Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 47–61
http://www.ssca.org.in/journal

Comparison of Cause Specific Rate Functions of Panel Count Data with Multiple Modes of Recurrence

Sankaran P. G.[1]**, Ashlin Mathew P. M.**[2] **and Sreedevi E. P.**[1]
[1]*Department of Statistics, Cochin University of Science and Technology, Cochin*
[2]*Department of Statistics, St. Thomas College (Autonomous), Thrissur*

Received: 11 September 2022; Revised: 09 November 2023; Accepted: 15 November 2023

Abstract

Panel count data refer to the data arising from studies concerning recurrent events where study subjects are observed only at distinct time points. If these study subjects are exposed to recurrent events of several types, we obtain panel count data with multiple modes of recurrence. In the present paper, we propose a nonparametric test to compare cause specific rate functions of panel count data with more than one mode of recurrence. We carry out simulation studies to evaluate the performance of the test statistic in a finite sample setup. The proposed test is illustrated using two real-life panel count data sets, one arising from a medical follow-up study on skin cancer chemo prevention trial and the other on a warranty database for a fleet of automobiles.

Key words: Cause specific rate functions; Chi-Square test; Kernel estimation; Panel count data; Recurrent events.

AMS Subject Classifications: 62N01, 62N03

1. Introduction

Panel count data arise from longitudinal studies on recurrent events where each subject is observed only at discrete time points. In many situations, continuous observation is impossible due to cost, feasibility or other practical considerations. As a result, the number of occurrence of the events between consecutive observation times are only available; the exact recurrence times remain unknown (Kalbfleisch and Lawless (1985); Sun and Tong (2009); Zhao *et al.* (2011)). Panel count data is also termed interval count data or interval censored recurrent event data (Lawless and Zhan (1998); Thall and Lachin (1988)). In panel count data, the number of observation times and observation time points may vary for each subject. If each subject is observed only once, the number of recurrences of the event up to the observation time is only available. This special case of panel count data is commonly known as current status data.

The standard methods in the analysis of panel count data are focused on the rate function or the mean function of the underlying recurrent event process. Thall and Lachin

Corresponding Author: Sankaran P. G.
Email: sankaran.pg@gmail.com

(1988) and Lawless and Zhan (1998) considered the analysis of panel count data using rate functions. An estimator for the mean function based on isotonic regression theory was developed by Sun and Kalbfleisch (1995). Wellner and Zhang (2000) discussed likelihood based nonparametric estimation methods for the mean function and proposed a nonparametric maximum likelihood estimator (NPMLE) and a nonparametric maximum pseudo-likelihood estimator (NPMPLE) for the same. They also showed that NPMPLE is exactly the one studied in Sun and Kalbfleisch (1995). Some recent research works in this area include Zhou *et al.* (2017), Xu *et al.* (2018), Wang *et al.* (2019), Jiang *et al.* (2020) and Wang and Lin (2020) among others.

When an individual (subject) in the study is exposed to the risk of recurrence due to several types of events at each point of observation, we obtain panel count data with multiple modes of recurrence. Such data naturally arise from survival and reliability studies where the interest is focused on the recurrence of competing events which can be observed only at discrete time points. For example, consider the data on skin cancer chemo prevention trial discussed in Sun and Zhao (2013). The cancer recurrences of 290 patients with a history of non-melanoma skin cancers are observed at different monitoring times. The types of cancers are classified into basal cell carcinoma and squamous cell carcinoma and the recurrences due to both types of cancers at each monitoring time are observed for each individual. Covariate information on age, gender, number of prior tumours and DFMO status is also observed for each individual. As a result, we obtain panel count data with multiple modes of recurrence. A detailed analysis of the data is given in Section 4.

Even though recurrent event data exposed to multiple modes of recurrence is studied by many authors in literature (Cook and Lawless, 2007), panel count data with multiple modes of recurrence is less explored in literature. Sreedevi and Sankaran (2021) derived an expression for the cause specific mean functions and developed a nonparametric test for comparing the effect of different causes on recurrence times based on the developed estimators. Sankaran *et al.* (2020) considered non parametric estimation of cause specific rate functions and studied their properties. When study subjects are exposed to multiple modes of recurrence, it is important to test whether the effect of different modes are identical on the lifetime (Gray (1988)). Many authors including Aly *et al.* (1994) and Sankaran *et al.* (2010) addressed the above testing problem for right censored data. When the current status data is only available, Sreedevi *et al.* (2012) developed a test for independence of time to failure and cause of failure. Comparison of cumulative incidence functions of current status data with continuous and discrete observation times is studied by Sreedevi *et al.* (2014) and Sreedevi *et al.* (2019) respectively. Even though current status data can be considered as a special case of panel count data, the estimation procedures are different for both data types and the aforementioned tests cannot be used in the present situation.

The test proposed by Sreedevi and Sankaran (2021) can be used for comparing the mean functions of panel count data with more than one recurrence mode. However, there are several advantages in using the rate functions for the analysis of panel count data compared to the mean functions. Often, we assume that the mean function follows a non-homogeneous Poisson process, but this assumption is not required for analysing rate functions directly. In addition, rate functions are not constrained by the non decreasing property as of mean functions and hence it is easy to understand the changing recurrence patterns with rate functions. This motivated us to propose a test to compare the cause specific rate functions proposed by Sankaran *et al.* (2020).

The paper is organized as follows. In Section 2, we discuss the estimation of the cause specific rate functions and then propose a non parametric test to compare the rate functions of panel count data with multiple modes of recurrence. We also discuss the asymptotic properties of the proposed test statistic. In Section 3, we report the results of the simulation study conducted to evaluate the performance of the proposed test in finite samples. We illustrate the practical usefulness of the method by applying it to two real data sets in Section 4. Finally, Section 5 summarizes the major conclusions of the study with a discussion on future works.

2. Inference procedures

We study cause specific rate functions and their properties in detail in this section. Further, a non parametric test for comparing cause specific rate functions is presented.

2.1. Cause specific rate functions

Consider a study on n individuals from a homogeneous population who are exposed to the recurrent events due to $\{1, 2, ..., J\}$ possible causes. Assume that the event process is observed only at a sequence of random monitoring times. Consequently, the counts of the event recurrences due to each cause in between the observation times are only available; the exact recurrence times remain unknown. As a result, we observe the cumulative number of recurrences up to every observation time due to each cause. Define a counting process $N_j = \{N_j(t); t \geq 0\}$ where $N_j(t)$ denotes the number of recurrences of the event due to cause j up to time t. Define $\mu_j(t) = E(N_j(t))$ as the mean function of the recurrent event process due to cause j which are termed as cause specific mean functions. Define $r_j(t)dt = d\mu_j(t) = EdN_j(t)$ as the rate function of the recurrent event process due to cause j, for $j = 1, 2, ..., J$. $r_j(t)$ is referred to as the cause specific rate function. By studying cause specific rate functions, one can easily understand the difference in recurrence patterns due to various causes (modes) of recurrence.

In panel count data, we can note that the number of observation times as well as observation time points may be different for each individual. Let M_i be an integer-valued random variable denoting the number of observation times for $i = 1, 2, .., n$. Also let $T_{i,p}$ denote the p^{th} observation time for i^{th} individual for $p = 1, 2, ..M_i$ and $i = 1, 2, .., n$. Assume that the number of recurrences due to different causes is independent of the number of observation times as well as observation time points. Let $N_{i,p}^j$ denote the number of recurrences of the event observed for i^{th} individual due to cause j , for $p = 1, 2, ..., M_i$, $i = 1, 2, ..., n$ and $j = 1, 2, ..., J$. Now we observe n independent and identically distributed copies of $\{M_i, T_{i,p}, N_{i,p}^1, ..., N_{i,p}^J\}$, $p = 1, 2, ..., M_i$. The observed data will be of the form $\{m_i, t_{i,p}, n_{i,p}^1, ..., n_{i,p}^J\}$, $p = 1, 2, ..., m_i$ and $i = 1, 2, ..., n$.

Sankaran *et al.* (2020) introduced various estimators for cause specific rate functions and established their practical utility through numerical illustrations. The empirical estimators for the cause specific rate functions $r_j(t)$'s are defined as

$$\widehat{r_j(t)} = \frac{\sum_{i=1}^{n}\left[\sum_{p=1}^{m_i}\frac{(n_{i,p}^{j}-n_{i,p-1}^{j})I(t_{i,p}<t\leq t_{i,p-1})}{(t_{i,p}-t_{i,p-1})}\right]}{\sum_{i=1}^{n} I(t\leq t_{i,p})} \qquad j=1,2,...,J. \tag{1}$$

In this definition, the numerator gives the average number of recurrences due to cause j and the denominator is the number of individuals at risk at time t. Hence the estimators $\widehat{r_j(t)}$'s are the average of rate functions due to cause j, over all individuals. The cause specific mean functions can be directly estimated from Equation (1). When $J = 1$, Equation (1) reduces to the empirical estimator of the rate function given in Sun and Zhao (2013) and the expression is given by

$$\widehat{r(t)} = \frac{\sum_{i=1}^{n}\left[\sum_{p=1}^{m_i}\frac{(n_{i,p}-n_{i,p-1})I(t_{i,p}<t\leq t_{i,p-1})}{(t_{i,p}-t_{i,p-1})}\right]}{\sum_{i=1}^{n} I(t\leq t_{i,p})} \tag{2}$$

where $n_{i,p}$ denotes the number of recurrences of the event observed for i^{th} individual due to all possible modes of recurrence up to time p, for $p = 1, 2, ..., M_i$, $i = 1, 2, ..., n$. By definition, $\widehat{r(t)} = \sum_{j=1}^{J}\widehat{r_j(t)}$. In practice, the estimators of cause specific rate functions presented in Equation (1) change only at the observed time points. Accordingly, Sankaran *et al.* (2020) proposed a smoothed version of the estimators of cause specific rate functions using kernel estimation techniques and also studied the asymptotic properties.

Let $K(t)$ be a non-negative kernel function symmetric about $t = 0$ with $\int_{-\infty}^{\infty} K(t)dt = 1$. Also, let $h_n > 0$ be the bandwidth parameter. Let $b_1 < b_2 < ... < b_l$ are the distinct observed time points in the set $\{T_{i,p},\ p = 1, 2, ..., M_i,\ i = 1, 2, ..., n\}$. Define $\widehat{r_{qj}} = \widehat{r_j(b_q)}$, for $q = 1, 2, ..., l$, $j = 1, 2, ..., J$. Now, the kernel estimators of $r_j(t)$'s are given as

$$\widehat{r_j^*(t)} = \sum_{q=1}^{l} w_q(t)\widehat{r_{qj}} \qquad j = 1, 2, ..., J. \tag{3}$$

where

$$w_q(t) = \frac{w_q^*(t, h_n)}{\sum_{u=1}^{l} w_u^*(t, h_n)} \qquad q = 1, 2, ..., l.$$

and

$$w_q^*(t, h_n) = h_n^{-1} K\left(\frac{t-b_q}{h_n}\right)$$

with

$$K(t) = (2\pi)^{-1/2}\exp(-t^2/2).$$

The smoothed estimators $\widehat{r_j^*(t)}$ of the cause specific rate functions are weighted average of $\widehat{r_j(t)}$'s. Smoothed estimators of overall rate functions can also be constructed in a similar way (Sun and Zhao (2013)). Clearly, $\widehat{r^*(t)} = \sum_{j=1}^{J}\widehat{r_j^*(t)}$, where $\widehat{r^*(t)}$ is the kernel estimator of the overall rate function. In practice, the bandwidth h_n for which the MSE is minimum is selected to employ smoothing.

The asymptotic properties of the estimators $\widehat{r_j^*(t)}$'s are studied and derived in Sankaran *et al.* (2020). Without loss of generality, assume that the kernel function $K(x)$ satisfies the following mild regularity conditions.
C1 : $K(x)$ is bounded ie $\sup\{K(x), x \in R\} < \infty$
C2 : $|xK(x)| \to 0$ as $|x| \to \infty$
C3 : $K(x)$ is symmetric about 0, ie $K(-x) = K(x)$, $x \in R$
Also suppose that, as $n \to \infty$ the bandwidth parameter h_n satisfies the conditions (i) $h_n \to 0$ (ii) $nh_n \to \infty$ and (iii) $nh_n^2 \to \infty$.
Under the assumptions C1, C2 and C3, Sankaran *et al.* (2020) showed that for fixed t, the estimators $\widehat{r_j^*(t)}$'s are asymptotically normal with mean $\lambda_j(t) = E(\widehat{r_j^*(t)})$ and standard deviation $\sigma_j(t) = \text{s.d}(\widehat{r_j^*(t)})$ for $j = 1, 2, ..., J$.

2.2. Test statistic

In this study, we focus on comparing the cause specific rate functions due to various recurrence modes. This may be helpful in selecting the appropriate treatment for a group of patients in a clinical study or to evaluate a newly introduced system in reliability experiments. To develop a test statistic, we now consider the hypothesis,

$$H_0 : r_j(t) = r_{j\prime}(t) \text{ for all } t > 0,\ j \neq j\prime = 1, 2, ..., J$$

against

$$H_1 : r_j(t) \neq r_{j\prime}(t) \text{ for some } t > 0 \quad \text{and} \quad j \neq j\prime = 1, 2, ..., J. \tag{4}$$

Since $r(t) = \sum_{j=1}^{J} r_j(t)$, the above hypothesis can also be written as

$$H_0 : r_j(t) = \frac{r(t)}{J} \text{ for all } t > 0, \quad j \neq j\prime = 1, 2, ..., J$$

against

$$H_1 : r_j(t) \neq \frac{r(t)}{J} \text{ for some } t > 0 \quad \text{and} \quad j \neq j\prime = 1, 2, ..., J. \tag{5}$$

To test H_0 against H_1, we choose $\widehat{r_j^*(t)}$ as the smoothed estimators for the cause specific rate functions defined in Equation (3). A smoothed estimator for the overall rate function $r(t)$ specified in Equation (2) is constructed by omitting the information on the mode of recurrence. Let $\widehat{r^*(t)}$ denote smoothed estimator of the overall rate function. A similar procedure of estimating the overall mean function by ignoring the cause of recurrence information is used in Sreedevi and Sankaran (2021) for comparing cause specific mean functions.

To develop a test statistic for comparing cause specific rate functions, consider the function

$$v_j(t) = \int_0^t w(u) \left[\widehat{r_j^*(u)} - \frac{\widehat{r^*(u)}}{J} \right] \text{d}u \qquad \text{for all } j = 1, 2, \ldots, J \tag{6}$$

where $w(.)$ is an appropriate data dependent weight function which is used to compensate the effect of censoring. The weight functions are also employed to increase the efficiency of the test statistic and to set it asymptotically distribution free (Pepe and Mori (1993)). The function $v_j(.)$ is similar to the one proposed by Sreedevi and Sankaran (2021) to compare the cause specific mean functions of panel count data. Now to test the null hypothesis given in Equation (4), we propose the test statistic

$$Z(\tau) = v'(\tau)\hat{\sum}(\tau)^{-1}v(\tau) \tag{7}$$

where τ is the largest monitoring time in the study and $v(\tau) = (v_1(\tau), \ldots, v_k(\tau))'$; $\sum(\tau)^{-1}$ is the generalized inverse $\hat{\sum}(\tau)$, where $\hat{\sum}(\tau)$ is a consistent estimator of $\sum(\tau)$, the variance-covariance matrix of $v(\tau)$. The matrix $\sum(\tau)$ involves variances of $\widehat{r_j^*(\tau)}$ and $\widehat{r(\tau)}$ and co-variances between $\widehat{r_j^*(\tau)}$ and $\widehat{r_{j'}^*(\tau)}$ for $j \neq j' = 1, 2, \ldots, J$ and that between $\widehat{r_j^*(\tau)}$ and $\widehat{r(\tau)}$. The bootstrap procedure is used to find the estimate of the variance-covariance matrix, since the expression for $\sum(\tau)$ is complex. To find the asymptotic distribution of $Z(\tau)$ given in Equation(7), consider the quantity

$$v_j(t) = \int_0^t w(u)\left[\widehat{r_j^*(u)} - \frac{\widehat{r^*(u)}}{J}\right] \mathrm{d}u \quad \text{for all } j = 1, 2, \ldots, J$$

which can be written as

$$\begin{aligned} v_j(t) = &\int_0^t w(u)\left[\widehat{r_j^*(u)} - r_j(u)\right] \mathrm{d}(u) + \int_0^t w(u)\left[r_j(u) - \frac{r(u)}{J}\right] \mathrm{d}u \\ &+ \int_0^t w(u)\left[\frac{r(u)}{J} - \frac{\widehat{r^*(u)}}{J}\right] \mathrm{d}u, \quad j = 1, 2, \ldots J \end{aligned}$$

Now under H_0, $r_j(t) = r(t)/J$ for all t, we get

$$v_j(t) = \int_0^t w(u)\left[\widehat{r_j^*(u)} - r_j(u)\right] \mathrm{d}u + \int_0^t w(u)\left[\frac{r(u)}{J} - \frac{\widehat{r^*(u)}}{J}\right] \mathrm{d}u, \quad j = 1, 2, \ldots, J$$

Now from the asymptotic properties of the kernel estimators of cause specific rate functions discussed in Sankaran *et al.* (2020) it follows that, under H_0 for any $\tau > 0$, the limiting distribution of $v(\tau) = (v_1(\tau), \ldots, v_J(\tau))'$ is a $J-$ variate normal with mean zero vector and variance-covariance matrix $\sum(\tau)$, where τ is the largest monitoring time in the study. Accordingly, under the regularity conditions stated in Section 2.1, the quadratic form $Z(\tau)$ follows a χ^2 distribution with $(J-1)$ degrees of freedom. We reject H_0, if $Z(t) \geq \chi^2_{\alpha,(J-1)}$ where $\chi^2_{\alpha,(J-1)}$ is the ordinate value of chi-square distribution with $(J-1)$ degrees of freedom at α level.

3. Simulation studies

We conduct simulation studies to evaluate the performance of the proposed test statistic in finite samples. The situation with two modes of recurrence is considered here. We generate panel count data of the form $\{m_i, t_{i,p}, n^1_{i,p}, n^2_{i,p}\}$ for $p = 1, 2, ..., m_i$ and $i = 1, 2, ..., n$ to carry out simulation. The number of observation times m_i for each individual is generated

Table 1: Empirical Type I error and power of the test in percentage for the weight functions $w(.) = 1, w(.) = n$ and $w(.) = \widehat{r^*(t)}$

		n					n		
$(\theta_1, \theta_2, \theta_3)$	α	100	200	500	$(\theta_1, \theta_2, \theta_3)$	α	100	200	500
				$w(t) = 1$					
(1,1,1)	5	5.8	5.4	5.1	(1,1,2)	5	5.6	5.2	4.9
	1	2	1.7	1.3		1	1.7	1.4	1.1
(1,2,1)	5	65.8	71.4	79.5	(1,2,2)	5	66.8	74.8	80.7
	1	63.7	67.2	73.1		1	65.2	73.1	75.2
(1,3,1)	5	74.5	81.9	86.4	(1,3,2)	5	81.5	87.7	92.4
	1	73.0	78.6	83.1		1	79.4	85.6	91.6
(1,4,1)	5	90.3	92.1	97.2	(1,4,2)	5	96.5	98.2	99.9
	1	87.4	91.8	94.5		1	96.8	98.2	99.1
(1,5,1)	5	98.9	100	100	(1,5,2)	5	100	100	100
	1	98.4	99.7	100		1	99.8	100	100
				$w(t) = n$					
(1,1,1)	5	4.5	4.7	5.2	(1,1,2)	5	4.4	4.8	5.1
	1	2	1.7	1.3		1	1.4	1.3	0.9
(1,2,1)	5	67.1	73.2	78.4	(1,2,2)	5	70.4	79.5	84.7
	1	66.7	69.2	74.1		1	68.1	74	79
(1,3,1)	5	79.6	83.9	86.4	(1,3,2)	5	85.2	89.3	94.7
	1	73.0	78.6	83.1		1	80.5	87.2	93.7
(1,4,1)	5	94.3	98.1	99.9	(1,4,2)	5	99.9	100	100
	1	87.4	96.8	97.2		1	99.8	99.9	100
(1,5,1)	5	100	100	100	(1,5,2)	5	100	100	100
	1	100	100	100		1	99.8	100	100
				$w(t) = \widehat{r^*(t)}$					
(1,1,1)	5	4.7	5.2	5	(1,1,2)	5	5.5	4.8	5.1
	1	0.7	1.2	0.9		1	1.3	1.2	1
(1,2,1)	5	73.2	81.0	85.7	(1,2,2)	5	76.9	84.1	85.4
	1	71.1	78.9	84.3		1	71.0	77.2	84.2
(1,3,1)	5	89.5	92.5	98.4	(1,3,2)	5	88.8	91.4	97.5
	1	83.2	88.6	96.9		1	85.0	87.3	96.0
(1,4,1)	5	99.9	100	100	(1,4,2)	5	100	100	100
	1	99.7	100	100		1	99.8	99.8	100
(1,5,1)	5	100	100	100	(1,5,2)	5	100	100	100
	1	100	100	100		1	100	100	100

from a discrete uniform distribution $U(1,10)$ for $i = 1,2,...,n$. Thus the maximum number of observations for each individual is restricted up to 10. Then we generate gap times between each observation from uniform distribution $U(0,5)$. The discrete observation time points $t_{i,p}$ for $p = 1,2,...,m_i$ and $i = 1,2,...,n$ are generated using the above-mentioned time gaps. A bivariate Poisson distribution with parameters $(\theta_1,\theta_2,\theta_3)$ is employed to generate recurrent processes $n^1_{i,p}$ and $n^2_{i,p}$.

The joint mass function of the bivariate Poisson distribution with parameters $(\theta_1,\theta_2,\theta_3)$ is given by

$$f(x,y) = \exp\{-(\theta_1+\theta_2+\theta_3)\}\frac{{\theta_1}^x}{x!}\frac{{\theta_2}^y}{y!}\sum_{k=0}^{min(x,y)}\binom{x}{k}\binom{y}{k}k!\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k. \tag{8}$$

The marginal distribution of X and Y is Poisson distribution with $E(X) = \theta_1 + \theta_3$, $E(Y) = \theta_2 + \theta_3$ and $\text{cov}(X,Y) = \theta_3$ gives a measure of dependence between random variables X and Y. Sankaran *et al.* (2020) used a similar procedure to generate panel count data with multiple modes of recurrence.

In the above simulation framework, if we set $\theta_1 = \theta_2$ and assign a non-zero value for θ_3, it corresponds to a situation where the cause specific rate functions are identical. Accordingly, the null hypothesis H_0 will be true. When the difference between θ_1 and θ_2 increases, the difference between the two rate functions also increases which results in a situation where the null hypothesis is false. Hence the parameter combination with $\theta_1 = \theta_2$ gives the type I error of the test and all other choices of parameter combinations give the power of the proposed test. We carry out simulation studies for different combinations of $(\theta_1,\theta_2,\theta_3)$ to calculate the empirical type I error and power of the test. For this purpose, observations of different sample sizes n = 100 or n = 200 or n = 500 are simulated and the process is repeated 1000 times. We employ three different choices of weight functions similar to Sreedevi and Sankaran (2021) which are (i) $w(t) = 1$, (ii) $w(t) = n$, the number of individuals in the study and (iii) $w(t) = \widehat{r^*(t)}$, the smoothed estimator of overall rate function.

Table 1 gives the type I error and the power of the proposed test statistic in percentage for significance level $\alpha = 0.05$ and $\alpha = 0.01$. From Table 1, we can see that type I error of the test approaches the chosen significance level. The test is efficient in terms of power also. Also, as the difference between θ_1 and θ_2 increases, the power of the test also increases.

4. Data analysis

The proposed inference procedures are illustrated using two real-life data sets in this section.

4.1. Skin cancer chemo prevention trial data

We consider the data arising from the skin cancer chemo prevention trial given in Sun and Zhao (2013) for demonstration. The study was conducted to test the effectiveness of the DFMO (DIfluromethylornithire) drug in reducing new skin cancers in a population with a history of non-melanoma skin cancers, basal cell carcinoma and squamous cell carcinoma.

The data consists of 290 patients with a history of non-melanoma skin cancers. The observation and follow-up times differ for each patient. The data has the counts of two types of recurring events basal cell carcinoma and squamous cell carcinoma which we treat here as two modes of recurrence (Sreedevi and Sankaran (2021)).

Table 2: Test statistic values of the proposed test for different weight functions

Weight function	Test statistic	p -value
1	26.97	$< .0005$
n	31.92	$< .0005$
$\widehat{r^*(.)}$	37.74	$< .0005$

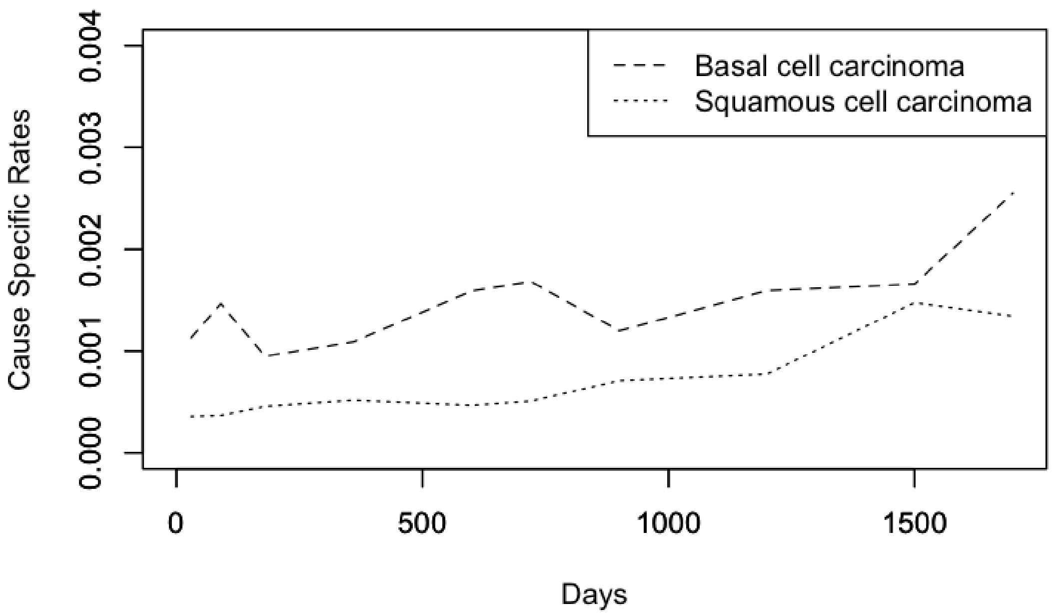

Figure 1: Kernel estimates of cause specific rate functions due to basal cell carcinoma and squamous cell carcinoma for $h_n = 1.76$

In the data set, the number of observations on an individual varies from 1 to 17 and the time of observation varies from 12 to 1766 days. The cause specific rate functions due to basal cell carcinoma and squamous cell carcinoma are estimated using Equation (3). Further, the proposed procedures are applied to evaluate the test statistic. Table 2 gives the chi-square test statistic values of the proposed test for different weight functions. From the value of the test statistic, it is clear that we reject the null hypothesis and conclude that the rate functions due to basal cell carcinoma and squamous cell carcinoma are significantly different.

The plots of the kernel estimators with bandwidth parameter value $h_n = 1.76$ is given in Figure 1. The bandwidth value $h_n = 1.76$ is chosen, which minimizes the MSE of the estimates, $\widehat{r_j^*(t)}$ for $j = 1, 2$.

From Figure 1, it can be noted that the recurrence rate of basal cell carcinoma is greater than the recurrence rate of squamous cell carcinoma at all time points, which clearly indicates the rejection of H_0. Since the rate functions are not monotonic, the change points of recurrence patterns can also be easily identified from the graph.

4.2. Automobile warranty claims data

We apply the proposed methods to the automobile warranty claims data studied in Somboonsavatdee and Sen (2015). The data set comprises the recurrent failure history of a fleet of automobiles. The outcome of interest is the repeated mileages at failure for multiple vehicles of a certain model and make, obtained from the warranty claim database which also includes the labour code associated with the failure. In the data, the source and specifics are masked for de-identification purposes.

Table 3: Test statistic values of the proposed test for different weight functions

Weight function	Test statistic	p -value
1	49.15	$< .0005$
n	68.96	$< .0005$
$\widehat{r^*(.)}$	79.55	$< .0005$

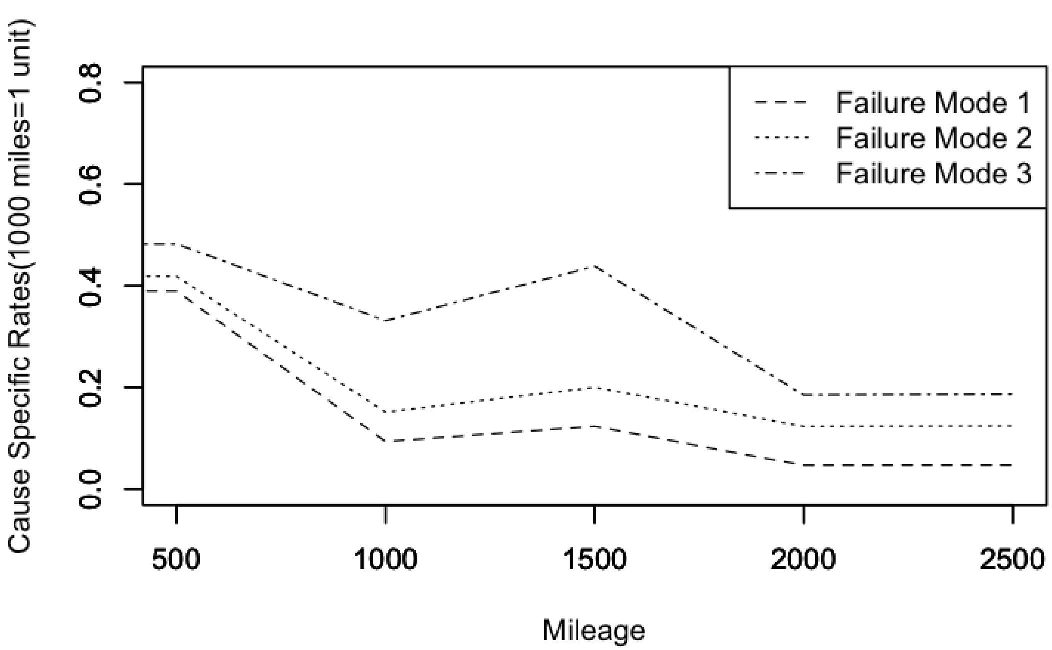

Figure 2: Kernel estimates of cause specific rate functions due to three modes of recurrences for $h_n = 1.67$

The database consists of the recurrent failure history of 456 vehicles subjected to type I censoring at 3000 miles. Fourteen different labor codes of the warranty claims of each vehicle were recorded with mileage at filing. Due to the absence of a specific description of the component associated with labor code, the grouping was determined on the basis of the rate of failures. The fourteen individual labor codes were combined into three broad groups of failure modes FM1, FM2 and FM3, where FM1 comprises labor codes with shape parameters ranging between 0.2 and 0.36, FM2 covers labor codes with shape parameter estimates between 0.4 and 0.55, whereas FM3 combines the remaining codes that have the slowest rate of growth with shape parameter estimates varying between 0.7 and 0.93. Table IV in Somboonsavatdee and Sen (2015) presents the data on 172 vehicles that have at least one documented record of warranty claim for repair.

We observed the recurrent failure history data at 1000, 2000 and 3000 mileages at which the number of failures due to each mode is noted, thereby making the recurrent event data as a panel count data with multiple modes of recurrence. The complete data set used in our study is given in Table 4 in Appendix.

Table 3 gives the chi-square test statistic values of the proposed test for different weight functions for automobile warranty data. For all choices of weight functions, we reject the null hypothesis and conclude that the rate functions due to the three modes of failure are significantly different. The plots of the kernel estimators with bandwidth parameter value $h_n = 1.67$ is given in Figure 2. The bandwidth value $h_n = 1.67$ is chosen as it minimizes the MSE of the estimates. From Figure 2, it can be noted that the recurrence rates of each mode of recurrence (FM1, FM2 and FM3) are distinct at all observed miles, which clearly indicates the rejection of H_0.

5. Conclusion

In the present paper, we developed non parametric inference procedures for the analysis of panel count data with multiple modes of recurrence based on cause specific rate functions. We proposed a test statistic to test the equality of cause specific rate functions. A simulation study was carried out by generating the data from a bivariate Poisson process to assess the performance of the proposed test in finite samples. Two real-life data sets, one from skin cancer chemo prevention trial (Sun and Zhao (2013)) and the other from automobile warranty claims (Somboonsavatdee and Sen (2015)) were analysed to demonstrate the practical utility of the procedures.

The nature of dependence between time to failure and cause of failure is important for modelling competing risks data. Even though the problem is studied under right censoring, it is unexplored for panel count data. We can use either cause specific mean functions or cause specific rate functions to tackle this problem. Works in this direction will be done separately. Regression analysis of panel count data with multiple modes of recurrence using rate functions is also under investigation.

Acknowledgements

The first author would like to thank Science Engineering and Research Board, DST, Government of India and the third author acknowledges the gratitude to Kerala State Council for Science Technology and Environment for the financial support provided to carry out this research work.

References

Aly, E.-E. A., Kochar, S. C., and McKeague, I. W. (1994). Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *Journal of the American Statistical Association*, **89**, 994–999.

Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, **16**, 1141–1154.

Jiang, H., Su, W., and Zhao, X. (2020). Robust estimation for panel count data with informative observation times and censoring times. *Lifetime Data Analysis*, **26**, 65–84.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, **80**, 863–871.

Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*, **26**, 549–565.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, **12**, 737–751.

Sankaran, P. G., Ashlin Mathew, P. M., and Sreedevi, E. P. (2020). Cause specific rate functions for panel count data with multiple modes of recurrence. *Journal of the Indian Statistical Association*, **58**, 193–211.

Sankaran, P. G., Nair, N. U., and Sreedevi, E. P. (2010). A quantile based test for comparing cumulative incidence functions of competing risks models. *Statistics & Probability Letters*, **80**, 886–891.

Somboonsavatdee, A. and Sen, A. (2015). Parametric inference for multiple repairable systems under dependent competing risks. *Applied Stochastic Models in Business and Industry*, **31**, 706–720.

Sreedevi, E. P. and Sankaran, P. G. (2021). Nonparametric inference for panel count data with competing risks. *Journal of Applied Statistics*, **48**, 3102–3115.

Sreedevi, E. P., Sankaran, P. G., and Dewan, I. (2019). Comparison of cumulative incidence functions of current status competing risks data with discrete observation times. *Communications in Statistics-Theory and Methods*, **48**, 5766–5776.

Sreedevi, E. P., Sankaran, P. G., and Dhanavanthan, P. (2012). A nonparametric test for independence of time to failure and cause of failure of current status competing risks data. *Calcutta Statistical Association Bulletin*, **64**, 167–180.

Sreedevi, E. P., Sankaran, P. G., and Dhanavanthan, P. (2014). A nonparametric test for comparing cumulative incidence functions of current status competing risks data. *Journal of Statistical Theory and Practice*, **8**, 743–759.

Sun, J. and Kalbfleisch, J. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, **5**, 279–289.

Sun, J. and Zhao (2013). *Statistical Analysis of Panel Count Data*. Springer.

Sun, L. and Tong, X. (2009). Analyzing longitudinal data with informative observation times under biased sampling. *Statistics & Probability Letters*, **79**, 1162–1168.

Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, **83**, 339–347.

Wang, J. and Lin, X. (2020). A Bayesian approach for semiparametric regression analysis of panel count data. *Lifetime Data Analysis*, **26**, 402–420.

Wang, W., Wu, X., Zhao, X., and Zhou, X. (2019). Quantile estimation of partially varying coefficient model for panel count data with informative observation times. *Journal of Nonparametric Statistics*, **31**, 932–951.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics*, **28**, 779–814.

Xu, D., Zhao, H., and Sun, J. (2018). Joint analysis of interval-censored failure time data and panel count data. *Lifetime Data Analysis*, **24**, 94–109.

Zhao, X., Balakrishnan, N., and Sun, J. (2011). Nonparametric inference based on panel count data. *Test*, **20**, 1–42.

Zhou, J., Zhang, H., Sun, L., and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Analysis*, **23**, 560–584.

ANNEXURE

Table 4: Automobile warranty data

ID	MIL	FM1	FM2	FM3	TOTAL	ID	MIL	FM1	FM2	FM3	TOTAL
1	1000	1	1	0	2	**45**	1000	2	0	0	2
1	3000	1	0	0	1	**46**	1000	0	1	0	1
2	1000	0	0	2	2	**47**	1000	1	0	0	1
3	3000	0	0	1	1	**47**	3000	0	1	0	1
4	2000	0	0	1	1	**48**	1000	1	1	0	2
5	1000	1	1	1	3	**49**	1000	0	1	0	1
5	2000	1	0	0	1	**50**	1000	0	0	1	1
6	1000	0	0	1	1	**51**	3000	0	0	1	1
7	1000	1	0	0	1	**52**	1000	0	0	1	1
8	1000	1	0	0	1	**53**	2000	1	0	0	1
9	1000	0	1	0	1	**54**	1000	0	1	0	1
10	2000	0	0	2	2	**55**	1000	1	0	0	1
11	1000	1	0	0	1	**56**	1000	0	1	0	1
12	1000	1	0	0	1	**57**	1000	0	2	0	2
13	3000	0	0	1	1	**57**	2000	1	0	1	2
14	1000	0	1	1	2	**58**	1000	0	0	1	1
15	1000	0	1	0	1	**59**	1000	0	1	0	1
15	2000	0	1	0	1	**60**	1000	0	1	0	1
16	2000	0	1	1	2	**61**	2000	1	0	0	1
16	3000	0	1	0	1	**62**	1000	0	1	0	1
17	1000	1	2	1	4	**63**	2000	0	0	1	1
17	2000	1	0	0	1	**64**	1000	0	0	1	1
18	3000	0	0	1	1	**65**	1000	1	0	0	1
19	1000	0	1	0	1	**66**	1000	2	1	0	3
20	1000	1	0	0	1	**67**	1000	1	0	0	1
21	1000	0	1	0	1	**67**	3000	0	0	1	1
22	3000	0	1	0	1	**68**	1000	0	1	0	1
23	1000	1	0	0	1	**69**	2000	0	1	0	1
24	3000	1	0	0	1	**70**	1000	1	0	0	1
25	1000	0	1	0	1	**71**	1000	1	0	0	1
26	1000	1	0	1	2	**72**	2000	0	0	2	2
26	2000	1	2	0	3	**73**	1000	1	0	0	1
26	3000	0	2	0	2	**73**	2000	0	0	1	1
27	3000	0	1	0	1	**74**	1000	1	0	1	2
28	2000	0	0	1	1	**74**	2000	0	0	1	1
29	1000	1	0	1	2	**75**	1000	1	0	0	1
30	3000	0	2	0	2	**76**	1000	0	0	1	1
31	2000	0	1	0	1	**77**	1000	0	1	1	2
32	2000	0	1	0	1	**78**	1000	0	1	0	1
33	3000	0	0	1	1	**79**	3000	0	0	1	1
34	1000	0	1	0	1	**80**	1000	1	0	0	1
35	1000	0	0	1	1	**81**	1000	0	0	1	1
35	2000	1	0	0	1	**82**	1000	1	0	0	1
36	1000	0	1	0	1	**83**	1000	0	0	1	1
37	1000	1	0	0	1	**84**	2000	0	0	1	1
37	2000	0	0	1	1	**85**	1000	0	2	0	2
38	1000	1	1	0	2	**86**	1000	0	0	1	1
39	1000	0	2	0	2	**86**	2000	0	2	0	2
40	1000	0	2	0	2	**87**	1000	1	0	0	1
41	3000	0	0	1	1	**88**	2000	0	0	1	1
42	1000	0	0	1	1	**88**	3000	0	0	1	1
43	1000	0	0	1	1	**89**	3000	1	0	0	1
44	3000	0	1	0	1	**90**	1000	0	0	2	2

ID	MIL	FM1	FM2	FM3	TOTAL	ID	MIL	FM1	FM2	FM3	TOTAL
90	3000	0	0	1	1	**135**	1000	0	0	1	1
91	1000	0	1	0	1	**136**	1000	0	0	1	1
92	1000	0	1	0	1	**137**	1000	0	0	1	1
93	1000	0	0	1	1	**138**	1000	0	0	1	1
94	1000	1	1	0	2	**138**	3000	1	0	0	1
95	1000	1	0	0	1	**139**	1000	1	0	0	1
96	2000	0	0	1	1	**140**	1000	1	0	0	1
97	2000	0	0	1	1	**141**	3000	0	1	0	1
98	1000	0	1	0	1	**142**	1000	0	1	1	2
98	2000	1	1	1	3	**143**	1000	1	0	0	1
99	1000	1	0	0	1	**143**	3000	0	0	1	1
100	1000	1	0	1	2	**144**	1000	0	1	0	1
101	1000	0	0	1	1	**144**	2000	0	0	2	2
102	1000	1	0	0	1	**145**	1000	0	1	0	1
103	1000	1	0	0	1	**146**	1000	1	0	1	2
104	2000	0	0	1	1	**146**	3000	0	0	1	1
105	1000	1	0	0	1	**147**	1000	0	1	0	1
106	1000	0	0	2	2	**148**	3000	0	0	1	1
107	3000	0	1	0	1	**149**	1000	1	0	0	1
108	1000	1	0	0	1	**150**	1000	1	0	0	1
108	3000	0	0	1	1	**151**	1000	0	0	1	1
109	2000	0	0	1	1	**152**	1000	1	0	0	1
109	3000	0	1	0	1	**153**	1000	0	1	0	1
110	1000	1	0	1	2	**154**	3000	1	0	0	1
111	1000	1	0	0	1	**155**	1000	0	1	0	1
112	1000	0	1	0	1	**156**	3000	0	1	0	1
113	1000	0	0	1	1	**157**	2000	0	0	1	1
114	1000	1	0	0	1	**158**	3000	0	0	1	1
115	1000	0	1	1	2	**159**	1000	0	0	1	1
116	1000	1	0	0	1	**160**	3000	0	0	1	1
117	2000	0	1	1	2	**161**	1000	0	1	2	3
118	2000	0	0	1	1	**161**	2000	0	1	2	3
119	2000	1	0	0	1	**161**	3000	1	0	2	3
120	1000	1	0	1	2	**162**	2000	1	0	0	1
121	1000	0	0	1	1	**163**	1000	0	1	0	1
121	3000	0	0	1	1	**164**	3000	0	0	1	1
122	1000	1	0	1	2	**165**	1000	0	2	2	4
123	2000	0	1	0	1	**165**	2000	0	1	1	2
124	1000	1	0	0	1	**165**	3000	0	1	1	2
125	2000	0	0	1	1	**166**	1000	1	0	1	2
126	1000	2	0	1	3	**167**	1000	0	1	0	1
126	3000	0	0	2	2	**167**	3000	0	1	0	1
127	2000	0	0	1	1	**168**	1000	0	1	0	1
128	2000	0	1	0	1	**169**	1000	1	0	0	1
129	1000	2	3	1	6	**169**	2000	0	0	4	4
129	2000	0	0	1	1	**169**	3000	0	1	0	1
130	1000	0	1	0	1	**170**	1000	0	1	0	1
131	1000	1	0	0	1	**170**	2000	0	0	1	1
132	3000	0	0	1	1	**171**	1000	0	0	1	1
133	2000	1	0	1	2	**172**	2000	0	0	1	1
134	2000	0	1	1	2						