STATISTICS IN TRANSITION new series, June 2023
Vol. 23, No. 3, pp. 61–76, https://doi.org/10.59170/stattrans-2023-036
Received – 25.10.2021; accepted – 05.09.2022

A nonparametric analysis of discrete time competing risks data: a comparison of the cause-specific-hazards approach and the vertical approach

Bonginkosi Duncan Ndlovu[1], **Sileshi Fanta Melesse**[2], **Temesgen Zewotir**[3]

ABSTRACT

Nicolaie et al. (2010) have advanced a vertical model as the latest continuous time competing risks model. The main objective of this article is to re-cast this model as a nonparametric model for analysis of discrete time competing risks data. Davis and Lawrance (1989) have advanced a cause-specific-hazard driven method for summarizing discrete time data nonparametrically. The secondary objective of this article is to compare the proposed model to this model. We pay particular attention to the estimates for the cause-specific-hazards and the cumulative incidence functions as well as their respective standard errors.

Key words: vertical model; total hazards; relative hazards; cause-specific-hazards; cumulative incidence functions.

1. The first section

Competing risks have come to refer to survival analysis experiments where subjects may fail by more than one mode of failure. The vertical model (Nicolaie et al., 2010) is the latest competing risks model that has been advanced additional to the models proposed by Prentice et al.(1978) and Larson and Dinse (1985). The model proposes total hazards and relative hazards for modelling competing risks data. It is the only competing risks model that is capable of handling the standard competing risks data as well as data that has unknown failure causes for some subjects, that is, the model is invariant to the presence or absence of unknown failure causes (Nicolaie et al., 2015). Furthermore, Nicolaie et al. (2018) have extended the model for analysis of data that has a sizable proportion of cured subjects. Both these topics, i.e., handling data with missing failure causes and data that comes with cured subjects, have not received satisfactory attention in discrete time. Whilst the vertical model possesses these attractive features, it cannot be naively applied to discrete time competing risks data because the model was introduced as a continuous time model. The main objective of this article is to modify this model and present it as a nonparametric discrete time competing risks model additional to the nonparametric model suggested by Davis and Lawrance (1989). The complication with discrete time data is excessive number of ties. Continuous time competing risks models are premised on the factorization of

[1]Department of Statistics, Durban University of Technology, South Africa. E-mail: bongi@dut.ac.za
[2]Department of Statistics, University of KwaZulu-Natal, South Africa. E-mail: melesse@ukzn.ac.za
[3]Department of Statistics, University of KwaZulu-Natal, South Africa. E-mail: zewotir@ukzn.ac.za. ORCID: https://orcid.org/0000-0003-1438-8571.

the full likelihood function into cause-specific likelihood functions assumption. This is not possible in the presence of a disproportionately large number of ties. The model suggested by Davis and Lawrance (1989) is one the models that have been advanced specifically for analysis of discrete time competing risks data. In fact, this model is the first truly discrete time competing risks model to be advanced in the competing risks literature. This model was followed by the multinomial model (Ambrogi et al., 2009, Tutz and Schmid, 2016) as the first regression model for analysis of discrete time data. Here, data are modelled with discrete time version of cause-specific-hazards. Concerns have been expressed about this model owing to estimation of a significantly larger number of cause-specific-hazard parameters simultaneously. Recently, Lee et al. (2018) have advanced an alternate regression model, which addresses these reservation regarding the multinomial model (Ambrogi et al., 2009, Tutz and Schmid, 2016) where the cause-specific-hazards are estimated individually via the application of a binomial distribution within the GEE framework. Both these models, that is, the multinomial model (Ambrogi et al.,2009, Tutz and Schmid, 2016) and the binomial model (Lee et al., 2018) give rise to a regression model for the cumulative incidence function, which has become notorious for complicating the assessment of covariate effects. Berger et al. (2020) have since proposed a discrete time subhazard regression model for the cumulative incidence function to address the limitations of the cause-specific-hazard denominated regression model for the cumulative incidence function. The Davis and Lawrance (1989) model proposes nonparametric discrete time cause-specific-hazards for modelling data. We shall refer to this model as the cause-specific-hazards model. If the vertical model proposes total hazards and relative hazards for characterizing data, it implies that the standard summary statistics are now obtained from the total hazard and relative hazard estimates in the place of the more familiar cause-specific-hazard estimates as suggested by the cause-specific-hazards model. The most logical question that follows is whether the two estimation methods produce the same estimates for the same quantity. For example, if one method proposes direct estimation of cause-specific-hazards from data and the other one suggests that the estimates for the same quantities are now derived from total and relative hazard estimates, then how do the two methods compare, and more importantly, how do the standard errors for the two estimation methods compare. As a secondary objective of this article, we attempt to address these questions. With the cause-specific-hazards and the cumulative incidence functions as the most widely quoted pair for summarizing competing risks data, we will focus on these quantities and their respective standard errors to address these pertinent questions.

As already highlighted, Davis and Lawrance (1989) proposes cause-specific-hazards for modelling data. Let $\tilde{T}$ and D represent time to failure and failure type respectively, where $D \in \{1,2,\dots J\}$ and J is the number of failure causes. Also, let C denote the time to censoring. Observed competing risks data can be represented by $y_i = (t_i, D_i)$ for $i = 1,\dots n$ where $T_i = \min\{\tilde{T}_i, C_i\}$, such that $T_i = t_i$ is a failure time due to failure type j or censoring time according to whether $D_i = j$ or 0. It is assumed that time to failure and censoring time are in discrete units, i.e., $\tilde{T}, C \in \{1,\dots q\}$ where q is a positive integer. The definition of discrete time cause-specific-hazards for nonparametric purposes is given by

$$\lambda_j(t) = P(T = t; D = j | T \geq t) \tag{1}$$

for $t = 1,2,\ldots q$ and $j = 1,2,\ldots J$. Suppose that at time t, $n_{(t)}$ and $d_{(jt)}$ denote the number at risks and the number of failures due to failure type j, respectively. Davis and Lawrance (1989) have shown that the observed data likelihood function is a kernel of a multinomial likelihood function:

$$\mathscr{L} = \sum_{s=1}^{q}\sum_{j=1}^{J} d_{(js)} \log\lambda_j(s) + (n_{(s)} - d_{(s)})\log(1-\lambda(s)) \tag{2}$$

where $d_{(s)} = \sum_{j=1}^{J} d_{(js)}$ and $\lambda(s) = \sum_{j=1}^{J} \lambda_j(s)$. As such, the MLE for $\lambda_j(s)$ is given by

$$\hat{\lambda}_j(s) = d_{(js)}/n_{(s)}$$

for $s = 1,2,\ldots q$ and $j = 1,2,\ldots J$. The estimates for the cumulative incidence functions are then given by

$$\hat{F}_j(t) = \sum_{s=1}^{t} \hat{S}(s\text{-}1)\hat{\lambda}_j(s) \tag{3}$$

for $t = 1,2,\ldots q$ and $j = 1,2,\ldots J$, where $\hat{S}(t) = \prod_{s=1}^{t}(1-\hat{\lambda}(s))$.

The vertical model proposes a factorization of the bivariate distribution of failure time and failure type into a marginal distribution for failure time and a distribution for failure type conditional on failure time as characterized via total hazards and failure type probabilities conditional on failure time (relative hazards), respectively. For analysis of discrete time competing risks data nonparametrically, we propose the following definition for discrete time total hazards:

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

for $t = 1,2,\ldots q$. The total hazard $\lambda(t)$ is the probability of failure, by any cause, at time t given survival to time t. On the other hand, the relative hazard $\pi_j(t)$ is the probality that a failure is attributable to cause j given that a failure has occurred at time t. The definition of relative hazards is given by

$$\pi_j(t) = P(D=j|T=t)$$

$t = 1,2,\ldots q$ and for $j = 1,2,\ldots J$. The term "relative hazards" comes from:

$$\begin{aligned}\pi_j(t) = P(D=j|T=t) = \frac{P(D=j, T=t)}{P(T=t)} &= \frac{P(D=j, T=t, T\geq t)/P(T\geq t)}{P(T=t, T\geq t)/P(T\geq t)}\\ &= \frac{\lambda_j(t)}{\sum_{j=1}^{J}\lambda_j(t)}\end{aligned} \tag{4}$$

It follows from (4)that:

$$\lambda_j(t) = \lambda(t)\pi_j(t) \tag{5}$$

Thus, the cause-specific-hazard estimates are now estimated indirectly from total hazard and relative hazard estimates via (5). All failures contribute to the estimation of the total hazards, then, the total hazards are apportioned to cause-specific-hazards via relative hazards. This formulation become very convenient in the presence of subjects with missing failure causes because these subjects also contribute to the estimation of total hazards. The expression for the cumulative incidence function is also given in terms of total and relative hazards:

$$F_j(t) = \sum_{s=1}^{t} S(s\text{-}1)\lambda(s)\pi_j(s)$$

$t = 1,2,\ldots q$ and for $j = 1,2,\ldots J$.

This concludes the exercise of re-framing the vertical model as a discrete time nonparametric competing risks model. To determine the summary statistics, i.e., the estimates for cause-specific-hazards and cumulative incidence functions we require the estimates for total hazards and relative hazards. Let $\theta = (\pi^T, \lambda^T)^T$ where $\lambda = (\lambda(1), \lambda(2) \ldots \lambda(q))^T, \pi = (\pi_1^T, \pi_2^T \ldots \pi_{J\text{-}1}^T)^T$, and $\pi_j = (\pi_j(1), \pi_j(2) \ldots \pi_j(q))^T$. In Section 2 we demonstrate the estimation of total hazards and relative hazards, that is, we determine $\hat{\theta}$. This is followed by the application of the proposed model in Section 3. We derive the standard errors for the cumulative incidence function estimates in Section 4. This concludes the first part of our twofold objectives. In Section 5, we address the second part of our objective, that is, to prove that the estimates for cause-specific-hazards and cumulative incidence function as well as the corresponding standard errors are identical by the proposed model or the cause-specific-hazards model. We conclude the article with a discussion in Section 6.

2. Estimation

It is straightforward to determine the MLE's for the total hazards and relative hazards. The observed data likelihood function which is specified in terms of total hazards and relative hazards is differentiated with respect to these quantities. When the vertical model is assumed, $P(T_i = t_i, D_i = j)$, the contribution of subject i that failed at time t_i due to failure cause j to the observed data likelihood function is now replaced by $P(D_i = j|T_i = t_i)P(T_i = t_i)$ while a censored subject i continues to contribute $P(T_i > t_i)$. Define an indicator variable d_{ij} such that d_{ij} is 1 or 0 according to whether subject i failed by cause j or not and let $d_i = \sum_{j=1}^{J} d_{ij}$ where d_i indicates failure by any cause for subject i. The observed data log-likelihood function can be written as:

$$\begin{aligned}
\mathscr{L}(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log P(D_i = j|T_i = t_i)P(T_i = t_i) + (1-d_i)\log P(T_i > t_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_j(t_i) + \sum_{i=1}^{n} d_i \log P(T_i = t_i) + (1-d_i)\log P(T_i > t_i) \\
&= \mathscr{L}(\pi) + \mathscr{L}(\lambda)
\end{aligned}$$

We can ignore $\mathscr{L}(\pi)$ because the estimates for the relative hazards can be obtained from (4), that is:

$$\hat{\pi}_j(t) = \frac{\hat{\lambda}_j(t)}{\sum_{j=1}^{J} \hat{\lambda}_j(t)} = \frac{d_{(jt)}/n_{(t)}}{\sum_{j=1}^{J} d_{(jt)}/n_{(t)}} = \frac{d_{(jt)}}{d_{(t)}}$$

The log-likelihood function $\mathscr{L}(\lambda)$ is a failure time log-likelihood function. It is straightforward to show that $\mathscr{L}(\lambda)$ can be written as:

$$\mathscr{L}(\lambda) = \sum_{s=1}^{q} d_{(s)} \log \lambda(s) + (n_{(s)} - d_{(s)}) \log(1 - \lambda(s))$$

Naturally, $\frac{\partial \mathscr{L}(\lambda}{\partial \lambda(s)} = 0$ yields an MLE for $\lambda_j(s)$ given by

$$\hat{\lambda}(s) = \frac{d_{(s)}}{n_{(s)}}$$

The estimates for total hazards and relative hazards can be plugged into appropriate equations to recover the estimates for cause-specific-hazards and cumulative incidence functions. In the next section we demonstrate the application of the proposed model.

3. Application

We apply the proposed model to data that comes with Ecdat R package (Croissant and Graves, 2020). In these data 3343 recently unemployed individuals are tracked the moment they lost their jobs until they are re-employed into part-time employment (339), full-time employment (1073) or are censored (1255). Of the remaining 676, 574 were re-employed but the type employment was not recorded. It is not clear with the other 102 subjects if they were censored or were re-employed. We have excluded the 674 individuals to leave a final sample of 2667 that were considered for analysis. Failure times assume values in $\{1, 2, \ldots 26, 27, 28\}$ where time is measured in bi-weekly units. There are some covariates that come with data such as unemployment benefits, disregard rate, replacement rate, etc., which are naturally ignored.

In the application of the proposed model we have computed the relative hazard and total hazard estimates, respectively from:

$$\hat{\pi}_j(t) = \frac{d_{(jt)}}{d_{(t)}} \text{ and } \hat{\lambda}(t) = \frac{d_{(t)}}{n_{(t)}}$$

The variances are respectively given by

$$\mathrm{V}(\hat{\pi}_j(t)) = \frac{\hat{\pi}_j(t)(1 - \hat{\pi}_j(t))}{d_{(t)}} \text{ and } \mathrm{V}(\hat{\lambda}(t)) = \frac{\hat{\lambda}(t)(1 - \hat{\lambda}(t))}{n_{(t)}}$$

These estimates are listed in Table 1 together with corresponding standard errors. We have labelled full-time re-employment as cause 1 and part-time re-employment as cause 2.

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 Cause-Specific-Hazards Model (with standard errors)

	Model I Nonparametric Vertical Model		**Model II** Cause-Specific-Hazards Model	
	$\hat{\pi}_1$	$\hat{\lambda}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
T1	0.752(0.022)	0.147(0.007)	0.110(0.006)	0.037(0.004)
T2	0.761(0.028)	0.104(0.006)	0.079(0.006)	0.025(0.003)
T3	0.763(0.034)	0.081(0.006)	0.062(0.006)	0.019(0.003)
T4	0.727(0.051)	0.048(0.005)	0.035(0.005)	0.013(0.003)
T5	0.748(0.037)	0.098(0.008)	0.074(0.006)	0.025(0.004)
T6	0.762(0.066)	0.037(0.006)	0.028(0.005)	0.009(0.003)
T7	0.779(0.039)	0.107(0.009)	0.083(0.009)	0.024(0.005)
T8	0.625(0.098)	0.029(0.006)	0.019(0.005)	0.011(0.004)
T9	0.825(0.060)	0.055(0.008)	0.045(0.008)	0.001(0.004)
T10	0.060(0.204)	0.009(0.004)	0.005(0.003)	0.005(0.003)
T11	0.838(0.066)	0.054(0.009)	0.046(0.009)	0.009(0.004)
T12	0.700(0.145)	0.021(0.007)	0.015(0.005)	0.006(0.004)
T13	0.756(0.075)	0.075(0.013)	0.057(0.011)	0.018(0.006)
T14	0.833(0.062)	0.099(0.016)	0.083(0.014)	0.017(0.007)
T15	0.864(0.073)	0.081(0.017)	0.069(0.015)	0.011(0.006)
T16	0.769(0.117)	0.059(0.016)	0.046(0.014)	0.014(0.008)
T17	0.889(0.105)	0.050(0.016)	0.044(0.015)	0.006(0.006)
T18	0.778(0.139)	0.059(0.019)	0.046(0.017)	0.013(0.009)
T19	0.667(0.192)	0.044(0.018)	0.029(0.015)	0.015(0.010)
T20	1.000(0.000)	0.025(0.014)	0.025(0.014)	0.000(0.000)
T21	0.571(0.187)	0.071(0.026)	0.041(0.019)	0.030(0.017)
T22	0.800(0.179)	0.067(0.029)	0.053(0.026)	0.013(0.013)
T23	0.000(0.000)	0.016(0.016)	0.000(0.000)	0.016(0.016)
T24	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
T25	0.000(0.000)	0.019(0.018)	0.000(0.000)	0.019(0.018)
T26	1.000(0.000)	0.045(0.031)	0.045(0.031)	0.000(0.000)
T27	0.833(0.152)	2.000(0.073)	0.167(0.068)	0.033(0.033)

Since $\hat{\pi}_1(t)+\hat{\pi}_2(t)=1$, we have only listed $\hat{\pi}_1(t)$. We have also listed the cumulative incidence function estimates together with corresponding standard errors in Table 2.

The cumulative incidence function estimates are obtained from:

$$\hat{F}_j(t)=\sum_{s=1}^{t}\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)$$

Table 2: Maximum likelihood estimates for the Cumulative Incidence Function from the Vertical Model (with standard errors)

Nonparametric Vertical Model		
	$\hat{F}_1$	$\hat{F}_2$
T1	0.110(0.006)	0.036(0.004)
T2	0.178(0.007)	0.058(0.005)
T3	0.225(0.008)	0.072(0.005)
T4	0.249(0.009)	0.082(0.005)
T5	0.299(0.009)	0.098(0.006)
T6	0.316(0.009)	0.103(0.006)
T7	0.364(0.010)	0.117(0.007)
T8	0.374(0.010)	0.123(0.007)
T9	0.397(0.011)	0.128(0.007)
T10	0.399(0.011)	0.130(0.007)
T11	0.421(0.011)	0.134(0.007)
T12	0.427(0.011)	0.137(0.008)
T13	0.452(0.012)	0.145(0.008)
T14	0.485(0.012)	0.152(0.008)
T15	0.511(0.013)	0.156(0.009)
T16	0.526(0.013)	0.162(0.009)
T17	0.539(0.014)	0.162(0.009)
T18	0.553(0.014)	0.166(0.009)
T19	0.562(0.015)	0.169(0.009)
T20	0.569(0.015)	0.169(0.009)
T21	0.579(0.016)	0.178(0.011)
T22	0.592(0.016)	0.181(0.011)
T23	0.592(0.016)	0.185(0.012)
T24	0.000(0.000)	0.000(0.000)
T25	0.592(0.016)	0.189(0.012)
T26	0.602(0.017)	0.189(0.012)
T27	0.637(0.021)	0.196(0.014)

The standard errors for the cumulative incidence function estimates as derived in the next section are given by

$$\mathrm{V}(\hat{F}_j(t)) = \sum_{s=1}^{t} \mathrm{Var}\big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\big) + 2\sum_{s=1}^{t-1}\sum_{k=s+1}^{t} \mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)\big)$$

where:

$$\mathrm{V}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)) = \big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\hat{\pi}_j(s)\big)^2 \left(\sum_{l=1}^{s-1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} + \frac{n_{(s)} - d_{(s)}}{d_{(s)} n_{(s)}} + \frac{d_{(s)} - d_{(js)}}{d_{(s)} d_{(js)}} \right)$$

and,

$$\mathrm{Cov}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)) = \big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)\big) \times \left(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} - \frac{1}{n_{(s)}}\right)$$

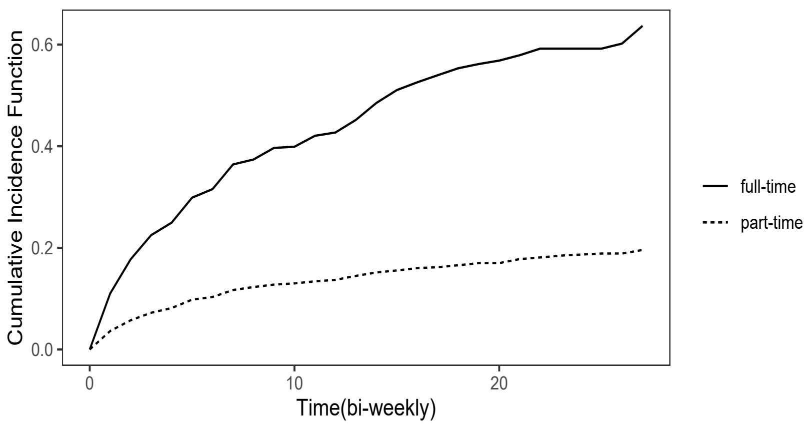

Figure 1: The cumulative incidence function of exit to full-time and part-time

In Figure 1 we have plotted the cumulative incidence function estimates from the proposed model because the proposed model and the cause-specific-hazards model produce identical estimates for cumulative incidence function as will be shown in Section 5. Clearly, the plot suggest that unemployed subjects are more likely to exit the state of unemployment to full-time employment than to part-time employment.

4. Cumulative Incidence Function Standard Errors

The expression of standard errors for the cumulative incidence function estimates under the vertical model is given by

$$\mathrm{V}(\hat{F}_j(t)) = \sum_{s=1}^{t} \mathrm{Var}\big(S(s\text{-}1)\lambda(s)\pi_j(s)\big) + 2\sum_{s=1}^{t\text{-}1}\sum_{k=s+1}^{t} \mathrm{Cov}\big(S(s\text{-}1)\lambda(s)\pi_j(s), S(k\text{-}1)\lambda(k)\pi_j(k)\big)\Big|_{\theta=\hat{\theta}}$$

Let $Q_s\big(\lambda_1,\dots\lambda_{s\text{-}1},\lambda_s,\pi_j(s)\big) = S(s\text{-}1)\lambda(s)\pi_j(s)$
and $Q_k\big(\lambda_1,\dots\lambda_{s\text{-}1},\lambda_s\dots\lambda_{k\text{-}1},\lambda_k,\pi_j(k)\big) = S(k\text{-}1)\lambda(k)\pi_j(k)$, where $s < k$. We begin by com-

puting the partial derivatives:

$$\frac{\partial Q_s}{\partial S(s\text{-}1)} = \lambda(s)\pi_j(s)$$
$$\frac{\partial Q_s}{\partial \lambda(s)} = S(s\text{-}1)\pi_j(s)$$
$$\frac{\partial Q_s}{\partial \pi_j(s)} = S(s\text{-}1)\lambda(s)$$
$$\frac{\partial Q_k}{\partial S(s\text{-}1)} = \frac{Q_k}{S(s\text{-}1)}$$
$$\frac{\partial Q_k}{\partial \lambda(s)} = -\frac{Q_k}{1-\lambda(s)}$$

Assuming that $d_{(1)}, d_{(2)} \dots d_{(s)}$ are uncorrelated (Dinse and Larson, 1986), then $\mathrm{Cov}(\lambda(l), \lambda(m)) = 0$ when $l \neq m$ for $l = 1,2\dots q$ and $m = 1,2\dots q$. It, therefore, follows that:

$$\begin{aligned}
\mathrm{V}\big(Q_s\big) &= \begin{pmatrix} \lambda(s)\pi_j(s) \\ S(s\text{-}1)\pi_j(s) \\ S(s\text{-}1)\lambda(s) \end{pmatrix} \begin{pmatrix} \mathrm{V}(S(s\text{-}1)) & 0 & 0 \\ 0 & \mathrm{V}(\lambda(s)) & 0 \\ 0 & 0 & \mathrm{V}(\pi_j(s)) \end{pmatrix} \begin{pmatrix} \lambda(s)\pi_j(s) \\ S(s\text{-}1)\pi_j(s) \\ S(s\text{-}1)\lambda(s) \end{pmatrix} \\
&= \big(\lambda(s)\pi_j(s)\big)^2 \mathrm{Var}(S(s\text{-}1)) + (S(s\text{-}1)\pi_j(s))^2 \mathrm{V}(\lambda(s)) \\
&+ (S(s\text{-}1)\lambda(s))^2 \mathrm{V}(\pi_j(s)) \\
&= \big(S(s\text{-}1)\lambda(s)\pi_j(s)\big)^2 \sum_{l=1}^{s\text{-}1} \frac{\lambda(l)}{n_{(l)}(1-\lambda(l))} + \frac{1-\lambda(s)}{\lambda(s)n_{(s)}} + \frac{1-\pi_j(s)}{d_{(s)}\pi_j(s)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{V}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)) &= \big(S(s\text{-}1)\lambda(s)\pi_j(s)\big)^2 \sum_{l=1}^{s\text{-}1} \frac{\lambda(l)}{n_{(l)}(1-\lambda(l))} + \frac{1-\lambda(s)}{\lambda(s)n_{(s)}} \\
&+ \frac{1-\pi_j(s)}{d_{(s)}\pi_j(s)}\Bigg|_{\theta=\hat{\theta}} \\
&= \big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\hat{\pi}_j(s)\big)^2 \Bigg(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)}-d_{(l)})} \\
&+ \frac{n_{(s)}-d_{(s)}}{d_{(s)}n_{(s)}} + \frac{d_{(s)}-d_{(js)}}{d_{(s)}d_{(js)}} \Bigg)
\end{aligned}$$

We now consider:

$$\begin{aligned}\operatorname{Cov}\big(Q_s,Q_k\big) &= (S(s\text{-}1))^2\lambda(s)\pi_j(s)\frac{S(k\text{-}1)\lambda(k)\pi_j(k)}{S(s\text{-}1)}\sum_{l=1}^{s\text{-}1}\frac{\lambda(l)(1-\lambda(l))}{n_{(l)}}\\ &\quad - S(s\text{-}1)\pi_j(s)\frac{S(k\text{-}1)\lambda(k)\pi_j(k)}{1-\lambda(s)}\frac{\lambda(s)(1-\lambda(s))}{n_{(s)}}\\ &= S(s\text{-}1)\lambda(s)\pi_j(s)S(k\text{-}1)\lambda(k)\pi_j(k)\left(\sum_{l=1}^{s\text{-}1}\frac{\lambda(l)(1-\lambda(l))}{n_{(l)}}-\frac{1}{n_{(s)}}\right)\end{aligned}$$

Thus,

$$\begin{aligned}\operatorname{Cov}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)) &= \operatorname{Cov}\big(Q_s,Q_k\big)\big|_{\theta=\hat{\theta}}\\ &= \big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)\big)\\ &\quad\times\left(\sum_{l=1}^{s\text{-}1}\frac{d_{(l)}}{n_{(l)}(n_{(l)}-d_{(l)})}-\frac{1}{n_{(s)}}\right)\end{aligned}$$

5. Proofs

In this section we demonstrate that estimates for cause-specific-hazards and cumulative incidence function together with corresponding standard errors derived from the proposed model and the cause-specific-hazards model are identical. Let $\hat{\lambda}_j^V(t)$ and $\hat{\lambda}_j^C(t)$ denote the estimates for the cause-specific-hazards via the proposed model and the cause-specific-hazards model, respectively. Likewise, let $\hat{F}_j^V(t)$ and $\hat{F}_j^C(t)$ represent the estimates for the cumulative incidence function that are produced by the proposed model and the cause-specific-hazards model, respectively.

Beginning with the estimates for the cause-specific-hazards:

$$\hat{\lambda}_j^V(t)=\hat{\pi}_j(t)\hat{\lambda}(t)=\frac{d_{(jt)}}{d_{(t)}}\times\frac{d_{(t)}}{n_{(t)}}=\frac{d_{(jt)}}{n_{(t)}}=\hat{\lambda}_j^C(t)$$

It follows that the cumulative incidence function estimates by both models are identical, that is:

$$\hat{F}_j^V(t)=\sum_{s=1}^{t}\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)=\sum_{s=1}^{t}\hat{S}(s\text{-}1)\hat{\lambda}_j^C(s)=\hat{F}_j^C(t)$$

To determine the standard errors for cause-specific-hazard and cumulative incidence function estimates, we apply the delta method. We begin with standard errors for the cause-specific-hazard estimates. The standard error for $\hat{\lambda}_j^C(s) = \frac{d_{(s)}}{n_{(s)}}$ is well known and it is given by

$$\mathrm{V}(\hat{\lambda}_j^C(s)) = \frac{\hat{\lambda}_j(s)(1-\hat{\lambda}_j(s))}{n_{(s)}}$$

We now determine the expression for the variance of $\hat{\lambda}_j^V(s) = \hat{\lambda}(s)\hat{\pi}_j(s)$. Since

$$\frac{\partial \mathscr{L}}{\partial \lambda(l) \partial \pi_j(m)} = 0$$

for $l = 1, \ldots q; j = 1, 2, \ldots J; m = 1, \ldots q$, thus:

$$\begin{aligned}
\mathrm{V}(\hat{\lambda}_j^V(s)) &= \left(\frac{\partial \lambda_j^V(s)}{\partial \lambda(s)}, \frac{\partial \lambda_j^V(s)}{\partial \pi_j(s)} \right) \begin{pmatrix} \mathrm{V}(\lambda(s)) & 0 \\ 0 & \mathrm{V}(\pi_j(s)) \end{pmatrix} \\
&\times \left(\frac{\partial \lambda_j^V(s)}{\partial \lambda(s)}, \frac{\partial \lambda_j^V(s)}{\partial \pi_j(s)} \right)^T \Bigg|_{\theta=\hat{\theta}} \\
&= \pi_j(s)^2 \mathrm{V}(\lambda(s)) + \lambda(s)^2 \mathrm{V}(\pi_j(s)) \Big|_{\theta=\hat{\theta}} \\
&= \hat{\pi}_j(s)^2 \mathrm{V}(\hat{\lambda}(s)) + \hat{\lambda}(s)^2 \mathrm{V}(\hat{\pi}_j(s))
\end{aligned}$$

where the partial derivatives are given by

$$\begin{aligned}
\frac{\partial \lambda_j^V(s)}{\partial \lambda(s)} &= \pi_j(s) \\
\frac{\partial \lambda_j^V(s)}{\partial \pi_j(s)} &= \lambda(s)
\end{aligned}$$

Therefore, the expression for $\mathrm{V}(\hat{\lambda}_j^V(s))$ is given by

$$\begin{aligned}
\mathrm{V}(\hat{\lambda}_j^V(s)) &= \hat{\pi}_j(s)^2\mathrm{V}(\hat{\lambda}(s)) + \hat{\lambda}(s)^2\mathrm{V}(\hat{\pi}_j(s)) \\
&= \hat{\pi}_j(s)^2\frac{\hat{\lambda}(s)(1-\hat{\lambda}(s))}{n_{(s)}} + \hat{\lambda}(s)^2\frac{\hat{\pi}_j(s)(1-\hat{\pi}_j(s))}{d_{(s)}} \\
&= \hat{\pi}_j(s)\hat{\lambda}(s)\left(\frac{\hat{\pi}_j(s)(1-\hat{\lambda}(s))}{n_{(s)}} + \frac{\hat{\lambda}(s)(1-\hat{\pi}_j(s))}{d_{(s)}}\right) \\
&= \hat{\lambda}_j(s)\left(\frac{d_{(s)}\hat{\pi}_j(t) - d_{(s)}\hat{\pi}_j(t)\hat{\lambda}(s) + n_{(s)}\hat{\lambda}(s) - n_{(s)}\hat{\lambda}(s)\hat{\pi}_j(s)}{n_{(s)}d_{(s)}}\right) \\
&= \hat{\lambda}_j(s)\left(\frac{d_{(s)}\hat{\pi}_j(t) - d_{(js)}\hat{\lambda}(s) + n_{(s)}\hat{\lambda}(s) - d_{(s)}\hat{\pi}_j(s)}{n_{(s)}d_{(s)}}\right) \\
&= \hat{\lambda}_j(s)\frac{n_{(s)} - d_{(js)}}{n_{(s)}}\frac{\hat{\lambda}(s)}{d_{(s)}} = \frac{\hat{\lambda}_j(s)(1-\hat{\lambda}_j(s))}{n_{(s)}} \\
&= \mathrm{V}(\hat{\lambda}_j^C(s))
\end{aligned}$$

Gaynor et al. (1993) showed in continuous time when competing risks data are analyzed nonparametrically, that the full log-likelihood function is a kernel of a multinomial log-likelihood function as in (2), where the continuous time cause-specific-hazards are approximated with discrete time cause-specific-hazards $\lambda_j(t)$ at failure times. Therefore, the expression for $\mathrm{V}(\hat{F}_j^C(t))$ that is derived for continuous time competing risks data equally applies in discrete time:

$$\mathrm{V}(\hat{F}_j^S(t)) = \sum_{s=1}^{t}\mathrm{Var}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big) + 2\sum_{s=1}^{t\text{-}1}\sum_{k=s+1}^{t}\mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}_j(k)\big)$$

where,

$$\begin{aligned}
\mathrm{Var}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big) &= \mathrm{Var}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big) \\
&= \big(\hat{\lambda}_j(s)\hat{S}(s\text{-}1)\big)^2\left(\frac{n_{(s)} - d_{(js)}}{d_{(js)}n_{(s)}} + \sum_{l=1}^{s\text{-}1}\frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})}\right) \qquad (6)
\end{aligned}$$

and,

$$\begin{aligned}
\mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}_j(k)\big) &= \mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}_j(k)\big) \\
&= \big(\hat{\lambda}_j(s)\hat{S}(s\text{-}1)\hat{\lambda}_j(k)\hat{S}(k\text{-}1)\big) \\
&\quad \times\left(-\frac{1}{n_{(s)}} + \sum_{l=1}^{s\text{-}1}\frac{d_{(l)}}{n_{(n)}(n_{(l)} - d_{(l)})}\right) \qquad (7)
\end{aligned}$$

To show that $\mathrm{V}(\hat{F}_j^C) = \mathrm{V}(\hat{F}_j^V)$, we need to demonstrate that:

$$\mathrm{V}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big) = \mathrm{V}\big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\big)$$

and,

$$\mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}_j(k)\big) = \mathrm{Cov}\big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s), \hat{S}(k\text{-}1)\hat{\lambda}_j(k)\big)$$

Now,

$$\mathrm{V}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)) = \big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\hat{\pi}_j(s)\big)^2 \Bigg(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} + \frac{n_{(s)} - d_{(s)}}{d_{(s)} n_{(s)}} + \frac{d_{(s)} - d_{(js)}}{d_{(s)} d_{(js)}} \Bigg) \tag{8}$$

Note that:

$$\frac{(n_{(s)} - d_{(s)})}{d_{(s)} n_{(s)}} + \frac{(d_{(s)} - d_{(js)})}{d_{(js)} d_{(s)}} = \frac{d_{(js)} n_s - d_{(s)} d_{(js)} + n_{(s)} d_{(s)} - n_{(s)} d_{(js)}}{d_{(s)} n_{(s)} d_{(js)}} = \frac{n_{(s)} - d_{(js)}}{n_{(s)} d_{(js)}}$$

Substituting this result in (8) and using the fact that $\hat{\lambda}_j(s) = \hat{\pi}_j(s)\hat{\lambda}(s)$, we now have:

$$\begin{aligned}\mathrm{V}\big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\big) &= \big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big)^2 \Bigg(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_l(n_{(l)} - d_{(l)})} + \frac{(n_{(s)} - d_{(js)})}{d_{(js)} n_{(s)}} \Bigg) \\ &= \mathrm{V}\big(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\big)\end{aligned}$$

We now consider:

$$\begin{aligned}\mathrm{Cov}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)) &= \big(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)\big) \\ &\quad \times \Bigg(\sum_{l=1}^{s\text{-}1} \frac{d_{(l)}}{n_{(l)}(n_{(l)} - d_{(l)})} - \frac{1}{n_{(s)}} \Bigg)\end{aligned} \tag{9}$$

If we replace $\hat{\lambda}(.)\hat{\pi}_j(.)$ with $\hat{\lambda}_j(.)$ in the RHS of (9), then:

$$\mathrm{Cov}(\hat{S}(s\text{-}1)\hat{\lambda}(s)\hat{\pi}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}(k)\hat{\pi}_j(k)) = \mathrm{Cov}(\hat{S}(s\text{-}1)\hat{\lambda}_j(s)\hat{S}(k\text{-}1)\hat{\lambda}_j(k))$$

This completes the proof that; $\mathrm{V}(\hat{F}_j^C(t)) = \mathrm{V}(\hat{F}_j^V(t))$.

6. Conclusion

We have presented the vertical model as a nonparametric model for analysis of discrete time competing risks data. We also demonstrated that the proposed model and the cause-specific-hazards model produce identical estimates. We focussed on the estimates for the cause-specific-hazards and the cumulative incidence functions. We also showed that the standard errors for the estimates of these quantities were identical under both models. Indeed, it is a roundabout way of estimating the cause-specific-hazards, however, there are cases in practice where these quantities cannot be estimated directly from the data such as when some of the subjects have failed with unknown failure causes. Furthermore, the cause-specific-hazards are not appropriate for application in the presence of a sizable proportion of cured subjects. Nicolaie et al. (2015) have extended the model to handle missing failure causes and the same authors, Nicolaie et al. (2018) have upscaled the model to handle cured subjects. The cause-specific-hazards model cannot handle these data complications. The proposed model, therefore, offers a possibility that the proposed model can also be upscaled to handle these challenges in discrete time. Ndlovu et al. (2020) have presented the vertical model as a nonparametric model for analysis of discrete time data that comes with missing failure causes. Another data complication that has not been explored as yet in the literature is the possibility that data might come with missing failure causes as well as cured subjects.

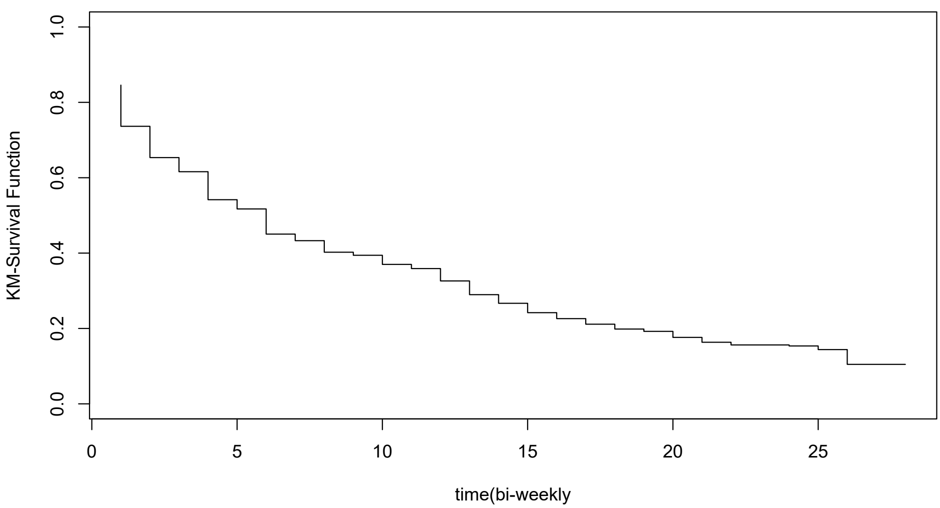

Figure 2: The KM-Survival Function

In clinical trials of a drug for treatment of some cancer, for example, implicit in the study is the expectation that data from that study may have a significant proportion of cured subjects if the drug proves to be effective against the cancer. It is also possible that the failure causes for some of the failures may not be recorded. The subjects with missing failure causes and cured subjects are distinct subjects because cured subjects are assumed to be mixed with censored subjects, whereas missing failure causes relate to subjects that failed. It is, therefore, not inconceivable, that data may come with missing failure causes and cured subjects. In fact, this very data set that was used for illustrative purpose in this article has missing failure causes and it also presents some evidence that there is a portion of cured subjects, albeit, minimal.

In Figure 2 we have plotted the KM survival function estimate for 3343 subjects. It is evident from the plot that the survival function does not approach zero fast enough, i.e. there is a portion of cured subjects. This means the cause-specific-hazards and cumulative incidence function estimates that were obtained from the proposed model are understated and the extent of bias is directly proportional to the relative size of cured subjects. This is an area that requires further exploration and our opinion is that the vertical model is a strong candidate for handling such data.

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, pp. 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, pp. 449–466.

Croissant, Y. and Graves, S., (2020). Ecdat: Data Sets for Econometrics. R package version 0.3 – 7

Davis, T. P. and Lawrance, A. J., (1989). The likelihood for competing risk survival analysis. *Board of the Foundation of the Scandinavian Journal of Statistics*, 16, pp. 23–28.

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

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Straus, D. J., Clarkson, B. D., and Brennan, M. F., (1993). On the use of cause-specific failure and conditional failure probabilities. examples from clinical oncology data. *American Statistical Association*, 88, pp. 400–409.

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 (Applied Statistics)*, 34, pp. 201–211.

Lee, M., Feuer, E., and Fine, J., (2018). On the analysis of discrete time competing risks data. *Biometrics*. doi.org/10.1111/biom.12881.

Ndlovu, B. D., Melesse, S., and Zewotir, T., (2020). A nonparametric vertical model: Anapplication to discrete time competing risks data with missing failure causes. *South African Journal of Statistics*, 6, pp. 534–545.

Nicolaie, M., van Houwelingen, H. C., and Putter, H., (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine*, 29, pp. 1190–1205.

Nicolaie, M. A., Taylor, J., and Legrand, C., (2018). Vertical modeling: analysis of competing risks data with a cure fraction. Lifetime Data Analysis