South African Statistical Journal
Vol. 54, No. 2, 231–241
https://doi.org/10.37920/sasj.2020.54.2.7

A NONPARAMETRIC VERTICAL MODEL: AN APPLICATION TO DISCRETE TIME COMPETING RISKS DATA WITH MISSING FAILURE CAUSES

Bonginkosi D. Ndlovu[1]
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
e-mail: *bongi@dut.ac.za*

Sileshi F. Melesse
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Temesgen Zewotir
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Discrete time competing risks data continue to arise in social sciences, education etc., where time to failure is usually measured in discrete units. This data may also come with unknown failure causes for some subjects. This occurs against a background of very limited discrete time analysis methods that were developed to handle such data. A number of continuous time missing failure causes models have been proposed over the years. We select one of these continuous time models, the vertical model (Nicolaie et al., 2015), and present it as a nonparametric model that can be applied to discrete time competing risks data with missing failure causes. The proposed model is applied to real data and compared to the MI. It was found that the proposed model compared favorably with the MI method.

Key words: Discrete Time Competing Risks, Missing Failure causes, Nonparametric Vertical Model, Relative hazards, Total hazards.

1. Introduction

Discrete time competing risks data arise in survival analysis experiments when subjects are exposed to multiple risks of failure and the time to failure evolves discretely. Time to failure $\tilde{T}$ is measured in discrete units if $\tilde{T} \in \{1, \ldots, q\}$ for some positive integer q. This occurs when $\tilde{T}$ is inherently discrete or when $\tilde{T}$ is originally continuous but observed failure times have been grouped into intervals. The distinguishing feature of this data is an excessive number of event/censoring ties. In education, for example, students may exit the system either by graduating or withdrawing. There are often a substantial number of graduation ties because students can only graduate at the end of the year or semester. Even though students can withdraw continuously throughout the year or a semester, and in most instances informally, the authorities are able to ascertain a withdrawal if a student

[1] Corresponding author.
MSC2010 subject classifications. 62-07, 62G05.

fails to re-register. Thus students graduate or withdraw in larger numbers at the end of the year or semester, hence the excessive number of ties. Therefore, graduation times are inherently measured in discrete units, whereas withdrawal times can be viewed as having been grouped into semesters or years. Discrete time competing risk data is also summarized via what has become the standard summary quantities for competing risks data, i.e., the cause-specific-hazards and the cumulative incidence functions. Typically, the data is modelled with cause-specific-hazards and, if the objective of a study is to evaluate the effects of certain covariates, the cause-specific-hazards are popularly modelled on these covariates via the multinomial model (Ambrogi et al., 2009; Tutz and Schmid, 2016). The cause-specific-hazards are estimated by applying the multinomial distribution to the data. The cumulative incidence function estimates are thereafter derived from the cause-specific-hazard estimates. When the assessment of covariate effects is not the primary objective of a study, the summary quantities can be estimated nonparametrically (Davis and Lawrance, 1989). There are instances, however, when discrete time data may come with missing failure causes. The discrete time models that we have just described are no longer applicable in the presence of missing failure causes. Analysis of competing risks data with missing failure causes has received minimal attention in discrete time. This is not, however, unique to analysis of data with missing failure causes since, in general, discrete time models, both in univariate and competing risks settings, are less developed in comparison to continuous time models. One of the options that may be considered in discrete time is to apply the existing discrete time models after editing the data, either by excluding the affected cases or treating them as an additional category of failure. Both these approaches are not ideal because they may lead to biased estimates. A model based approach, where the missing failure causes are imputed, see for example, Bakoyannis et al. (2010) and Lee et al. (2014), may also be considered.

Analysis of data with missing failure causes is a topic that has been widely discussed in continuous time, see for example, Dinse (1982); Dewanji (1992); Goetghebeur and Ryan (1990, 1995); Lu and Tsiatis (2001); Lee et al. (2014); and Nicolaie et al. (2015). In response to the limited options in discrete time for handling data that comes with this complication, we consider the vertical competing risks model (Nicolaie et al., 2015) with a view to present it as a nonparametric discrete time model. The vertical model was originally introduced as an ordinary competing risks model in continuous time (Nicolaie et al., 2010). Nicolaie et al. (2015) demonstrated that the ordinary vertical model (Nicolaie et al., 2010), as is, can serve both as an ordinary competing risks model and a missing failure causes model. We therefore wish to extend this flexibility of the vertical model to the discrete time domain in this article.

In the absence of missing failure causes the standard method (Davis and Lawrance, 1989) entails estimating discrete time cause-specific-hazards nonparametrically from observed data. Let $\tilde{T}$ and D denote the time to failure and failure type respectively, where $\tilde{T}$ is in discrete units and $D \in \{1, \ldots, J\}$. Observed discrete time data on the pair $(\tilde{T}; D)$ can be represented by (t_i, Δ_i) in the absence of missing failure causes. Implicit in this data representation is time to censoring, C, also in discrete units, such that $\Delta_i = I(\tilde{T}_i < C_i)D_i$ and $T_i = \tilde{T}_i \wedge C_i$. Consider a definition of a discrete time cause-specific-hazard that is given by

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

where $t \in \{1, \ldots, q\}$ is a set of observed discrete failure times. Let $d_{(js)}$ and n_s denote the number of cause j failures and the number at risk at $T = s$, respectively. The MLE of $h_j(s)$, the cause j discrete time cause-specific-hazard is given by $\hat{h}_j(s) = d_{(js)}/n_s$ for $s = 1, \ldots, q$ and $j = 1, \ldots, J$.

The cumulative incidence function estimates are then derived from

$$\hat{F}_j(t) = \sum_{s:s\leq t} \hat{h}_j(s)\hat{S}(s-1), \tag{1}$$

where $\hat{S}(t) = \prod_{s=1}^{t}(1-\hat{h}(s))$ and $h(t) = \sum_{j=1}^{J} h_j(t)$.

Nicolaie et al. (2010) presented an alternate method for estimating both the cause-specific-hazards and the cumulative incidence functions. The model proposes modelling observed competing risks data with total hazards and conditional failure type probabilities (relative hazards). This development follows from the vertical model assumption which posits that a given joint distribution of failure time and failure type can also be expressed in terms of a distribution for failure type conditional on failure time and a marginal distribution for time to failure, that is

$$P(\tilde{T}; D) = P(\tilde{T})P(D|\tilde{T}). \tag{2}$$

The immediate consequence of (2) is the following re-expression of a given cause-specific-hazard $h_j(t)$ in terms of a total hazard $\lambda(t)$, a measure that summarizes the marginal failure time distribution, and a relative hazard $\pi_j(t)$:

$$h_j(t) = \lambda(t)\pi_j(t).$$

For the purposes of modelling discrete time competing risks data nonparametrically, it is most logical to summarize the marginal failure time distribution with discrete total hazards, that is, $\lambda(t) = P(T = t|T \geq t)$ for $(t = 1, \ldots, q)$.

The definition of relative hazards is given by $\pi_j(t) = P(D = j|T = t)$ for $t = 1, \ldots, q$ and $j = 1, \ldots, J$. The term "relative hazards" originates from

$$\pi_j(t) = \frac{P(T = t, D = j)}{P(T = t)} = \frac{P(T = t, D = j|T \geq t)P(T \geq t)}{\sum_{j=1}^{J} P(T = t, D = j|T \geq t)P(T \geq t)} = \frac{h_j(t)}{h(t)}.$$

Note that $h(t)$ is the conditional probability of failure by any of the known failure causes at time t, whereas the total hazard $\lambda(t)$ quantifies the conditional probability of failure at time t by any cause (known or unknown). The relative hazards $\pi_j(t)$ expresses the probability that a failure is due to cause j given that a failure has occurred. The conditional probability of failure by cause j at time t, that is, the cause-specific-hazard $h_j(t)$ is now determined as the conditional probability of failure at time t by any cause, and this failure is apportioned to cause j with probability $\pi_j(t)$, the proportion of cause j failures at time t. Note that $\lambda(t) \neq h(t)$ unless the missing failure causes are absent. In fact, in the presence of missing failure causes, $\lambda(t) - h(t)$ is the hazard function of the subjects with missing failure causes at time t.

When the vertical model is assumed, the full likelihood function splits according to (2) into a time to failure likelihood function with total hazards as parameters and a likelihood function for relative hazards. The total hazards and the relative hazards are, therefore, estimated separately. The cause-specific-hazards can no longer be estimated directly from observed data when some of the subjects have missing failure causes, but they can be indirectly estimated from $\hat{\lambda}_j(t) = \hat{\lambda}(t)\hat{\pi}_j(t)$ for $t = 1, \ldots, q$ and $j = 1, \ldots, J$. Likewise, the cumulative incidence function estimates are now obtained from

$$\hat{F}_j(t) = \sum_{s:s\leq t} \hat{S}(s-1)\hat{\lambda}(s)\hat{\pi}_j(s).$$

The standard errors for $\hat{F}_j(t)$ are derived in the Appendix.

Up to this point, we have re-formulated the continuous time vertical model as a nonparametric discrete time model by proposing discrete time total hazards as the summary quantities for the failure time distribution. Collect the parameters that define the discrete time vertical model in $\boldsymbol{\theta} = (\boldsymbol{\lambda}^T, \boldsymbol{\pi}^T)^T$, where $\boldsymbol{\lambda} = (\lambda(1), \ldots, \lambda(q))^T$, $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^T, \ldots, \boldsymbol{\pi}_{J-1}^T)$, and $\boldsymbol{\pi}_j = (\pi_j(1), \ldots, \pi_j(q))^T$. In the next section, Section 2, we consider the estimation of $\boldsymbol{\theta}$ in the presence of missing failure causes, which in essence, is the proposed nonparametric vertical model. This is followed by an application of the proposed model to real discrete time competing risks data in Section 3, together with the MI method. We conclude with a discussion in Section 4. We derive the standard errors for the cumulative incidence function estimates in the Appendix.

2. Estimation

We assume two failure causes in this article. The MLE of $\boldsymbol{\theta}$ is determined by maximizing the observed data likelihood function with respect to $\boldsymbol{\theta}$. When data comes with missing failure causes for some subjects, we introduce a missingness indicator R to distinguish the subjects that failed with known failure causes from the subjects that failed with missing failure causes. Let R assume values 0 or 1 according to whether a subject has failed with a missing failure cause or not. In the presence of missing failure causes, observed data now becomes $(t_i, \Delta_i, R_i = 1)$ for subject i that failed with a known failure cause or censored, otherwise $(t_i, R_i = 0)$ when the failure cause is missing. We have assumed that $R = 1$ for a censored subject because the censoring status is always known. We also make the usual assumption about censoring, that is, C is independent of $(\tilde{T}; D)$. Additionally, we assume that missingness does not depend on failure causes. Put formally, $P(R = 0|\Delta, \Delta > 0, T) = P(R = 0|\Delta > 0, T)$, a consequence of MAR (missing at random) (Rubin, 1976). With these assumptions, see, for example, Betancur (2013), it can easily be shown that the observed data log-likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{2} d_{ij} \log P(T_i = t_i, D_i = j) + d_{i*} \log P(T_i = t_i) + d_{i0} \log P(T_i > t_i),$$

where $d_{ij} = I(\Delta_i = D_i = j)$, $d_{i*} = I(\Delta_i > 0)$ and $d_{i0} = I(D_i = 0)$. Assuming the factorization principle that is proposed by the vertical model, the above log-likelihood function can also be written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \sum_{j=1}^{2} d_{ij} \log P(D_i = j | T_i = t_i) + \Big(\sum_{j=1}^{2} d_{ij} + d_{i*} \Big) \log P(T_i = t_i) + d_{i0} \log P(T_i > t_i) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{2} d_{ij} \log P(D_i = j | T_i = t_i) + \sum_{i=1}^{n} \delta_i \log P(T_i = t_i) + (1 - \delta_i) \log P(T_i > t_i) \\
&= \mathcal{L}(\boldsymbol{\pi}) + \mathcal{L}(\boldsymbol{\lambda}),
\end{aligned}$$

where δ_i, the censoring indicator, is written as $\delta_i = \sum_{j=1}^{2} d_{ij} + d_{i*}$. The log-likelihood function $\mathcal{L}(\boldsymbol{\lambda})$ is the standard univariate failure time log-likelihood function. We then transform the censoring indicator δ_i into a duration dependent indicator variable, that is, $\delta_{is} = \sum_{j=1}^{2} d_{ijs} + d_{i*s} = d_{is} + d_{i*s}$, where $d_{ijs} = 0$ when $s \leq t_i - 1$ and $d_{ijt_i} = d_{ij}$, as well as $d_{i*s} = 0$ when $s \leq t_i - 1$ and $d_{i*t_i} = d_{i*}$. It easily follows that $\mathcal{L}(\boldsymbol{\lambda})$ can be re-written as

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{i=1}^{q} \delta_{(s)} \log \lambda(s) + (n_s - \delta_{(s)}) \log(1 - \lambda(s)),$$

where $\delta_{(s)}$ is the total number of failures at time s regardless of whether the failure causes are known or unknown, i.e., $\delta_{(s)} = \sum_{i=1}^{n} d_{ijs} + d_{i*s}$. This yields

$$\hat{\lambda}(s) = \frac{\delta_{(s)}}{n_s},$$

for $s = 1, \ldots, q$ as the MLE for $\lambda(s)$. Thus, when some of the subjects have failed with missing failure causes, these subjects, together with subjects with known failure causes, contribute towards the estimation of total hazards. In the absence of missing failure causes, the missingness indicator variable falls away so that $\delta_{(s)} = d_{(s)} = \sum_{i=1}^{n} d_{ijs}$, and the model serves as an ordinary nonparametric competing risks model. The conditional failure type log-likelihood $\mathcal{L}(\boldsymbol{\pi})$ can be written as

$$\mathcal{L}(\boldsymbol{\pi}) = \sum_{i=1}^{q} d_{(1s)} \log \pi_1(s) + d_{(2s)} \log(1 - \pi_1(s)).$$

The MLE of $\pi_j(s)$ is then given by

$$\hat{\pi}_j(s) = \frac{d_{(js)}}{d_{(s)}},$$

for $s = 1, \ldots, q$ and $j = 1, 2$. The presence or absence of missing failure causes does not affect the estimation of relative hazards, because only the subjects with known failure causes contribute towards their estimation. Clearly, the vertical model can handle competing risks data in the presence or absence of missing failure causes without any structural adjustment to the model. It is the only model within the joint distribution of time to failure and failure time framework that can serve as an ordinary competing risks model as well as a missing failure causes model.

We have demonstrated that this model can accommodate both ordinary competing risks data without missing failure causes as well as data that comes with missing failure causes. We compare the proposed model to the MI method (Bakoyannis et al., 2010), the subject of Section 3.

3. The Multiple Imputation Model

The MI model is also premised on the MAR assumption. In essence, the objective is to impute the failure types for subjects with missing failure causes. This is achieved by modelling the failure type distribution for the subjects with missing failure causes, i.e., $P(\Delta = 1 | R = 0, \Delta > 0, T)$, followed by the estimation of the model so that it can be used to impute the missing failure causes. Under the MAR assumption:

$$\begin{aligned} P(\Delta_i = 1 | R_i = 0, \Delta_i > 0, T_i) &= P(\Delta_i = 1 | R_i = 1, \Delta_i > 0, T_i) \\ &= P(\Delta_i = 1 | \Delta_i > 0, T_i) = \phi(T_i). \end{aligned}$$

It means that the failure type distribution, $\phi(T)$, for the subjects with missing failure causes is the same as the failure type distribution for subjects with complete information regarding failure type because the failure type distribution does not depend on whether the subject's failure type is known or unknown according to the (MAR) assumption. Therefore, $\phi(T)$ can be estimated from the data for subjects with complete information. Given the nature of the response variable Δ, where $\Delta = 1, 2$, a binomial distribution is the most natural model for $\phi(T)$. The steps of the MI model can be summarized as follows:

1. Let $\text{logit}(\phi(T = t)) = \beta_0 + \sum_{s=1}^{q} \beta_{1s}\mathbf{1}_s(t)$, where $\mathbf{1}_s(t) = 1$ when $s = t$ or zero otherwise. Fit this model to complete cases for whom $\Delta > 0$ to estimate $\boldsymbol{\beta} = (\beta_0, \beta_{11}, \ldots, \beta_{1q})^T$.

2. Draw $\boldsymbol{\beta}_m^* = (\beta_{m0}, \beta_{m11}, \ldots, \beta_{m1q})^T$, from $\mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}))$. Use
$$\hat{\phi}_m(T = t) = 1/\exp(\hat{\beta}_{m0} + \hat{\beta}_{m1t})$$
to estimate $\phi_m(T = t)$, then, randomly select 1 or 0 from $\mathcal{B}(1, \hat{\phi}_m(T = t))$, where 1/0 indicates cause 1/2, to impute the failure types for subjects with missing failure causes. Repeat this exercise M times to create M complete data sets.

3. For the m^{th} complete data set, compute $\hat{F}_{mj}(t)$, $j = 1, 2$ from (1). Repeat this exercise M times for each complete data set, and finally compute $\underline{\hat{F}}_j(t)$, $j = 1, 2$, where $\underline{\hat{F}}_j(t) = \sum_{m=1}^{M} \hat{F}_{mj}(t)$. The variance expression, for $\hat{F}_j(t)$, is given in the Appendix.

4. Application

We apply the model to the *Unemployment Duration Data* available under the `R` package `Ecdat` (Croissant, 2015) as "**UnempDur**". The total sample consists of 3343 subjects where 1073 exit to full-time employment, 339 exit to part-time employment, and 1255 are censored, 574 subjects failed with missing failure causes, and 102 were excluded because of incomplete information. Thus, 3241 subjects are considered in the analysis. The time in two-week interval is $t \in \{1, 2, 3, \ldots, 28\}$. For the purposes of analysis we consider $t \in \{1, 2, 3, \ldots, 20\}$ by collapsing the event/censoring time $t \geq 20$ into one interval because there are relatively few events beyond that time point.

We have applied the proposed model together with the MI method where $M = 8$. Under the MI method, each time the missing failure causes have been imputed, an ordinary nonparametric model (Davis and Lawrance, 1989) is applied to the complete data set to obtain the estimates for the cause-specific-hazards. These estimates are then fed into (1) to obtain the cumulative incidence function estimates $\hat{F}_{mj}(t)$ from the m^{th} complete data set. We have displayed the total and relative hazard estimates from the proposed model and the cause-specific hazard estimates from the MI method. The standard errors for the cause-specific-hazard estimates are computed from (3) – (5) when the MI model is assumed.

Recall that the cause-specific-hazards cannot be estimated directly from observed data in the presence of missing failure causes. These quantities can, however, be recovered from this relationship:
$$\hat{h}_j(t) = \hat{\lambda}(t)\hat{\pi}_j(t).$$
For example, at time $T = 9$, the estimate of $h_1(9)$ is $0.825 \times 0.071 = 0.0585$, which compares favorably with $\hat{h}_1(9) = 0.059$ according to the MI method. The reason these methods produce almost identical results is that the missing failure causes are imputed via the $\phi(T)$ which is identical to $\pi_j(T)$, the distribution of relative hazards, as noted by Nicolaie et al. (2015).

Table 1. Maximum likelihood estimates for the total and relative hazards from the Vertical Model as well as the cause-specific-hazards estimates from the MI method (with standard errors).

	Model I (Nonparametric Vertical Model)		**Model II** (MI Method)	
	$\hat{\pi}_1$	$\hat{\lambda}$	$\hat{h}_1$	$\hat{h}_2$
T1	0.752(0.022)	0.154(0.006)	0.116(0.007)	0.038(0.005)
T2	0.761(0.028)	0.129(0.006)	0.098(0.008)	0.031(0.005)
T3	0.763(0.034)	0.113(0.007)	0.085(0.008)	0.028(0.006)
T4	0.727(0.051)	0.057(0.005)	0.042(0.006)	0.015(0.004)
T5	0.748(0.037)	0.121(0.008)	0.091(0.009)	0.029(0.006)
T6	0.762(0.066)	0.045(0.006)	0.035(0.007)	0.011(0.004)
T7	0.779(0.039)	0.129(0.009)	0.101(0.012)	0.029(0.008)
T8	0.625(0.098)	0.039(0.006)	0.024(0.007)	0.015(0.006)
T9	0.825(0.060)	0.071(0.009)	0.059(0.010)	0.012(0.006)
T10	0.060(0.204)	0.020(0.005)	0.011(0.007)	0.009(0.006)
T11	0.838(0.066)	0.061(0.009)	0.052(0.011)	0.009(0.006)
T12	0.700(0.145)	0.029(0.007)	0.020(0.009)	0.009(0.007)
T13	0.756(0.075)	0.091(0.013)	0.068(0.014)	0.024(0.009)
T14	0.833(0.062)	0.112(0.016)	0.093(0.018)	0.019(0.010)
T15	0.864(0.073)	0.079(0.015)	0.069(0.016)	0.010(0.007)
T16	0.769(0.117)	0.093(0.018)	0.071(0.022)	0.022(0.015)
T17	0.889(0.105)	0.065(0.018)	0.058(0.019)	0.007(0.008)
T18	0.778(0.139)	0.065(0.019)	0.051(0.020	0.015(0.012)
T19	0.709(0.081)	0.297(0.038)	0.199(0.049)	0.098(0.041)

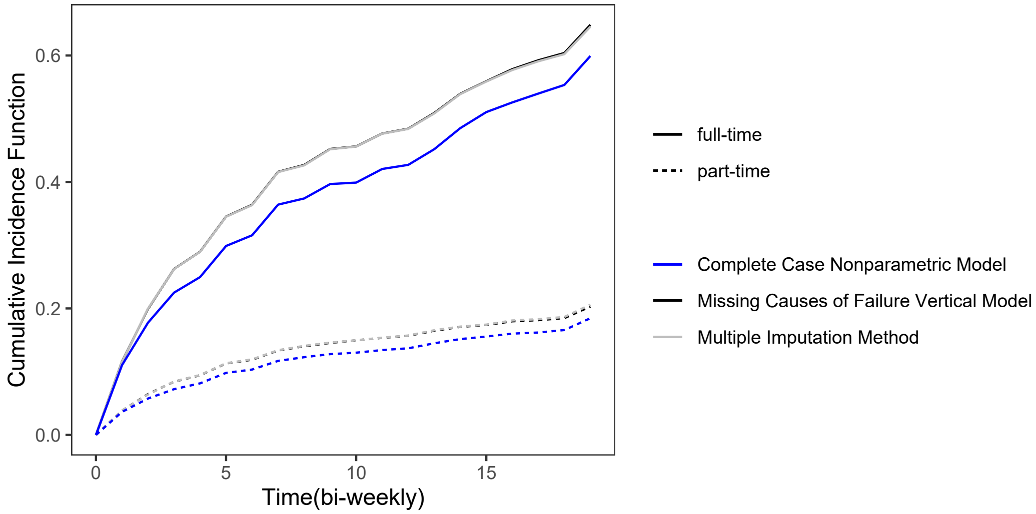

Figure 1. The cumulative incidence function of exit to full-time and part-time employment via the Missing Causes of Failure Vertical Model, the Complete Case Nonparametric Model, as well as the MI method.

We have plotted the cumulative incidence function estimates that were derived via the proposed model (Missing Causes of Failure Vertical Model), the MI method and the Complete Case Nonparametric model in Figure 1. It can be seen that the cumulative incidence function estimates via the proposed model and the MI method are almost identical. Therefore, the interpretation is that at each time point, there are more unemployed subjects that exit to full-time employment than part-time employment whether we consider the proposed model or the MI method. We also estimated the cumulative incidence function with the complete cases vertical model and included it in the plot, and, evidently, the complete cases model understates both the full-time and the part-time cumulative incidence function estimates.

5. Conclusions

We have presented the vertical model Nicolaie et al. (2015) as a nonparametric model with a particular focus on modelling discrete time competing risks data that comes with missing failure causes. This model is premised on the MAR (missing at random) assumption. We compared the proposed model to the MI method (Bakoyannis et al., 2010) which also assumes MAR, and found that the two models compared favorably from the cumulative incidence function plots. One of the advantages of the proposed model is the ease with which the model is applied in contrast to the comparable MI method, where the estimation exercise may require some programming. The advantage of the vertical model over other models within the joint distribution of time to failure and failure type framework, is that the model, as is, serves both as an ordinary competing risks model as well as a model that can handle data with missing failure causes.

Appendix

In this section, we apply the delta method to determine the expression of standard errors for the cumulative incidence function estimates. This expression is given by

$$\begin{aligned} V(\hat{F}_j(t)) &= \sum_{s=1}^{q} \mathrm{Var}\big(S(s-1)\lambda(s)\pi_j(s)\big) \\ &\quad + 2\sum_{s=1}^{q-1}\sum_{k=s+1}^{q} \mathrm{Cov}\big(S(s-1)\lambda(s)\pi_j(s), S(k-1)\lambda(k)\pi_j(k)\big)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \end{aligned}$$

Let

$$Q_s\big(\lambda_1,\ldots,\lambda_{s-1},\lambda_s,\pi_j(s)\big) = S(s{-}1)\lambda(s)\pi_j(s)$$

and

$$Q_k\big(\lambda_1,\ldots,\lambda_{s-1},\lambda_s,\ldots,\lambda_{k-1},\lambda_k,\pi_j(k)\big) = S(k{-}1)\lambda(k)\pi_j(k),$$

where $s < k$. We begin by computing the partial derivatives:

$$\begin{aligned} \frac{\partial Q_s}{\partial\lambda(l)} &= -\frac{S(s{-}1)\lambda(s)\pi_j(s)}{1-\lambda(l)} \quad l = 1,\ldots,s{-}1, \\ \frac{\partial Q_s}{\partial\lambda(s)} &= S(s{-}1)\pi_j(s), \\ \frac{\partial Q_s}{\partial\pi_j(s)} &= S(s{-}1)\lambda(s). \end{aligned}$$

Now,

$$\begin{aligned} &V\big(\hat{S}(s{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\big) \\ &= \sum_{l=1}^{s-1} \frac{\partial Q_s}{\partial\lambda(l)} V\big(\lambda(l)\big)\frac{\partial Q_s}{\partial\lambda(l)} + \frac{\partial Q_s}{\partial\lambda(s)} V\big(\lambda(s)\big)\frac{\partial Q_s}{\partial\lambda(s)} + \frac{\partial Q_s}{\partial\pi_j(s)} V\big(\pi_j(s)\big)\frac{\partial Q_s}{\partial\pi_j(s)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \sum_{l=1}^{s-1}\left(-\frac{Q_s}{1-\lambda_{(l)}}\right)^2 \frac{\lambda(l)(1-\lambda(l))}{n_l} + \left(\frac{Q_s}{\lambda(s)}\right)^2 \frac{\lambda(s)(1-\lambda(s))}{n_s} \\ &\quad + \left(\frac{Q_s}{\pi_j(s)}\right)^2 \frac{\pi_j(s)(1-\pi_j(s))}{d_{(s)}}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ &= \big(\hat{S}(s-1)\hat{\lambda}(s)\hat{\pi}_j(s)\big)^2\left(\sum_{l=1}^{s-1}\frac{\delta_{(l)}}{n_l(n_l-\delta_{(l)})} + \frac{(n_s-\delta_{(s)})}{\delta_{(s)}n_s} + \frac{(d_s-d_{(js)})}{d_{(js)}d_{(s)}}\right) \end{aligned}$$

and

$$
\begin{aligned}
&\operatorname{Cov}\big(\hat{S}(s-1)\hat{\lambda}(s)\hat{\pi}_j(s), \hat{S}(k-1)\hat{\lambda}(k)\hat{\pi}_j(k)\big)\\
&= \sum_{l=1}^{s-1} \frac{\partial Q_s}{\partial \lambda_l} \mathrm{V}\big(\lambda(l)\big) \frac{\partial Q_k}{\partial \lambda_l} + \frac{\partial Q_s}{\partial \lambda_s} \mathrm{V}\big(\lambda(s)\big) \frac{\partial Q_k}{\partial \lambda_s}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\\
&= \sum_{l=1}^{s-1} \left(-\frac{Q_s}{1-\lambda_{(l)}} - \frac{Q_k}{1-\lambda_{(l)}} \right)^2 \frac{\lambda(l)(1-\lambda(l))}{n_l} + \left(-\frac{Q_s}{\lambda(s)} \frac{Q_k}{1-\lambda(s)} \right) \frac{\lambda(s)(1-\lambda(s))}{n_s}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\\
&= \big(\hat{S}(s-1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k-1)\hat{\lambda}(k)\hat{\pi}_j(k)\big) \left(\sum_{l=1}^{s-1} \frac{\delta_{(l)}}{n_l(n_l-\delta_{(l)})} - \frac{1}{n_s} \right).
\end{aligned}
$$

We have assumed that $\operatorname{Cov}(\lambda(l), \pi_j(m)) = 0$, because

$$
\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \lambda(l) \partial \pi_j(m)} = 0
$$

for $j = 1, 2; l = 1, \ldots, q; m = 1, \ldots, q$.

The variance $\hat{V}(\hat{\Gamma})$, where $\hat{\Gamma}$ is estimated via the MI method i.e.,

$$
\hat{\Gamma} = \frac{\sum_{m=1}^{M} \hat{\Gamma}_m}{M},
$$

is a combination of within imputation variance W_V and between imputatation variance B_V, that is

$$
\hat{V}(\hat{\Gamma})) = W_V + (1 + M^{-1}) B_V, \tag{3}
$$

$$
W_V = \frac{1}{M} \sum_{m=1}^{M} \hat{V}(\hat{\Gamma}_m), \tag{4}
$$

$$
B_V = \sum_{m=1}^{M} \frac{(\hat{\Gamma}_m - \hat{\Gamma})(\hat{\Gamma}_m - \hat{\Gamma})^T}{M-1}. \tag{5}
$$

The variance of $\hat{F}_j(t)$ is, therefore,

$$
\hat{V}(\hat{F}_j(t)) = \frac{1}{m} \sum_{m=1}^{M} \hat{V}(\hat{F}_{mj}(t)) + \sum_{m=1}^{M} \frac{(\hat{F}_{mj}(t) - \underline{\hat{F}}_j(t))(\hat{F}_{mj}(t) - \underline{\hat{F}}_j(t))^T}{M-1}.
$$

References

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.

Bakoyannis, G., Siannis, F., and Touloumi, G. (2010). Modelling competing risks data with missing cause of failure. *Statistics in Medicine*, **29**, 3172–3185.

Betancur, M. M. (2013). *Regression Modeling with Missing Outcomes: Competing Risks and Longitudinal Data*. Ph.D. thesis, Université Paris-Sud, Orsay, France.

CROISSANT, Y. (2015). *Ecdat: Data Sets for Econometrics*. R package version 0.2-9. URL: *https://CRAN.R-project.org/package=Ecdat*

DAVIS, T. P. AND LAWRANCE, A. J. (1989). The likelihood for competing risk survival analysis. *Scandinavian Journal of Statistics*, **16**, 23–28.

DEWANJI, A. (1992). A note on a test for competing risks with missing failure. *Biometrika*, **79**, 855–857.

DINSE, G. E. (1982). Nonparametric estimation for partially-complete time and of failure data. *Biometrics*, **38**, 417–431.

GOETGHEBEUR, E. AND RYAN, L. (1990). A modified logrank test for competing risks with missing failure type. *Biometrika*, **77**, 207–211.

GOETGHEBEUR, E. AND RYAN, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, **82**, 821–833.

LEE, M., DINGHAM, J. J., AND HAN, J. (2014). Multiple imputation methods for nonparametric inference on cumulative incidence with missing cause of failure. *Statistics in Medicine*, **33**, 4605–4626.

LU, K. AND TSIATIS, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, **57**, 1191–1197.

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 Medical Research*, **24**, 891–908.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

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

Manuscript received 2019-09-01, revised 2020-03-05, accepted 2020-05-27.