

Sugata Sen Roy¹ and Moumita Chatterjee²

²*SQC and OR Unit, Indian Statistical Institute Hyderabad, Hyderabad 500007, India*

Abstract

Key words: Survival times; Alternating recurrent event; Bivariate copula; Cox's model.

AMS Subject Classifications: 62N01

In life-time data analysis, one often comes across the problem of recurrent events, that is events which repeat after a certain interval of time. Examples include, relapses of the disease in individuals suffering from cancer, times to first/post-repair failures of any machine and days of periodical unemployment among agricultural labourers. The time intervals between such repetitions are usually unequal and may often exhibit certain patterns. Comprehensive reviews of early studies in this area can be found in Kalbfleisch and Prentice (2002) and Clayton (1994).

However, very often the recurrent events are not instantaneous but have themselves a life-time of their own. This gives rise to a sequence of alternating life-times. For example, when one is looking at the time to relapse of cancer, one also has intervening times of treatment, *i.e.* time to cure, before one can again start to measure the time to the next relapse. Similarly, the recurrent time-lengths of unemployment are sandwiched between recurrent times of employment, and recurrent times to machine failure are separated by recurrent repair times. Thus one would be faced with alternating sets of times for two alternating complementary events. Data of these kind are termed as “recurrent episode” data or alternating state data (Hougaard, 2000).

There have been very few studies on such alternating recurrent events. Gap time methods have been applied by Lin, Sun and Ying (1999), Wang and Chang (1999) and Pena, Strawderman and Hollander (2002), while Yan and Fine (2008) used temporal process regression to analyse recurrent episode data in cystic fibrosis patients. Sen Roy and Chatterjee (2015) analysed data of this kind without considering the correlation between the alternating events.

In this paper we compare two techniques as developed by Chatterjee and Sen Roy (2018(a) and (b)) for modelling alternating sets of times under right-censoring, where the censoring occurs primarily because of either loss to follow-up or withdrawal or the termination of the study. The first, a parametric method, uses a bivariate distribution to take account of the dependence between the two event times. The joint distribution is obtained from the marginals through a bivariate copula. Since the moments of the bivariate distribution are likely to exhibit some pattern, the parameters of the model are linked through functional relationships to capture this dependence over the cycles. In the second method, the semiparametric Cox Proportional Hazard (PH) model is used, with the dependence over the cycles and across the two events structured through an autoregressive type model.

The paper is organized as follows. Section 2 introduces the model and discusses the underlying censoring mechanism. Section 3 discusses the first method, including the choice of an appropriate copula function. The second method is discussed in Section 4. A motivation for the study was the clinical trial data on patients suffering from cystic fibrosis as reported by Fuchs *et al.* (1994). Section 5 contains a comparative study of this data based on the two methods. Some concluding remarks are made in Section 6.

2. The model

Consider two events E and F , where F is the complement of E . We also use E and F to denote the two alternating states *i.e.* an individual in state E is faced with the event F , after which he is in state F and facing event E . Assume for the moment that all individuals start in state E . We then have an alternating sequence of E 's and F 's. Let X and Y denote respectively the lengths of times in states E and F *i.e.* X is the time to the event F and Y is the time to the event E .

Suppose there are n individuals in the study. If for the i^{th} individual, $i = 1, \dots, n$, there are m_i recurrences of each event, we have for this individual a sequence $X_{1i}, \dots, X_{m_i i}$ of E -times each culminating in event F and a sequence $Y_{1i}, \dots, Y_{m_i i}$ of F -times each culminating in event E , assuming of course that there is no censoring. (X_{ji}, Y_{ji}) then gives the complete cycle for the j^{th} recurrence with time length $X_{ji} + Y_{ji}$.

Before constructing the likelihood function some assumptions made above need to be relaxed. Firstly, it is not necessary that all individuals are in state E when they enter the study. Similarly, an individual need not necessarily be in state F with an event E occurring at the termination of his observation. In fact the observations need not end with any event and an individual may voluntarily withdraw from the study or may be lost to follow-up while in any of the two states. For example, in a study on time to cure and time to relapse, the death of an individual during the study period may be taken as a loss to follow-up. All these can be looked upon as censored observations. To incorporate these into the model we define

for the i^{th} individual the following indicator functions :

$$\delta_{1ji}^* = \begin{cases} 1 & \text{for all } j = 2, \dots, m_i \\ 1 & \text{if } j = 1 \text{ and the } i^{th} \text{ individual starts in state E} \\ 0 & \text{if } j = 1 \text{ and the } i^{th} \text{ individual starts in state F} \end{cases}$$

$$\delta_{2ji}^* = \begin{cases} 1 & \text{for all } j = 1, \dots, (m_i - 1) \\ 1 & \text{if } j = m_i \text{ and the } i^{th} \text{ individual withdraws in state F} \\ & \text{or ends by experiencing event E} \\ 0 & \text{if } j = m_i \text{ and the } i^{th} \text{ individual withdraws in state E} \\ & \text{or ends by experiencing event F} \end{cases}$$

While δ_{1ji}^* is an indicator of whether the i^{th} individual enters the study in the E or the F state, δ_{2ji}^* indicates whether the individual leaves the study either during or at the end of the E or F state. Further to mark whether the event is observed or censored in the last stage, we use δ_{1ji} and δ_{2ji} to indicate whether the i^{th} individual exits with an E-event or an F-event or is censored while in state E or in state F respectively, where

$$\delta_{1ji} = \begin{cases} 1 & \text{for all } j = 1, \dots, (m_i - 1) \\ 1 & \text{if } j = m_i \text{ and the } i^{th} \text{ individual experiences event F} \\ 0 & \text{if the } i^{th} \text{ individual withdraws while in state E at } j = m_i \end{cases}$$

$$\delta_{2ji} = \begin{cases} 1 & \text{for all } j = 1, \dots, (m_i - 1) \\ 1 & \text{if } j = m_i \text{ and the } i^{th} \text{ individual experiences event E} \\ 0 & \text{if the } i^{th} \text{ individual withdraws while in state F at } j = m_i \end{cases}$$

3. Method 1

In modelling these recurrent event-times, it is obvious that for a given individual, the X_{ji} 's and Y_{ji} 's will be dependent, both over X and Y as also over the respective sequences. However, joint modelling of (X_{ji}, Y_{ji}) for $j = 1, \dots, m_i$ would be difficult, if not impossible. An alternative would be to begin by assuming that the sequence of X_{ji} 's and Y_{ji} 's have the same distributional form, say, $F_X(x, \theta_j)$ and $F_Y(y, \eta_j)$ respectively, where $\theta_j = (\theta_{1j}, \dots, \theta_{pj})'$ and $\eta_j = (\eta_{1j}, \dots, \eta_{qj})'$ are respectively the p and q dimensional parameter vectors which characterize the two distribution functions. To accommodate the relationships between the X_{ji} 's, let, for each $k = 1, \dots, p$, $g_k(h, \alpha_k)$ be continuous and differentiable functions of h which are characterized by the r_k -dimensional parameter vectors α_k 's. Then for $j = 1, \dots, m_i$, the parameters θ_j are linked through $\theta_{kj} = g_k(j, \alpha_k)$, $k = 1, \dots, p$. Similarly, the relationships between the Y_{ji} 's are modelled through the parameters as $\eta_{lj} = h_l(j, \beta_l)$, $l = 1, \dots, q$ and $j = 1, \dots, m_i$, with β_l 's the s_l -dimensional parameter vectors characterizing the continuous and differentiable functions h_l for $l = 1, \dots, q$.

The g_k and h_l functions are not predetermined. For a given data-set, estimates of the parameters θ_j are first obtained by separately fitting $F_X(x, \theta_j)$ for each of the cycles, $j = 1, \dots, m_i$. Then plotting each estimated θ_{kj} against j , an idea of the functional form of $g_k(j, \alpha_k)$ for $k = 1, \dots, p$ can be envisaged. The functional forms of the elements of h_l can be obtained similarly.

The g_k and h_l functions allow us to capture the pattern that the two events exhibit over the cycles. However, the inter-relationship between the X_{ji} and the Y_{ji} are still to be accounted for, and this we do through the use of a copula.

As is well-known, a copula is a function through which a multivariate distribution function can be built from the univariate marginal distributions of each of the variables. A bivariate copula can thus be used in our model to find a functional relationship between X_{ji} and Y_{ji} for each j . Thus if C denotes the copula function, then under the assumption of continuity of F_{X_j} and F_{Y_j} , it follows from Sklar (1959) that the joint distribution of X_{ji} and Y_{ji} is

$$F_{X_j Y_j}(x, y, \zeta_j) = C(F_{X_j}(x, \theta_j), F_{Y_j}(y, \eta_j), \xi_j), \quad (1)$$

where ξ_j is a dependence parameter and $\zeta_j = (\theta_j', \eta_j', \xi_j)'$.

Let $C'_U(u, v)$, $C'_V(u, v)$ and $C''_{UV}(u, v)$ denote respectively the first and second derivatives of C with respect to u and v . Defining the marginal probability functions of X_{ji} and Y_{ji} respectively as $f_{X_j}(x, \theta_j)$ and $f_{Y_j}(y, \eta_j)$, the joint probability density function of (X_{ji}, Y_{ji}) is then

$$f_{X_j Y_j}(x_{ji}, y_{ji}, \zeta_j) = C''_{F_X F_Y}(F_{X_j}(x_{ji}, \theta_j), F_{Y_j}(y_{ji}, \eta_j), \xi_j) f_{X_j}(x_{ji}, \theta_j) f_{Y_j}(y_{ji}, \eta_j). \quad (2)$$

Since by our assumption, X_{ji} precedes Y_{ji} , the survival and hazard functions of X_{ji} can be obtained directly from the marginal distribution function as

$$S_{X_j}(x_{ji}, \theta_j) = 1 - F_{X_j}(x_{ji}, \theta_j) \quad (3)$$

$$\text{and } \lambda_{X_j}(x_{ji}, \theta_j) = -\frac{d}{dx_{ji}} \ln S_{X_j}(x_{ji}, \theta_j), \quad (4)$$

respectively. However, the survival and hazard functions for Y_{ji} will be conditioned on $X_{ji} = x_{ji}$. Since the conditional distribution of $Y_{ji}|X_{ji} = x_{ji}$ is

$$F_{Y_j|X_j}(y_{ji}, \zeta_j|x_{ji}) = C'_{F_X}(F_{X_j}(x_{ji}, \theta_j), F_{Y_j}(y_{ji}, \eta_j), \xi_j) \quad (5)$$

these are given respectively by

$$S_{Y_j|X_j}(y_{ji}, \zeta_j|x_{ji}) = 1 - F_{Y_j|X_j}(y_{ji}, \zeta_j|x_{ji}) \quad (6)$$

$$\text{and } \lambda_{Y_j|X_j}(y_{ji}, \zeta_j|x_{ji}) = -\frac{d}{dy_{ji}} \ln S_{Y_j|X_j}(y_{ji}, \zeta_j|x_{ji}). \quad (7)$$

However, like the θ_{ji} 's and η_{ji} 's, the dependence parameters ξ_j 's are also likely to be related over the cycles for $j = 1, \dots, m_i$. This is accounted for through the relation $\xi_j = u(j, \kappa)$, where $\kappa = (\kappa_1, \dots, \kappa_d)'$ is the d -dimensional parameter vector characterizing the continuous and differentiable function u .

We define the $r = \sum_{k=1}^p r_k$ dimensional vector $\alpha = (\alpha'_1, \dots, \alpha'_p)'$ and the $s = \sum_{l=1}^q s_l$ dimensional vector $\beta = (\beta'_1, \dots, \beta'_q)'$ and write $\nu = (\alpha', \beta', \kappa)'$ as the $(r + s + d)$ -dimensional vector of unknown parameters. The likelihood function is then constructed by considering for each individual and each cycle the marginal survival and hazard functions of the E-times

X_{ji} and the conditional survival and hazard times of the F -times $Y_{ji}|X_{ji} = x_{ji}$. This leads to

$$\begin{aligned}
 l(\nu) &= l(\alpha, \beta, \kappa | (x_{ji}, y_{ji}), j = 1 \dots, m_i, i = 1, \dots, n) \\
 &= \prod_{i=1}^n \prod_{j=1}^{m_i} \left[\left\{ S_{X_j}(x_{ji}, \mathbf{g}(j, \alpha)) \left\{ \lambda_{X_j}(x_{ji}, \mathbf{g}(j, \alpha)) \right\}^{\delta_{1ji}} \right\}^{\delta_{1ji}^*} \right. \\
 &\quad \left. \left\{ S_{Y_j|X_j}(y_{ji}, \mathbf{g}(j, \alpha), \mathbf{h}(j, \beta), u(j, \kappa)) \left\{ \lambda_{Y_j|X_j}(y_{ji}, \mathbf{g}(j, \alpha), \mathbf{h}(j, \beta), u(j, \kappa)) \right\}^{\delta_{2ji}} \right\}^{\delta_{2ji}^*} \right] \quad (8)
 \end{aligned}$$

Building the score vector and the information matrix by taking first and second derivatives with respect to $\nu = (\alpha', \beta', \kappa)'$, the maximum likelihood (m.l.) estimates $\hat{\nu} = (\hat{\alpha}', \hat{\beta}', \hat{\kappa})'$ can now be obtained simultaneously by the Newton-Raphson method. This for any cycle j leads to the estimator $\hat{\zeta}_j = (\hat{\theta}'_j, \hat{\eta}'_j, \hat{\xi}_j)'$ of the model parameters and hence to the estimators $\hat{S}_{X_j}(x_{ji}, \hat{\theta}_j)$ and $\hat{S}_{Y_j|X_j}(y_{ji}, \hat{\zeta}_j|x_{ji})$ of the respective survival functions and their corresponding hazard functions.

The likelihood function (8) depends on the copula through equations (3-7). Hence it is necessary to choose an appropriate copula before maximizing the likelihood function. Owing to its simple form which is easy to work with, the Archimedean copula is a popular choice for modelling bivariate distributions (Nelson, 2006). Its added advantage is that it caters to a variety of dependence structures including Kendall's τ , which can be evaluated directly from its generator Genest and Mackay (1986). Shih (1998) remarks that in survival studies, the Clayton copula, a member of the Archimedean family, is probably the most widely used, primarily because of its mathematical convenience and attractive properties. In this paper we tried different copulas under the Archimedean set-up, but eventually found that the Clayton copula gives the best fit.

4. Method 2

Next let the E -times depend on a set of p covariates $\mathbf{z} = (z_1, \dots, z_p)$ while the F -times depend on a set of q covariates $\mathbf{w} = (w_1, \dots, w_q)$. Then the Cox PH models for the hazard functions for the j^{th} cycle of the E and F times are

$$\begin{aligned}
 \lambda_{Ej}(x_{ji}, \mathbf{z}_j) &= \lambda_{E0j}(x_{ji}) e^{\boldsymbol{\theta}'_j \mathbf{z}_j} \\
 \lambda_{Fj}(y_{ji}, \mathbf{w}_j) &= \lambda_{F0j}(y_{ji}) e^{\boldsymbol{\eta}'_j \mathbf{w}_j}
 \end{aligned}$$

where $\boldsymbol{\theta}_j = (\theta_{j1}, \dots, \theta_{jp})'$ and $\boldsymbol{\eta}_j = (\eta_{j1}, \dots, \eta_{jq})'$ are the p and q dimensional vectors of coefficients, and $\lambda_{E0j}(x)$ and $\lambda_{F0j}(y)$ are the baseline hazard functions. Of course, in our study, there are no exogenous covariates. However, to account for the dependence over the alternating sequences, we assume a first order autoregressive-type model where x_j depends on y_{j-1} and x_{j-1} , while y_j depends on x_j and y_{j-1} . Of course, this can be easily extended to a higher-order model, but for simplicity and because the number of cycles will generally be small, we restrict ourselves to the first order only. Defining $x_0 = 0$ and $y_0 = 0$, this leads

to the hazard model

$$\begin{aligned}\lambda_{Ej}(x_{ji}, \mathbf{z}_{ji}, \mathbf{y}_{ji}) &= \lambda_{E0j}(x_{ji})e^{(\mu_1 y_{(j-1)i} + \mu_2 x_{(j-1)i})} \\ &= \lambda_{E0j}(x_{ji})e^{\boldsymbol{\mu}' \mathbf{y}_{ji}} \\ \lambda_{Fj}(y_{ji}, \mathbf{w}_{ji}, \mathbf{x}_{ji}) &= \lambda_{F0j}(y_{ji})e^{(\gamma_1 x_{ji} + \gamma_2 y_{(j-1)i})} \\ &= \lambda_{F0j}(y_{ji})e^{\boldsymbol{\gamma}' \mathbf{x}_{ji}}\end{aligned}$$

where $\mathbf{y}_{ji} = (y_{(j-1)i}, x_{(j-1)i})'$, $\mathbf{x}_{ji} = (x_{ji}, y_{(j-1)i})'$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$.

The corresponding survival functions are

$$\begin{aligned}S_{Ej}(x_{ji}, \mathbf{z}_{ji}, \mathbf{y}_{ji}) &= [S_{E0j}(x_{ji})]e^{\boldsymbol{\mu}' \mathbf{y}_{ji}} \\ S_{Fj}(y_{ji}, \mathbf{w}_{ji}, \mathbf{x}_{ji}) &= [S_{F0j}(y_{ji})]e^{\boldsymbol{\gamma}' \mathbf{x}_{ji}}\end{aligned}$$

where $S_{E0j}(x_{ji})$ and $S_{F0j}(y_{ji})$ are the respective baseline survival functions.

For $m = \max m_i$, the likelihood function based on the m_i recurrences of each of the n individuals is then

$$\begin{aligned}& l(\boldsymbol{\mu}, \boldsymbol{\gamma} | (x_{ji}, y_{ji}), j = 1 \dots, m_i, i = 1, \dots, n) \\ &= \prod_{i=1}^n \prod_{j=1}^m \left[\left\{ S_{Eji}(x_{ji}, \mathbf{y}_{ji}) \left\{ \lambda_{Eji}(x_{ji}, \mathbf{y}_{ji}) \right\}^{\delta_{1ji}} \right\}^{\delta_{1ji}^*} \left\{ S_{Fji}(y_{ji}, \mathbf{x}_{ji}) \left\{ \lambda_{Fji}(y_{ji}, \mathbf{x}_{ji}) \right\}^{\delta_{2ji}} \right\}^{\delta_{2ji}^*} \right] \\ &= \prod_{i=1}^n \prod_{j=1}^m \left[\left\{ \lambda_{E0j}(x_{ji}) e^{\boldsymbol{\mu}' \mathbf{y}_{ji}} \right\}^{\delta_{1ji} \delta_{1ji}^*} \left\{ e^{-\Lambda_{E0j}(x_{ji}) e^{\boldsymbol{\mu}' \mathbf{y}_{ji}}} \right\}^{\delta_{1ji}^*} \right] \times \\ &\quad \left[\left\{ \lambda_{F0j}(y_{ji}) e^{\boldsymbol{\gamma}' \mathbf{x}_{ji}} \right\}^{\delta_{2ji} \delta_{2ji}^*} \left\{ e^{-\Lambda_{F0j}(y_{ji}) e^{\boldsymbol{\gamma}' \mathbf{x}_{ji}}} \right\}^{\delta_{2ji}^*} \right] \\ &= l_E(\boldsymbol{\mu}, |(x_{ji}, y_{ji}), j = 1 \dots, m, i = 1, \dots, n) l_F(\boldsymbol{\gamma} | (x_{ji}, y_{ji}), j = 1 \dots, m, i = 1, \dots, n), \quad (9)\end{aligned}$$

$\Lambda_{E0j}(x_{ji})$ and $\Lambda_{F0j}(y_{ji})$ being the baseline cumulative hazard functions of E and F , respectively.

Equation (9) shows that the likelihood function decomposes into distinct E and F components. Hence treating the two components separately, and interchanging the order of the product, the partial likelihoods for the two components are obtained as

$$pl_E(\boldsymbol{\mu}) = \prod_{j=1}^m \prod_{i=1}^{n_{Ej}} \left[\frac{e^{\boldsymbol{\mu}' \mathbf{y}_{ji}}}{\sum_{l \in \mathfrak{R}_{Ej}(x_{ji})} e^{\boldsymbol{\mu}' \mathbf{y}_{jl}}} \right]^{\delta_{1ji} \delta_{1ji}^*} \quad (10)$$

$$\text{and } pl_F(\boldsymbol{\gamma}) = \prod_{j=1}^m \prod_{i=1}^{n_{Fj}} \left[\frac{e^{\boldsymbol{\gamma}' \mathbf{x}_{ji}}}{\sum_{l \in \mathfrak{R}_{Fj}(y_{ji})} e^{\boldsymbol{\gamma}' \mathbf{x}_{jl}}} \right]^{\delta_{2ji} \delta_{2ji}^*} \quad (11)$$

where n_{Ej} and n_{Fj} are respectively the number of individuals in the risk sets $\mathfrak{R}_{Ej}$ and $\mathfrak{R}_{Fj}$ at the beginning of the E and F states of the j^{th} cycle.

In order to find the estimates of the parameters involved we treat the two equations (10) and (11) separately, and obtain their score vectors and information matrices. The Newton-Raphson method can then be used to solve for the estimates.

5. Numerical examples

5.1. The rhDNase data

In this section we analyse the cystic fibrosis data as reported by Fuchs et. al. (1994). Cystic fibrosis is a genetic disorder that can cause chronic infection. The infection produces thick, sticky mucus that clogs the respiratory organs and can recur several times in a patient's lifetime. Each infection can have different times to cure. In 1992, a trial was conducted to test the extent to which rhDNase could reduce the clogging as compared to placebo. Of the 647 patients, 322 were placed under the treatment and their times to cure as well as times to the next infection were recorded. Several recurrences are observed for some of the patients. However, data beyond the 3rd recurrence is sparse. Fuchs et. al. (1994) restricted their study to the first exacerbation only.

Therneau and Hamilton (1997) considered the successive exacerbation although as independent occurrences. Chang and Wang (1999) and Huang and Wang (2005) implemented the Gap time method to analyze the episode lengths but considered them independently of the frequency. Moreover, their assumption of conditional independence of the gap times did not match the empirical findings. However, none of the previous studies considered estimating the survival times of both infection and cure assuming a dependence structure.

In using our technique we assume the *onset of the infection* as event F (state E being the non-infection period) and *cure* as the event E (state F being the period of infection).

5.2. Method 1

In the first method, to guess the underlying distributions, the Kaplan-Meier (KM) estimator (Kleinbaum and Klein, 2012) is used to obtain rough fits to the data for both the infection and cure times for each of the cycles. Four alternative distributions, the lognormal (LN), the Loglogistic (LL), the Gamma (G) and the Weibull (W) are then fitted to each of these times and these are mapped on to the corresponding KM curves. The goodness of fit for each distribution is judged using a technique common in functional data analysis. This requires recording the survival probabilities at several selected time-points on each of the five fitted curves. The sum of squares (SSE) of the difference between these values for each of the four curves against the corresponding values on the KM curve are then computed. The curve with the minimum SSE is then selected. For our data, the Weibull gives the best fit for both the time to infection and the time to cure for most of the cycles and particularly for the first three. Since the data from the fourth cycle onwards is sparse, the Weibull distribution is selected to model the two event times.

Writing the marginal distribution of the j^{th} time to infection (X_j) as $W(\theta_{1j}, \theta_{2j})$, $\theta_{1j}, \theta_{2j} > 0$ and the marginal distribution of the j^{th} time to cure (Y_j) as $W(\eta_{1j}, \eta_{2j})$, $\eta_{1j}, \eta_{2j} > 0$, individual Weibull distributions are fitted to each cycle to get initial estimates of the scale parameters θ_{1j} and η_{1j} and the shape parameters θ_{2j} and η_{2j} for each j . A plot of each of these against j is shown in Figure 1. Although the plot for η_1 is not very conclusive, we assume the following relationships:

$$\begin{aligned} \theta_{1j} &= e^{(\alpha_{11} + \alpha_{12}j)} \quad \text{and} \quad \theta_{2j} = e^{(\alpha_{21} + \alpha_{22}j)}; \\ \eta_{1j} &= e^{(\beta_{11} + \beta_{12}j)} \quad \text{and} \quad \eta_{2j} = e^{(\beta_{21} + \beta_{22}j)}; \quad j = 1, \dots, 4. \end{aligned} \quad (12)$$

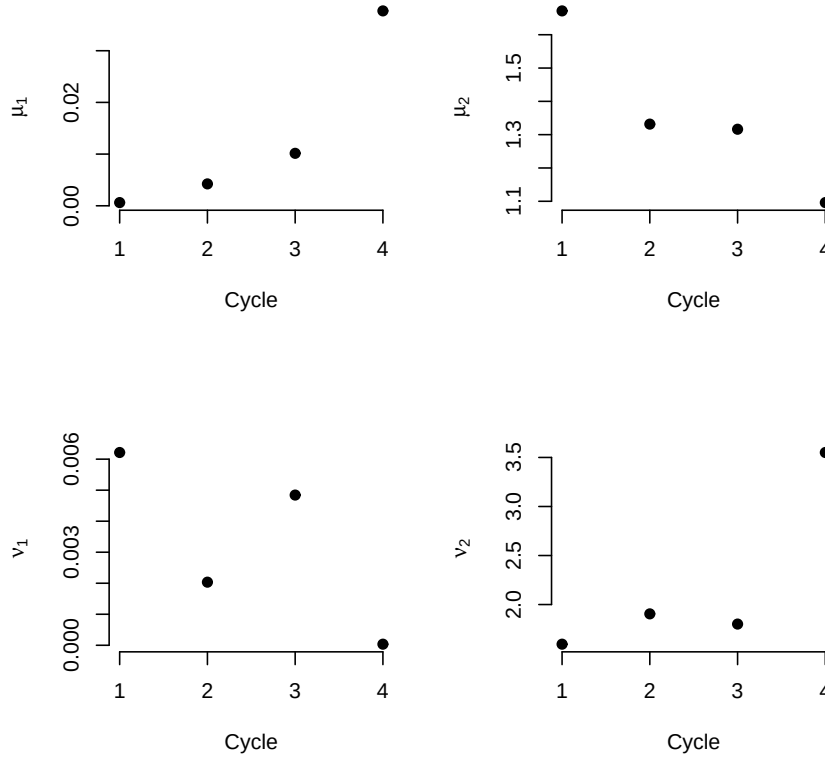


Figure 1: Functional forms of the estimated Weibull parameters over the stages

Next taking ξ_j as the dependence parameter, the joint distribution of (X_j, Y_j) is constructed for the j^{th} cycle using the Clayton copula.

Computing Kendall's $\hat{\tau}_j$ for each j , where $\hat{\tau}_j = (C_j - D_j)/(C_j + D_j)$, C_j and D_j being respectively the number of concordant and discordant pairs in the j^{th} cycle, the estimates of $\hat{\xi}_j$ are obtained from the relation $\hat{\xi}_j = 2\hat{\tau}_j/(1 - \hat{\tau}_j)$, which holds true for the Clayton copula. The values across the cycles indicate that a constant $\xi_j = -0.28$ can be assumed for all j . The negative value of $\hat{\xi}_j$ is justified since it is expected that the more severe is the case, the longer is the times to cure and the shorter is the time to the next relapse.

Writing the likelihood function in (8) in terms of $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})'$ and solving by Newton Raphson method, the estimates as shown in Table 1 are obtained. The corresponding estimated standard errors are also shown in brackets. Finally, the estimated hazard curves for infection and cure are obtained as shown in Figure 2.

The left panel relates to the time to infection, and both for placebo and treatment, the hazards are increasing over cycles, which means that more the recurrences the quicker is the chance of the next recurrence. For the placebo, although initially decreasing slightly, the hazards remain, more or less constant, over time. However, for the treatment, except for the first cycle, the hazards are initially low but increases over time. This means that with time the chances of catching an infection increases. What intrigues us is that for the

second and third cycle the hazards for treatment are greater than that of the placebo after a given time. This is counter-intuitive and needs to be more closely looked into. For time to cure, the right panel shows increasing hazards for both placebo and treatment over all the cycles, with the latter dominating the former in general. This implies that cure under treatment is faster than that under placebo. And as expected, the hazards for both placebo and treatment increases over cycles, implying that in either case, with successive recurrences, it takes longer to get cured.

Table 1: Estimates of the parameters and their standard errors along with 95% C.I.

Time to infection	Estimates	std. error	95% C.I.
α_{11}	-3.4152	0.6560	(-3.4868,-3.34354)
α_{12}	-0.5144	0.4103	(-0.5592,-0.4696)
α_{21}	0.294	0.1416	(0.2785,0.3095)
α_{22}	-0.2357	0.0814	(-0.2446,-0.2268)

Time to cure	Estimates	std. error	95% C.I.
β_{11}	2.8855	0.7307	(2.8057,2.9653)
β_{12}	2.3647	0.514	(2.3085,2.4208)
β_{21}	0.2239	0.1058	(0.2123,0.2355)
β_{22}	0.3011	0.0523	(0.2954,0.3068)

5.3. Method 2

In this section, we use the Cox model with the autoregressive-type event-time structures. Table 2 shows the autoregressive coefficients obtained by solving the partial likelihoods (10)-(11). Joint testing of $H_{01} : (\mu_1, \mu_2) = (0, 0)$ and $H_{02} : (\gamma_1, \gamma_2) = (0, 0)$ shows that the current times of relapse and cure are significantly dictated by their past values. In Figure 3, the hazard curves are plotted against the baseline hazard for both treatment and placebo.

Intersetingly, the left panel of Figure 3 corroborates the fact that hazard of infection is greater under the treatment than under the placebo. However, as in method 1, the cure-time is faster under treatment than under placebo for all the cycles.

Table 2: The estimates of the regression parameters and their standard errors

parameters	Estimates	standard errors
(μ_1, μ_2)	(0.09, 0.009)	(0.003, 0.001)
(γ_1, γ_2)	(0.02, 0.06)	(0.002, 0.004)

6. Some concluding remarks

In this paper we have considered two different methods for modelling alternating recurrent event times. The first models the dependence structure through a bivariate copula and relationships of the cycle parameters, while the second uses the Cox PH model with the dependence brought in through autoregressive structures. The methods are illustrated

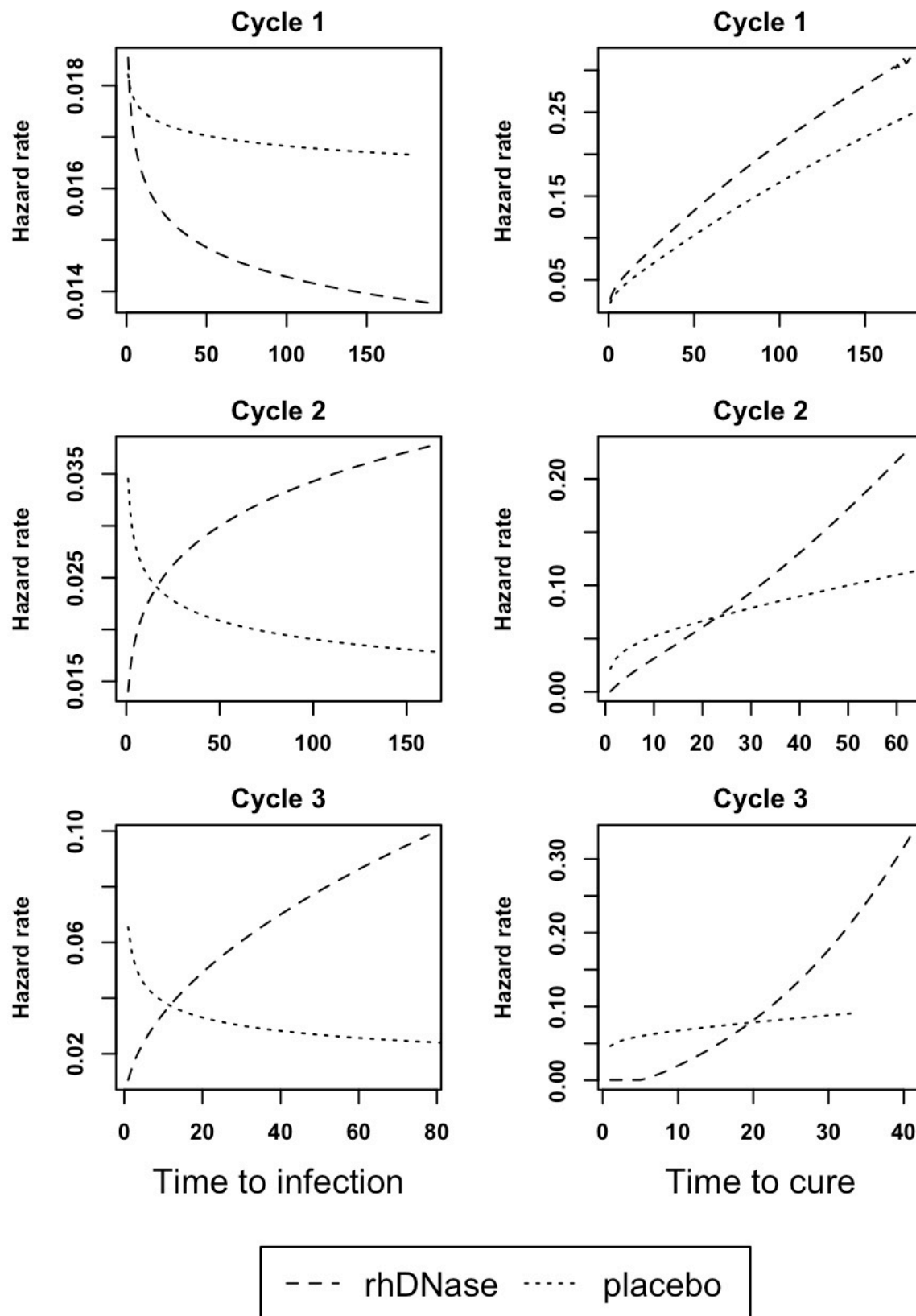


Figure 2: The hazard curves by method 1

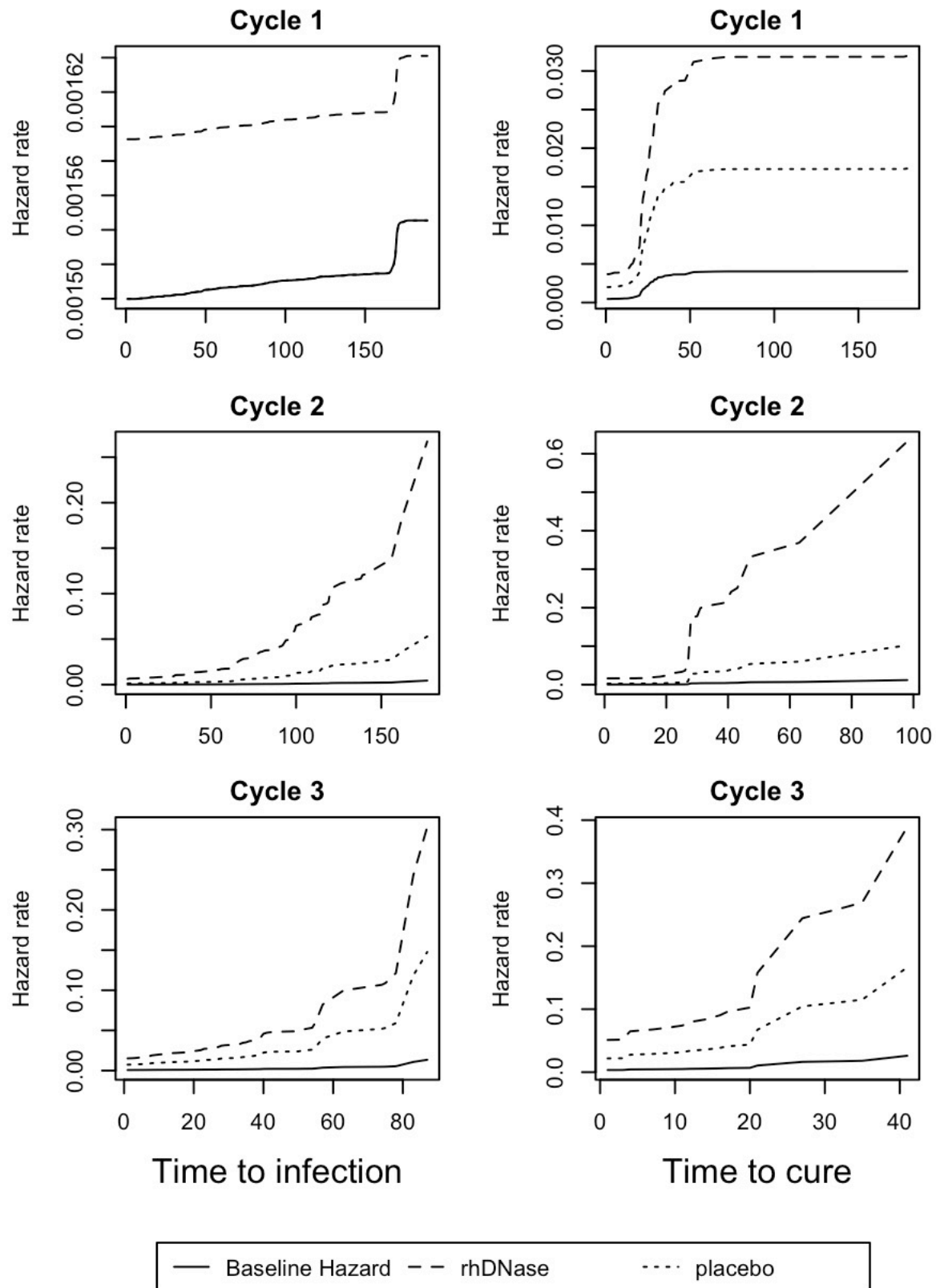


Figure 3: The hazard curves by method 2

through the cystic fibrosis data-set. Except for the first-cycle time to infection under treatment, the patterns in each of the two methods corroborate each other.

The first method, since it directly models the pattern over cycles, is likely to give smoother and sequenced hazard curves. However, if one is interested in the small twists and turns, the second method might be preferred.

The study can be extended to other forms of censoring, particularly interval censoring. Also covariates can be incorporated in the models to observe their impacts on survival times.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

- Chang, S. H. and Wang, M. C. (1999). Conditional regression analysis for recurrence-time data. *Journal of the American Statistical Association*, **94**, 1221–1230.
- Chatterjee, M. and Sen Roy, S. (2018a). Estimating the hazard functions of two alternating recurrent events in the presence of covariates. *Advances in Statistical Analysis*, **102**, 289–304.
- Chatterjee, M. and Sen Roy, S. (2018b). A copula-based approach for estimating the survival functions of two alternating recurrent events. *Journal of Statistical Computing and Simulation*, **88**, 3098–3115.
- Clayton, D. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research*, **3**, 244–262.
- Fuchs, H., Borowitz, D., Christiansen, D., Morris, E., Nash, M., Ramsey, B., Rosenstein, B., Smith, A., and Wohl, M. (1994). Effect of aerosolized re-combinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *New England Journal of Medicine*, **331**, 637–642.
- Genest, C. and MacKay, R. J. (1986). The joy of copulas: Bivariate distributions with univariate marginals. *The American Statistician*, **40**, 280–283.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. 1st ed. (NY): Springer-Verlag.
- Huang, C. Y. and Wang, M. C. (2005). Nonparametric estimation of the bivariate recurrence time distribution. *Biometrics*, **61**, 392–402.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. John Wiley & Sons.
- Kleinbaum, D. G. and Klein, M. (2012). *Survival Analysis: A Self-learning Text*. 3rd ed. Springer, New York.
- Lin, D. Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of gap time distributions for serial events with censored data. *Biometrika*, **86**, 59–70.
- Nelson, R. B. (2006). *An Introduction to Copulas*. 2nd ed. Springer, New York.
- Pena, E. A., Strawderman, R. L., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, **96**, 1299–1315.
- Sen Roy, S. and Chatterjee, M. (2015). Estimating the hazard functions of two alternating occurring recurrent events. *Journal of Applied Statistics*, **42**, 1547–1555.

- Shih, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, **85**, 189–200.
- Sklar, A. (1959), Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.
- Therneau, T. M. and Hamilton, S. A.. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine*, **16**, 2029–2047.
- Wang, M. C. and Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, **94**, 146–153.
- Yan, J. and Fine, J. P. (2008). Analysis of episodic data with application to recurrent pulmonary exacerbations in cystic fibrosis patients. *Journal of the American Statistical Association*, **103**, 498–510.