x}, \phi) = 1 - \exp(-\exp(\phi_{0jt} + \boldsymbol{x}^T\phi_{1j})),$$

for $t = 1, 2, \ldots, 19$, and $j = 1, 2$. The results of the analysis are displayed in Table 3. From $\hat{\phi}_{\mathtt{jui}} < 0$ for $j = 1, 2$, it can be inferred that the effect of `ui` is to reduce the prospects of re-employment for both part-time and full-time employment. The cumulative incidence function estimates from the proposed model and the subdistribution hazard model are plotted in Figure 3. Evidently, the proposed model and the subdistribution hazard model lead to the same conclusion. That is, the effect of `ui` is to reduce employment prospects for both part-time and full-time employment.

All these findings are in agreement with the view that unemployment benefits tend to discourage unemployed individuals from searching for work extensively as espoused by theory.

Table 3. Maximum likelihood estimates (with standard errors) for the Discrete Time Subdistribution Hazard Model (* denotes $p < 0.05$).

	Model III *Subdistribution Hazard Model*	
Coefficient	ϕ_1	ϕ_2
T1	-1.541(0.063)*	-2.303(0.114)*
T2	-2.059(0.079)*	-3.397(0.140)*
T3	-2.307(0.096)*	-3.719(0.171)*
T4	-2.902(0.137)*	-4.166(0.223)*
T5	-2.151(0.103)*	-3.561(0.176)*
T6	-3.167(0.179)*	-4.684(0.319)*
T7	-2.059(0.114)*	-3.716(0.210)*
T8	-3.619(0.260)*	-4.558(0.337)*
T9	-2.746(0.178)*	-4.740(0.381)*
T10	-5.018(0.578)*	-5.474(0.579)*
T11	-2.767(0.199)*	-4.890(0.450)*
T12	-3.944(0.379)*	-5.283(0.579)*
T13	-2.586(0.203)*	-4.233(0.357)*
T14	-2.236(0.186)*	-4.389(0.411)*
T15	-2.465(0.232)*	-4.896(0.579)*
T16	-2.935(0.318)*	-4.752(0.579)*
T17	-2.980(0.355)*	-5.711(0.999)*
T18	-2.973(0.379)*	-4.905(0.709)*
T19	-1.703(0.218)*	-3.319(0.337)*
ui	-0.904(0.062)*	-0.853(0.111)*
dr	-1.792(0.500)*	-0.294(0.790)
rr	-0.843(0.292)*	1.083(0.509)*

4. Conclusion

We set out to develop a discrete time vertical competing risks model. This is easily achieved by concentrating on the marginal failure time distribution where the total hazards are modelled on covariates via a link function within the GLM framework. This is a standard approach to locate the failure time distribution in the discrete time realm. The choice of a link function is often determined by the measurement units of time. Some authors, for example, prefer the logit link function when time is intrinsically measured in discrete units or a complementary log-log link function for grouped continuous survival times. The multinomial distribution becomes the most natural choice for modelling the conditional failure type probabilities or relative hazards on covariates. In contrast to continuous time, where tied events are an exception more than a rule, the number of tied events is larger in discrete time, a situation that is most welcome because it improves the stability of parameter estimation insofar as relative hazards are concerned. It was demonstrated that both the relative and

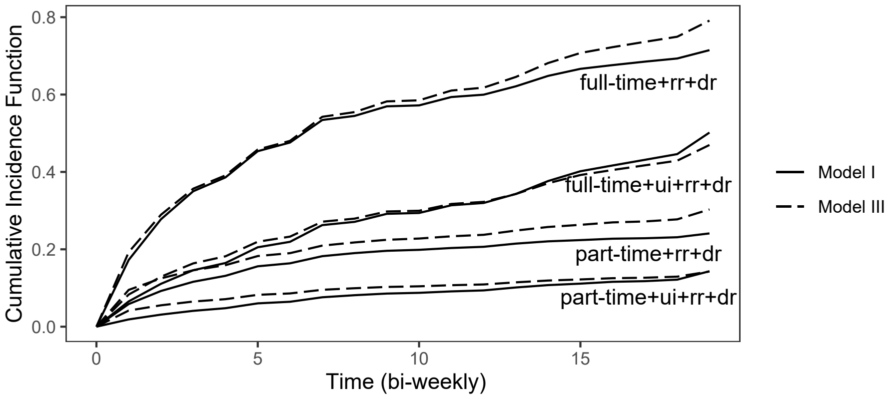

Figure 3. The cumulative incidence function of exit to full-time and part-time employment with the effect of `ui` via the Discrete Time Vertical Model and the Discrete Time Subdistribution Hazard Model.

total hazards can easily be estimated with standard software packages. This model was applied to real discrete time competing risks data together with a multinomial model. It was found that the two models compare favourably. Often, concerns are raised regarding the multinomial model and the sizeable number of parameters that require estimation. The estimation burden is somewhat reduced when the proposed model is applied because the number of parameters to be estimated are shared as relative hazard parameters and total hazards parameters. Otherwise, there is nothing between the models, they merely present different approaches for modelling discrete time competing risks data. In terms of their potential, the proposed model has an edge over the multinomial model. As already mentioned, it has been demonstrated in the continuous time settings that the vertical model can handle missing failure causes without any structural modification. The model has also been upscaled to handle data that come with cured subjects. The proposed model, therefore, presents a possibility that can be further explored in relation to extending the model for handling these data complications. The expression for the cumulative incidence function that is obtained under the multinomial model is well known for complicating the interpretation of covariate effects. The expression for the cumulative incidence function under the proposed model suffers similar shortcomings.

Appendix

In this section, we derive the expression for standard error of the cumulative incidence function estimate via the delta method.

Let $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_1^T, \ldots, \boldsymbol{\eta}_t^T)^T$, where $\boldsymbol{\eta}_s = (\eta_{1s}, \ldots, \eta_{J-1s})^T$ and $\eta_{ks} = \gamma_{0ks} + \boldsymbol{x}^T\gamma_{1k}$. Furthermore, let $\boldsymbol{\zeta}_t = (\zeta_1, \ldots, \zeta_t)^T$ where $\zeta_s = \beta_{0s} + \boldsymbol{x}^T\boldsymbol{\beta}_1$. The expression for the standard error is then given by

$$\begin{aligned}
\mathrm{Var}(F_j(t|\boldsymbol{x}, \boldsymbol{\theta}) &= \begin{bmatrix} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t} \\ \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t} \end{bmatrix}^T \begin{bmatrix} \mathrm{Var}(\boldsymbol{\eta}_t) & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}(\boldsymbol{\zeta}_t) \end{bmatrix} \begin{bmatrix} \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t} \\ \frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t} \end{bmatrix} \\
&= \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t}\right)^T \mathrm{Var}(\boldsymbol{\eta}_t) \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_t}\right) + \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t}\right)^T \mathrm{Var}(\boldsymbol{\zeta}_t) \left(\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_t}\right),
\end{aligned}$$

since $\mathrm{Cov}(\boldsymbol{\eta}_q, \boldsymbol{\zeta}_q) = \mathbf{0}$, because $\partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\eta}_q\partial\boldsymbol{\zeta}_q = 0$; see Yu et al. (2011).

The partial derivatives of $F_j(t|\boldsymbol{x}, \boldsymbol{\theta})$ w.r.t. ζ_s and η_{js} are given by

$$\begin{aligned}
\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_s} &= \lambda(s|\boldsymbol{x},\boldsymbol{\beta})\left\{S(s|\boldsymbol{x},\boldsymbol{\beta})\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma}) - \left(F_j(t|\boldsymbol{x},\boldsymbol{\theta}) - F_j(s|\boldsymbol{x},\boldsymbol{\theta})\right)\right\},\\
\frac{\partial F_j(t)}{\partial \eta_{js}} &= \pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})(1-\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})),\\
\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{ks}} &= -\pi_j(s|\boldsymbol{x},\boldsymbol{\gamma})\pi_k(s|\boldsymbol{x},\boldsymbol{\gamma}), \quad j \neq k.
\end{aligned}$$

The expression given in Ambrogi et al. (2009) for the cumulative incidence function is modified to give

$$\begin{aligned}
\mathrm{V}(\hat{F}_j(t|\boldsymbol{x},\hat{\boldsymbol{\theta}})) = &\sum_{s=1}^{t}\sum_{l=1}^{t}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_s}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \zeta_l}\mathrm{Cov}(\zeta_s,\zeta_l)\\
&+\sum_{j=1}^{J-1}\sum_{k=1}^{J-1}\sum_{s=1}^{t}\sum_{l=1}^{t}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{js}}\frac{\partial F_j(t|\boldsymbol{x},\boldsymbol{\theta})}{\partial \eta_{kl}}\mathrm{Cov}(\eta_{js},\eta_{kl})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.
\end{aligned}$$

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, **13**, 61–98.

Ambrogi, F., Biganzoli, E., and Boracchi, P. (2009). Estimating crude cumulative incidences through multinomial logit regression on discrete cause specific hazard. *Computational Statistics and Data Analysis*, **53**, 2767–2779.

Berger, M., Schmid, M., Schmitz-Valckenberg, Welchowski, T., and Bayermann, S. (2020). Subdistribution hazard models for competing risks in discrete time. *Biostatistics*, **21**, 449–466.

Bertoni, F. and Groh, A. P. (2014). Cross-border investments and venture capital exits in Europe. *Corporate Governance: An International Review*, **22**, 84–99.

Choi, K. C. (2002). *On the mixture models in survival analysis with competing risks and covariates*. Ph.D. thesis, Hong Kong Polytechnic University.

Cleric, R., Giraldo, A., and Meggiolaro, S. (2014). The determinants of academic outcomes in a competing risks approach: Evidence from Italy. *Studies in Higher Education*, **22**, 84–99.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society: Series B*, **34**, 187–220.

Croissant, Y. and Graves, S. (2020). *Ecdat: Data Sets for Econometrics*. R package version 0.3-7. URL: *https://CRAN.R-project.org/package=Ecdat*

Dinse, G. E. and Larson, M. G. (1986). A note on semi-Markov models for partially censored data. *Biometrika*, **73**, 379–386.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Hwang, R. and Chu, C. (2013). Forecasting forward defaults: A simple hazard model with competing risks. *Quantitative Finance*, **14**, 1467–1477.

KLEIN, J. P. AND ANDERSON, P. K. (2005). Regression modelling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, **61**, 223–229.

LARSON, M. G. AND DINSE, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society: Series C*, **34**, 201–211.

LEE, M., FEUER, E. J., AND FINE, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics*, **74**, 1468–1481.

MCCALL, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometric Society*, **64**, 647–682.

NICOLAIE, M. A., TAYLOR, J. M. G., AND LEGRAND, C. (2018). Vertical modeling: Analysis of competing risks data with a cure fraction. *Lifetime Data Analysis*, **25**, 1–25.

NICOLAIE, M. A., VAN HOUWELINGEN, H. C., AND PUTTER, H. (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine*, **29**, 1190–1205.

NICOLAIE, M. A., VAN HOUWELINGEN, H. C., AND PUTTER, H. (2015). Vertical modelling: Analysis of competing risks data with missing causes of failure. *Statistical Methods in Medicine*, **29**, 1190–1205.

PENG, Y. AND TAYLOR, J. M. G. (2014). Cure models. *In* KLEIN, J. P., VAN HOUWELINGEN, H. C., IBRAHIM, J. G., AND SCHEIKE, T. H. (Editors) *Handbook of Survival Analysis*, chapter 6. Chapman and Hall, Boca Raton, Florida, 113–134.

PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V., FLOURNOY, N., FAREWELL, V. T., AND BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

SCHEIKE, T. H. AND GERDS, M. J. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, **1**, 205–220.

SCHMID, M. AND BERGER, M. (2020). Advanced review: Competing risks analysis for discrete time-to-event data. *WIREs Computational Statistics*, **13**.

SINGER, J. D. AND WILLET, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford University Press, New York.

TUTZ, G. AND SCHMID, M. (2016). *Modeling Discrete Time-to-Event Data*. Springer, Cham.

VALLEJOS, C. AND STEEL, M. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society: Series A*, **180**, 613–631.

WELCHOWSKI, T. AND SCHMID, M. (2019). *discSurv: Discrete Time Survival Analysis*. R package version 1.4.1.
URL: *https://CRAN.R-project.org/package=discSurv*

YU, B., TIWARI, R. C., AND FEUER, E. J. (2011). Estimating the personal cure of cancer patients using population-based grouped cancer survival data. *Statistical Methods in Medical Research*, **20**, 261–274.

ZHIPING, T. (2011). *On competing risks data with covariates and long-term survivors*. Ph.D. thesis, Hong Kong Polytechnic University.

Manuscript received 2021-04-16, revised 2021-09-06, accepted 2021-10-29.