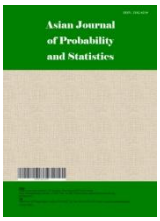

Asian Journal of Probability and Statistics

15(4): 144-162, 2021; Article no.AJPAS.77313
ISSN: 2582-0230

An Introspective Overview of the Dynamics of Recurrent Events Data Analysis

Anthony Joe Turkson [a*], Timothy Simpson [b] and John Awuah Addor [a]

[a] Takoradi Technical University, Department of Mathematics, Statistics and Actuarial Science, Ghana.
[b] Takoradi Technical University, Department of Computer Science, Ghana.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v15i430371
Editor(s):
(1) Dr. Dariusz Jacek Jakóbczak, Koszalin University of Technology, Poland.
Reviewers:
(1) G Abozaid, Aswan University, Egypt.
(2) Sevgi YURT ÖNCEL, Kırıkkale University, Türkiye.
(3) K Rajyalakshmi, Koneru Lakshmaiah Education Foundation, India.
Complete Peer review History, details of the editor(s), Reviewers and additional Reviewers are available here: https://www.sdiarticle5.com/review-history/77313

Original Research Article

Received 07 October 2021
Accepted 14 December 2021
Published 14 December 2021

Abstract

A recurrent event remains the outcome variable of interest in many biometric studies. Recurrent events can be explained as events of defined interest that can occur to same person more than once during the study period. This study presents an overview of different pertinent recurrent models for analyzing recurrent events.
Aims: To introduce, compare, evaluate and discuss pros and cons of four models in analyzing recurrent events, so as to validate previous findings in respect of the superiority or appropriateness of these models.
Study Design: A comparative studies based on simulation of recurrent event models applied to a tertiary data on cancer studies.
Methodology: Codes in R were implemented for simulating four recurrent event models, namely; The Andersen and Gill model; Prentice, Williams and Peterson models; Wei, Lin and Weissferd*;* and Cox frailty model. Finally, these models were applied to analyze the first forty subjects from a study of Bladder Cancer Tumors. The data set contained the first four repetitions of the tumor for each patient, and each recurrence time was recorded from the entry time of the patient into the study. An isolated risk interval is defined by each time to an event or censoring.
Results: The choice and usage of any of the models lead to different conclusions, but the choice depends on: risk intervals; baseline hazard; risk set; and correlation adjustment or simplistically, type of data and research

**Corresponding author: Email: anthony.turkson@ttu.edu.gh;*

question. The PWP-GT model could be used if the research question is focused on whether treatment was effective for the kt event since the previous event happened. However, if the research question is designed to find out whether treatment was effective for the kt event since the start of treatment, then we could use the PWP- TT. The AG model will be adequate if a common baseline hazard could be assumed, but the model lacks the details and versatility of the event-specific models. The WLW model is very suitable for data with diverse events for the same person, which underscores a potentially different baseline hazard for each type.
Conclusion: PWP-GT has proven to be the most useful model for analyzing recurrent event data.

Keywords: bladder cancer; cox model; recurrent events; survival analysis; simulation.

1 Introduction

Survival Analysis belongs to the class of statistical methods intended for studying and analyzing longitudinal data with the view of unearthing the timing and nature of occurrences of events. The method is used to find the survival probabilities of individuals from a point in time until the required event occurs. A recurrent event remains the outcome variable of interest in many biometric studies. Recurrent events allude to events of interest that repeatedly affect a given individual. These events may all be homogenous or heterogenous. Two features may define recurrent event data, these include events ordering and the exposure of an individual to risk for only one such event at a time.

This paper is devoted to the study of survival functions of recurrent events. Special emphasis is placed upon how recurrent events vary over time and across treatments in the population. There are many research settings in which individual subjects may repeatedly experience the event of interest within the study period, the occurrences of such events may either follow a probability distribution or may be purely due to chance. In order to mitigate and or eradicate the effects of these occurrences to life and properties, it is deemed expedient to subject events of such nature to empirical studies. Event data analyses have implications and applications to real life situations. Applying the theories behind recurrent events, reveal trends in recurrent event data which invariably leads to the introduction of interventions to curb further occurrences. It is useful for predictions in maintaining optimal maintenance policies in engineering; medical; and biomedical studies. It also has the added advantage of being used as a research design for other research works. The paper is organized as follows: it reviews the dynamics of recurrent events, it considers some theoretical underpinnings of the present study, discussed each of the four models in detail and presented the data layout of each of the models, we did some simulation studies to identify which of the models provided standard results, discussed the results and drew the needed conclusions.

1.1 Dynamics of Recurrent Events

Recurrent event connotes the occurrence of events more than once per subject over the follow up period. In other words, the event of interest can happen multiple times during the subjects' lifetime. In recurrent events, the subjects persist in the risk set pending the completion of the last interval. Subsequent reoccurrence is influenced by previous occurrences, therefore in the analysis, the correlation between the reoccurrence's must be investigated. What characterizes recurrent event is that the same event is observed multiple times with a single subject. Recurrent events data arise in diverse settings with each setting having its own goals; the goal in one setting may relate to the description of the connection of recurrent event rates with the individual's previous experience, or the connection with the history of the covariate that may involve the choices of treatment or repair activities performed on a machine to reduce the risk of reoccurrence. In studies associated with recurrent events, researchers are often interested in underscoring or exploring the effects of covariates on some features of the processes generating the recurrent event data. In this process, the assumption is that the events are independent, more so, there is this assumption that there are no tied event times, furthermore, as long as a subject is under observations, it contributes to the risk set till the event occurs. Referring to this model, the follow-up time is segmented into the time each event occurred, this model can also accommodate time–dependent covariates of any kind. Data obtained from such studies are often referred to as recurrent event data. The structure of the data obtained in such recurrent events forms a special case of multivariate survival data, with the event times ranked for each subject according to order of occurrences. Survival analysis of past and

current occurrences may affect future failure processes. For most requirements, event times are a key factor in the analysis, not only that, but the number of events can affect the process of recurrence [1].

Classical examples of recurrent event data used in modern literature follows: In medical studies, we could have recurring- migraines, seizures, heart attacks, strokes, cancers. HIV patients may experience recurrent opportunistic infections [2]. There could be repeated infections of malaria for individuals living in mosquito prone areas. There could be re-infection of sexually transmitted diseases, there could be multiple infection times among leukemia patients who are undergoing bone marrow transplant. In Reliability studies, there could be repeated breakdowns of machines. brake re-failures, circuit breakers and valve seats [3]. In quality control, researchers could investigate the rate at which production units fail to satisfy consumer taste. Of particular concern to recent researchers is the methodology associated with the modeling of such data. Various models have been recommended in the literature of survival analysis, they include but not limited to the following; marginal intensity models and conditional intensity models [4-5]. Research has shown that for recurrent event data, the mean number of events has more practicable interpretation than the event intensity or the hazard of recurrent event data, to that end researchers have modeled the mean and rate functions subject to the assumption that the covariates interact multiplicatively on the unspecified baseline rate functions. In [6], the simplest modeling approach called the counting process is proposed. Description of counting process could be done using time intervals, event indicators and strata [7]. In another development, semi-parametric additive model for the marginal recurrent event rates were developed under the assumption that the covariate influences were added to the unspecified function of the baseline rate. It has been assumed that in the proportional means and rate models, the covariates are exemplified with a fixed multiplicative effect on the mean as well as the rate function, however, in most of the applications, it is impracticable to consider that the influence of the covariates recorded at the commencement of the study will remain fixed over time, a more acceptable assumption proposed in [8] was that the influences of the covariates converge as time increases; bias and wrong conclusions are the inevitable outcomes of violating the proportional means and rates assumption under recurrent event data.

Reference [9] noted in health research in general, the possibility of two types of events: non-reversible events and reversible events. Non-reversible events allude to events which are chronic in appearance and characterized by their one-time occurrence to an individual. Specific examples include hypertension, AIDS, diabetes and cystic fibrosis. On the other hand, reversible events are acute in nature and are typified by their multiple occurrence to an individual. Reversible events could be further divided into multiple events and recurrent events. Multiple events refer to such repeated events that are somehow related but not exactly the same, cases in point are repeated hospitalization ascribable to myriad underlying reasons such road accident, fall, fever, among others. Distinct from multiple events, recurrent events connote repeated events, which are of the same type, for instance, severe exacerbations in asthmatic patients, seizures in epileptics, patients with low back pain, especially women, skin cancer, fibroids in women, myocardial infractions, and migraine.

They further noted that recurrent events data have two salient features which are: within-subject correlation and time varying covariates. Recurrent events within-subject are less likely independent and are found to be related. Two possible sources have been identified with this within-subject correlation, one relates to event dependency while the other, to event heterogeneity. Within-subject correlation based on event dependency describes a scenario characterized by an acceleration or a deceleration in the rate of subsequent occurrence of an event itself, a case in point is common with the occurrence of an initial heart attack to a subject, there is an increased chance of suffering a second heart attack due to damages that might be caused to some part of the heart by the first heart attack. Within- subject correlation based on heterogeneity describes the circumstances that are exemplified by higher possibility of frequent influences of the event on some subjects than others, underscored by some unconventional or immeasurable reasons. To correct estimation of standard error, it essential to properly adjust within-subject correlation; by treating the correlated observation as uncorrelated, the possibility to overestimate the amount of information provided by each observation is high, thereby resulting in improper estimates of standard errors. An equally significant concern of recurrent event analysis concerns with how to deal with time-dependence covariates. In numerous applications, some covariates are subject to vary over time. In the instance of asthma management, the outcome directly depends on the dosage and type of drugs, which are subject to vary during the course of time.

1.2 Risk Intervals

Risk Intervals define the hazard of obtaining the event of interest within a given time scale. Basically, three formulations for the risk intervals: Counting process; gap time; and total time formulation. The term, counting

process time or calendar time is applied interchangeably to further state the risk intervals to be applied in the regression analysis [10-11]. A counting process also describes a total time scale.

Gap time defines the time in reference to the prior event. In order words, the hazard process (that is, the clock) restarts to zero after every event.

Total time alludes to the time from a chosen point, usually the time from the start of treatment. This implies that the hazard process is defined from some starting point, for instance, the commencement of certain diseases.
In both gap time and counting process formulations, the subject is at risk for the same length of time. The risk interval for the initial event is the same for all other risk intervals.

A marginal or conditional model is determined by the risk interval. The gap time and counting process formulation are both conditional, which is to say that, the subject is at risk conditional on previous events. Total time falls within the marginal domain since the subject is at risk from the commencement of treatment and does not depend on any previous event.

Even though the need has been indicated in [12] to assess statistical methods subject to diverse event creation processes and correlation structures, little attention has been made to the time scale that is employed for further events. It was noted specifically that in simulation-based studies, data is regularly generated from the standpoint of the gap time, in which case the time and for that matter the risk process is reset after each event to simplifie the simulation process. But in many clinical applications the total time perspective is considered appropriate. They noted some discordances between the two approaches and subsequently derived a flexible simulation plan that could improve the accordance between the two approaches.

1.3 Risk Sets

The subjects who are at risk for the kt event are contained in the risk set. Three risk sets have been referred to in the literature: unrestricted; restricted; and semi-restricted. In defining the risk set, we incorporate the choice of baseline hazard. The risk set at any given point in time depends on two conditions: the subjects included in the set; and the risk interval (i.e., when those subjects were at risk). For unrestricted risk set, irrespective of the number of events that affected each subject, the risk intervals of the subjects may account for the risk set for any given event. An unrestricted risk set has a common baseline hazard function for all events. For a restricted risk set, contributions to the kt risk set are limited to include only the kt event risk intervals of those subjects who had already experienced $(k-1)t$ events. For example, only subjects who have experienced two events will be considered to be at risk for the third event. A restricted risk set has event-specific baseline hazard functions. Semi-restricted risk sets have event-specific baseline hazards functions but permit subjects who have encountered fewer than $(k-1)t$ events to be at risk for the kt event through the creation of ‘dummy' risk intervals. Thus, a subject who has encountered none or one event can be considered at risk of a third event. However, a semi-restricted risk set does not allow information from the kt event risk interval to contribute to the risk set for a previous or earlier event. This third kind of risk set applies to the counting process and total time with event-specific baseline hazards [13].

1.4 Theoretical Perspectives Underpinning the Study

Reference [14] underscored the fact that many different statistical techniques which exist for analyzing recurrent events data could be divided into naive and longitudinal techniques. The “Naive techniques” were explained to have been characterized either by ignoring the existence of recurrent events or disregarding the fact that the recurrent events within subjects were correlated, while the longitudinal techniques are characterized by the fact that the whole trend of recurrent events over time is analyzed considering that the recurrent events were correlated within subjects. Albeit the existence of many statistical techniques for analyzing recurrent event data, [14] expressed surprise that most researchers rather found it uneasy to select the proper technique to answer the research question of interest. The naive statistical technique is most often selected to analyze the outcomes of studies. They proposed an overview in view of basic applicable statistical approaches for analyzing recurrent event data; to enable comparison of results of the naive and longitudinal approaches; and to outline some recommendations in connection to analyzing recurrent event data, subject to a given research question.

Reference [15] aimed at evaluating the performance of known recurrent event models given specific data situation of a complex endpoint which exhibited the outlined features.

- For each event type, be it recurrent or terminal, a separate event processes exists which might be correlated or otherwise;
- A tendency of deviation by the event-specific treatment effects which associate with the diverse types of events;
- There is a surge in the instantaneous baseline risk for a subsequent event (fatal or non-fatal) after an event had occurred;
- The time the previous event occurred determines the instantaneous risk for a subsequent event; and
- The relative treatment effect for a subsequent event (in terms of the hazard ratio) may undergo alteration precedent to the occurrence of an event.

They identified three most frequently used models for analyzing recurrent event data. The first is the Andersen and Gill model underpinned by the basic Cox proportional hazards model. The simple axiomatic consideration is independence between all observed event times regardless of whether the event times corresponded to the same subject. The second is the Prentice, Williams, and Peterson model which emphasizes the order of events. There are two different time scales needed to handle this model, the gap time and the total time scales. The gap time perspective evaluates the time since the last event occurred, while the calendar or total time scale emphasizes the time of entry of subjects into the study. The third is the unconditional marginal model which was formulated by Wei, Lin, and Weissfeld. The order of occurrence of the events is disregarded in this model. As such, for each subsequent event, the tendency of all individuals to be at risk is not determined by a preceding event. The model by Wei et al. operates within the context of a total time scale. The focus of [15] was the comparison of the three most common models of analyzing recurrent data. The investigation was done using two different data sets - one recurrent non-fatal event and the other a fatal event. The comparison was accomplished in consideration of the statistical features of the models' treatment effect estimator and their appropriate interpretation. This comparison was done to enable them to proffer suggestions in the selection of a suitable analysis model which could address specific data structure of clinical trials with composite endpoints. A total of *n* subjects was assigned in a 1:1 ratio to the experimental group (E) and to the control group (C). The covariate X_i (which equals 1 whenever the patient belongs to the experimental group and 0 otherwise), was used to denote the assignment of subjects into groups. Each individual $i = 1,2 \ldots n$ could encounter up to $j = 1,2 \ldots k$ events of the same or of distinct types. In this case, k which was the maximum number of events considered per subject was restricted for simplicity sake. Their approach for the occurrence of the event exemplifies a multi-state model, that is to say, a subject entered the study at an initial state 0. Every time an event is experienced, the individual transits the previous state and enters a new event state. In the instance that the event observed was non-fatal, the subject could encounter more subsequent non-fatal events or the fatal events. The instantaneous risk to encounter a jt event at time t given that the subject has already experienced $j-1$ non-fatal events was parameterized. In [16], two methods of analyzing recurrent events were described: non-parametric and semiparametric methods. The attention was on the functions that were formulated in the analysis. It was indicated that the structure of data for recurrent events signified a distinct case of multivariate survival data, where ordering was considered in the failure times for a subject. Hence, recurrent event data is frequently analyzed applying techniques of multivariate survival analysis. Five models, which are Cox-based for recurrent event data were underscored in [17]: Andersen and Gill (AG); Wei, Lin and Weissfeld (WLW); Prentice, Williams and Peterson, total time (PWP-TT) and gap time (PWP-GT); and Lee, Wei and Amato (LWA). It was mentioned that some attempts have been made by some authors in comparing these models by detecting differences that arose in the event of fitting the models to real and simulated data. It was opined that no author had attempted to analytically discover the components of the models that are suitable for recurrent event data. They proposed a systematic way of classifying such Cox-based models considering four major components: risk intervals; baseline hazard; risk set, and correlation adjustment. In their perspective of risk interval and risk set, they conceptualized two new modified models known as: 'total time – restricted' (TT-R) and 'gap time – unrestricted' (GT-UR) models. In their study they determined the appropriateness of each model in respect of the salient components. The models were fitted to simulated data sets and to a real data set of childhood recurrent infectious diseases. It was concluded that the LWA model was not suitable for recurrent event data because it allowed a subject to be at risk several times for the same event. The WLW model was said to have overestimated treatment effect and was not recommended. It was then recommended that the PWP-GT and TT-R were useful models for analyzing recurrent event data. Reference [18] has underscored the fact that unsuitable or inefficient statistical techniques are still applied to analyze recurrent or repeated event data, notwithstanding

the numerous discussions or studies on the choice of efficient methods for analyzing such data. They referred the most popular technique for analyzing survival data to the Cox proportional hazards model, which as a result of the independence assumption, was only efficient for modeling the time to the first event. They indicated that this model was inefficient because data from later events were ignored. Another technique for modeling the number of events for each subject according to [18] was to fit Poisson or negative binomial models, which were incorporated into the generalized estimating equations (GEE) in recent times. The third approach they mentioned was the random effects models which considered the correlation of events, but not without its own inefficiencies like the other models. They climaxed their discussion by alluding to the fact that variants of the original Cox model have been proposed to analyze recurrent event data: Andersen-Gill (AG); Prentice, Williams and Peterson (PWP) (total and gap times); Wei, Lin and Weissfeld (WLW); and frailty models. A further analytical technique was to model the mean number of events or the rate of their occurrence. They concluded by showing that, multi-state models (MSM) have currently been extended for recurrent events. Reference [19] mentioned that over the past few decades, lots of statistical advancements had evolved in the field of recurrent event data analysis. The newly established techniques are more efficient than classical statistical methods (such as t-test, logistic regression, multiple linear regression, and Cox's Proportional Hazard regression), just to mention a few. It was revealed that regardless of the several superior methods available for analyzing recurrent event data, most researchers still relied on the classical statistical approaches to analyze their research questions even though the outcome of interest was recurrent in nature. It was indicated that relying on sub-superior suitability could result in loss in respect of internal validity and precision of the results.

Reference [20] considered outcome events that may occur repeatedly over the follow-up time for a given subject, known in the literature as "recurrent events." They surmised that modeling such data type could be performed via a Cox proportional hazard model with the data structure well-organized to assign each subject to a line of data that correspond to each recurrent event. They noted that a variant of this technique employed a stratified Cox proportional hazard model, which stratifies according to the order of occurrence of the recurrent events. It was concluded that independent of the method was applied, the researcher should consider adjusting the variances of estimated model coefficients to cater for the likely correlation among recurrent events on the same subject. Such controlled variance estimates are called "robust variance estimates." A parametric method for analyzing recurrent event data that encompass a frailty component was also described and illustrated.

Our purpose in this current study is to undertake another extensive investigation using a different dataset of recurrent Bladder Cancer Tumors [26-27] to compare, evaluate and discuss pros and cons of four models of recurrent events; The Andersen and Gill model; Prentice, Williams and Peterson models; Wei, Lin and Weissferd*;* and Cox frailty model. This will aid to further illustrate the characteristics of these four models so as to consolidate and validate previous findings in respect of the superiority or appropriateness of these models.

2 Materials and Methods

2.1 Study design

A simulation-based comparison of recurrent event models was applied to tertiary data on cancer studies. Data was extracted from the first forty (40) subjects on the study of recurrent Bladder Cancer Tumors [21-22] reproduced in the book authored by [20]. The entire data comprises 86 patients who suffered superficial bladder tumors, which were removed at the time of patients' recruitment or entry into the study. The group assignment of patients are as follows: 48 were randomized into the placebo group, and 38 were randomized into the group receiving thiotepa. Multiple recurrences of tumors were encountered by many patients during the study, and new tumors were removed at each visit. The data set is characterized by the first four repetitions of the tumor for each patient, and each repetition time was recorded from the time of entry of patients into the study. A separate risk interval is defined by each time to an event or censoring. Extending the database as mentioned above for purposes of exemplification will assist the researchers to further shed more light on the points that have been raised by earlier researchers.

2.2 Statistical Model Formulation

Four different models were applied to model recurrent event data, we propose that the final choice we make must depend on the type of data and research question.

The first model discussed was the counting process model, an assumption of dependency was coffered on each event. A subject contributed to the risk set for an event as once it was under observation at the time of the occurrence of the event. We described the data for each subject with recurrence events as data for multiple subjects where each subject experienced a delayed entry and was monitored until the occurrence of the next event. We ignored the order of the events, allowing each subject to be at risk for any event as long as they were under monitoring. Implicitly, a subject could be at risk for a subsequent event without having experienced prior events. We used the same time as that of total time and recognized that a subject may encounter a delayed entry or censored period before the subject be at risk for the event. We again noted that in the counting process formulations, subjects were at risk for the same duration of time.

The second model we considered was conditional (referred to in some literature as conditional 1 or A). It is so called because of the assumption that it was impossible for a subject to be at risk for a next event without experiencing the previous event. A strata variable was applied to represent the event number. We let the end of the time interval for the previous event marks the starting point of the time interval for a subsequent event. This model helped us to model the full-time course of the recurrent event.

The third model we considered was also conditional (referred to in some literature as conditional 2 or B). The only distinctive feature of this model from the conditional 1 model can be perceived in the manner of structure of the time intervals. We started each time interval at zero and ended at the length of time until the time of occurrence of the next event. Implicitly, the risk sets for each of the events were completely distinct. This model demonstrates relevance in modeling the time interval between each of the recurring event rather than the full-time course of the recurrent event process.

The fourth model we considered was the marginal model. We considered each event as an isolated procedure. The time for each event began at the commencement of the follow-up time for each subject. For all events, ach subject was considered to be at risk, independent on the number of events each subject actually experienced. Thus, available data for the specific events was modelled in this recurrent event modelling technique by considering each event separately.

2.3 Models of Recurrent Events

Starting from a time origin which is well defined, we will observe the points $T_1, T_2, \ldots, T_n$ for a given subject, where T_1 is the time to the subject's first failure, T_2 is the time from the origin to the subject's second failure with follow-ups continuing to a total follow-up time *C* that right censors the point process. There may also exist a covariate vector or covariate process *x*, which is a function of *u* with history $X(t) = \{x(u), 0 \le u < t\}$ for the subject. We represent by $\tilde{N}(t)$, the number of failures on the subject with follow up time *t*, and by $N(t)$, the corresponding observed number of failures in (0, t]. Due to censoring, $N(t)$ may be less than $\tilde{N}(t)$. Recurrence rates that condition on the preceding failure and covariate histories for the subject constitutes a reasonable starting point for modeling recurrent event data. For absolutely continuous event times, we can define the hazard or intensity process $\lambda(t)$ at a follow-up time t, given the covariate history $X(t)$, by

$$\lambda(t)dt = P[d\tilde{N}(t) = 1 \mid \tilde{N}(u), 0 \le u < t, X(t)]. \tag{1}$$

The expression in (1) assumes that jumps in $\tilde{N}$ are of unit size only, nevertheless, recurrent event data sometimes involve jumps of size greater than one, as more than one event is measured for an individual at a specific follow-up time. This particularly happens due to the fact that event times from a fundamental

continuous process are grouped. In addition, it is possible to assume continuous-time counting processes having jump sizes greater than one, a case in point may relate to queuing models where the arrival time of customers in continuous time are counted. However, where customers arrive in groups of various sizes, one natural approach to modeling counting processes with jumps that may exceed one is to model the mean jump in $\tilde{N}$ across time. Models that cater for such increments are presented below;

$$d\Lambda(t) = E[d\tilde{N}(t) \mid \tilde{N}(u), 0 \le u < t, X(t)]. \tag{2}$$

where Λ is the cumulative intensity process, Equation (2) is referred to as the failure intensity at time *t*, and is defined as the expected number of events on a subject, Equations (1) and (2) are equivalent to

$$\Lambda(t) = \int_0^t \lambda(u)\, du, \tag{3}$$

for cases of continuous-time processes with only unit jumps. Whenever reoccurrence times are restricted to be absolutely continuous, we assume that $d\tilde{N}(t) \le 1$. In many applications, attention is often placed on the assessment of the effects of covariates on the marginal recurrent event rates. The multiplicative rates model is written as:

$$E[d\tilde{N}(t) \mid X(t)] = \exp\{\beta_0' X(t)\}\lambda_0(t)dt. \tag{4}$$

Where β_0 is a p-vector of unknown regression parameters and λ_0 is an unspecified baseline rate function. The additive rates model takes the form;

$$E[d\tilde{N}(t) \mid X(t)] = \beta_0' X(t)dt + \lambda_0(t)dt. \tag{5}$$

Equation (5) was subjected to regression analysis in [23], where the regression parameters and the baseline rate were obtained. Additionally, it was shown that the proposed estimators were consistent and asymptotic Gaussian [24].

The Cox proportional hazards model could be generated from Equations (2) and (3):

$$\Lambda(t) = \int_0^t \lambda(u)\, du$$

$$d\Lambda(t) = d\Lambda_0(t)\, exp[(Y(t)'\beta] \Longrightarrow d\int_0^t \lambda(u)\, du = d\int_0^t \lambda(u)\, du.\, exp[(Y(t)'\beta], \tag{6}$$

where $Y(t)' = Y_1(t), Y_2(t), \ldots Y_P(t)$ are made up of functions of $X(t)$ and $[N(u); 0 \le u < t]$, and product terms with *t.*

When Equation (6) is calculably manipulated, we obtain

$$\lambda(t) = \lambda_0(t)\, exp[(X(t)'\beta]. \tag{7}$$

Where $X(t)' = X_1(t), X_2(t), \ldots X_i(t)$.

.

Five methods of modelling recurrent events have been discussed in [25-26]: Andersen and Gill model; Prentice, Williams and Peterson models; marginal means/rates model; frailty model; and multi-state models. We present brief description of four of the models.

2.3.1 The Andersen and Gill model (Counting process)

Andersen-Gill (AG) uses the counting process structure of data inputs. Each subject is represented as a series of observations with recurrent times given as $(t_0, t_1]\,(t_1, t_2]\,...\,(t_{n-1}, t_n]$. Each recurrent event for the it subject; $i = 0,1,2\,...\,n$; is assumed to follow the proportional hazards model. The outcome of interest is time since randomization for a treatment until an event occurs. It uses a common baseline hazard function for all events and estimates a global parameter for the factors of interest. The Andersen and Gill (AG) model assumes that the correlation between event times for a subject can be explained by past events, which implies that the time increments between events are conditionally uncorrelated, given the covariates. In simple terms we use this counting process when each event is assumed to be independent, moreover, a subject contributes to the risk set for an event as long as the subject is under observation at the time the event occurs. This model ignores the order of the events leaving each subject to be at risk for any event as long as they are still under observation at the time of the event. This further means that a subject could be at risk for a subsequent event without having experienced the prior event. The hazard function is given as:

$$\lambda_i(t) = \lambda_0(t)\exp\{\beta_k x_i(t)\}. \tag{8}$$

Under this model, the risk of recurrent event for a subject follows the Cox proportional hazards model assumption, but the number of recurrent events is not taken into consideration.

2.3.2 Prentice, Williams and Peterson models 1981 (Conditional 1 and 2)

The Prentice, Williams and Peterson (PWP) model do the analyses of ordered multiple events by stratification according to the prior number of events during the follow-up period. All subjects are deemed to be at risk for the first stratum, but only those with an event in the previous stratum are at risk for the successive one. It is deemed impossible for a subject to be at risk for a subsequent event without having experienced the previous event (in other words, one cannot be at risk of event 2, without being at risk of event 1). The model can incorporate both overall and event-specific effects for each covariate. In practice, the data may need to be limited to a specific number of recurrent events if the risk set becomes very small for later strata and event-specific estimates become too unreliable. Besides using the same outcome (total time: TT) as in the AG model, the PWP model can also be usually defined in terms of gap time (GT), which is the time since the previous event occurred. When using a gap or waiting-time scale, the time index is reset to zero after each recurrence of the event, with assumption of a renewal process. Gaps between events are often useful with infrequent events, when a renewal occurs after an event or when the interest lies on prediction of a next event. Hence, two stratified PWP models can be fitted: PWP-TT, which evaluates the effect of a covariate for the kt event since the entry time in the study; and the PWP-GT, which evaluates the effect of a covariate for the kt event since the time from the previous event. Unlike the AG model, the effect of covariates may vary from event to event in the stratified PWP models. Therefore, the PWP models might be preferable to the AG model when the effects of covariates are different in subsequent events. The PWP -TT model for the hazard function for the kt event for the it subject is given as:

$$\lambda_{ik}(t) = \lambda_0(t)\exp\{\beta_k x_i(t)\}. \tag{9}$$

The baseline hazards vary from event to event.

The PWP - GT model for the hazard function for the $(k-1)t$ event, a subject is not considered in the risk set for the kt event until it experiences the $(k-1)t$ event. The hazard function is given as:

$$\lambda_{ik}(t) = \lambda_0(t - t_{k-1})\exp\{\beta_k x_i(t)\}. \tag{10}$$

2.3.3 The marginal means/rates model (Wei, Lin and Weissferd, 1989)

This is an alternative model for analyzing recurrent events. This model can be interpreted in terms of the mean number of events when there are no time-dependent covariates. This approach does not specify dependence structures among recurrent event times within a subject. However, since the marginal means/rates model considers all recurrent events of the same subject as a single counting process and does not require time-varying covariates to reflect the past history of the process, this model is more flexible and parsimonious than AG

model. If no time-dependent covariates are included in the AG model to account for all the influence of the prior events on future recurrences, point estimates from the means/rates model and the AG model will be the same. Nevertheless, the covariance matrix estimate for the regression coefficients for the marginal means/rates model uses score residuals in the middle of the sandwich estimate, which corrects for the dependency structure. This approach can be of interest in many medical applications when the dependence structure is complex and unknown, especially when it cannot be characterized by including time-varying covariates, as in the AG model. In the hazard model, it is of interest to note that all the time intervals start at zero. The hazard model allows a separate underlying hazard for each event. When an event is set to zero, it means that subject is no longer at risk after the last given interval. The hazard is given by

$$\lambda_{ik}(t) = \lambda_0(t)\exp\{\beta_k x_i(t)\}. \quad (11)$$

2.3.4 The Cox frailty model

The frailty model also known as the random effects approach, introduces a random covariate into the model that induces dependence among the recurrent event times. The idea is that the random effect describes excess risk or frailty for distinct subjects, taking into account unmeasured heterogeneity that cannot be explained by observed covariates alone. The most commonly used frailty model is a shared frailty model with random effects assumed to follow a gamma distribution with mean equal to one and unknown variance. The model assumes that the recurrent event times are independent conditional on the covariates and random effects. When there is heterogeneous susceptibility to the risk of recurrent events, the frailty model can be applied. The Hazard function $\lambda_{ik}(t)$ for the recurrent time of the kt□ event of the subject it□ ($j = 1.2 \ldots k, i = 1,2 \ldots n$), conditional on the fraity Z_k, follows the proportional hazards model form and is given by:

$$\lambda_{ik}(t) = \lambda_{0k}(t)\, Z_i\{\exp\{\beta_k x_i(t)\}, t > 0. \quad (12)$$

Where $\lambda_{0k}(t)$ is the common baseline hazard function x_i is a vector of observable covariates, β is the vector of unknown regression coefficients, frailty Z_i is the unobserved common risk factors shared by all subjects in cluster i and is assumed to be fixed with unit mean and unknown variance.

2.4 Determining the Appropriate Model to Use

In addressing the question as to which model to choose in analyzing and modeling recurrent event data, there appears to be no end in sight as to which of the models is robust. Several authors have proposed one proportional hazards model over the other. Reference [27] had proposed that the choice of the most appropriate model depended on:

- Distribution of subsequent event times;
- Within person correlation of subsequent events;
- Frequency of occurrence; and
- Specific research questions.

In [17], a systematic way of exemplifying the Cox-based models using four salient components has been proposed. The components include risk intervals; baseline hazard; risk set, and correlation adjustment.

To help us to appreciate the model data layout, we will go through four different outlines for the start times and finish times of counting process, conditional 1, conditional 2 and marginal recurrent models. It should be noted that the choice of the appropriate model depends on how one defines the model characteristics, risk interval and risk set.

We will approach this by looking at a hypothetical data on two subjects 1 and 2. These subjects were observed for 33 weeks, the event of interest was time until reinfection of sexually transmitted disease. When a subject gets reinfested, he is immediately treated and discharged. It is assumed that subject 1 had reinfection on the following weeks of follow up: 10th, 15th and 31st weeks. Subject 2 experienced the event of interest at the 3rd and 12th weeks of follow-up and did not experience any event till the end of follow up. If a subject obtained an event of interest (E), $E = 1$, if a subject fails to obtain the event of interest, $E = 0$. It is assumed that each

recurrent event is independent of the previous event, thus a stratum number (from 1 to 4) was accorded each event. This number was used to track the number of separate events that have occurred within the follow-up time. The data layout for subjects1 and 2 are shown in Table 1 below under the four recurrent event models:

Table 1. Hypothetical data layout for two subjects (Subjects 1 and 2) with recurrent event times 10, 15, 31 and 3, 12 weeks respectively after 33 weeks of follow-up under four recurrent event models

Model	Subject 1			Subject 2		
	Time Interval	Event	Stratum	Time Interval	Event	Stratum
The Andersen and Gill model (Counting process)	(0,10]	1	1	(0,3]	1	1
	(10,15]	1	1	(3,12]	1	1
	(15,31]	1	1	(12,33]	0	1
	(31,33]	0	1			
Prentice, Williams and Peterson (Total time, counting process, Conditional 1 or A)	(0,10]	1	1	(0,3]	1	1
	(10,15]	1	2	(3,12]	1	2
	(15,31]	1	3	(12,33]	0	3
	(31,33]	0	4			
Prentice, Williams and Peterson (Gap time - Conditional 2 or B)	(0,10]	1	1	(0,3]	1	1
	(0,5]	1	2	(0,9]	1	2
	(0,16]	1	3	(0,21]	0	3
	(0,2]	0	4			
Wei, Lin and Weissferd (Marginal)	(0,10]	1	1	(0,3]	1	1
	(0,15]	1	2	(0,12]	1	2
	(0,31]	1	3	(0,33]	0	3
	(0,33]	0	4	(0,33]	0	4

2.5 R Codes for all Models

The software for simulation was run on an x64 windows-based machine with the following specification: operating system: MS Windows 10; processor: 1.4GHz, and RAM Size: 4 GB.

Software: All analysis were performed in R Software (R version 4.1.1). Packages installed were survival, survminer, and simrec

```
library(survival)
library(survminer)
library(simrec)
bcs <- read.csv ("flie_location/filename.csv" use.value.labels=TRUE)
```

2.5.1 AG model

```
AG <- coxph(Surv(Start,Stop, Event) ~ tx + num + size +cluster(id), robust=TRUE, data = bcs)
summary (AG)
Call:
coxph(formula = Surv(Start, Stop, Event) ~ tx + num + size, data = bcs,
   robust = TRUE, cluster = id)
 n= 80, number of events= 50
```

Stratification Models: The models below are stratified Models. The argument strata(intcount) identify the stratification variable to obtain their estimates. Estimates are obtained for event-specific covariates.

2.5.2 PWP-Total time model

```
PWP_TT <- coxph(Surv(Start,Stop, Event) ~ tx + num + size +cluster(id) + strata(int_count), data = bcs)
summary (PWP_TT)
Call:
coxph(formula = Surv(Start, Stop, Event) ~ tx + num + size +
```

```
   strata(int_count), data = bcs, cluster = id)
  n= 80, number of events= 50
```

2.5.3 PWP Gap time model

```
PWP_GP <- coxph(Surv(Stop - Start, Event) ~ tx + num + size +cluster(id) + strata(int_count), data = bcs)
summary (PWP_GP)
Call:
coxph(formula = Surv(Stop - Start, Event) ~ tx + num + size +
    strata(int_count), data = bcs, cluster = id)
  n= 80, number of events= 50
```

2.5.4 Marginal model

```
Marginal_M <- coxph(Surv(Start,Stop, Event) ~ tx + num + size +cluster(id) + strata(int_count), data = bcs)
summary (Marginal_M)
Call:
coxph(formula = Surv(Start, Stop, Event) ~ tx + num + size +
    strata(int_count), data = bcs, cluster = id)
  n= 80, number of events= 50
```

2.5.5 Frailty model

By default, gamma distribution is associated with the random effect for frailty model in R. However, gaussian distribution may be specified.

```
 Frailty_M <- coxph(Surv(Start,Stop, Event) ~ tx + num + size + frailty(id, distribution = "gamma"), data = bcs)
summary (Frailty_M)
Call:
coxph(formula = Surv(Start, Stop, Event) ~ tx + num + size +
    frailty(id, dist = "gamma"), data = bcs)
  n= 80, number of events= 50
```

The R codes for the various models are used to simulate the data in Table 2.

Table 2. The first forty (40) Subjects from Bladder Cancer Study (Byer,1980, and Wei Lin, Weissfield, 1989 reproduced in the book authored by [20]

id	int_count	Event	Start	Stop	tx	num	size
1	1	0	0	1	0	1	1
2	1	0	0	1	0	1	3
3	1	0	0	4	0	2	1
4	1	0	0	7	0	1	1
5	1	0	0	10	0	5	1
6	1	1	0	6	0	4	1
6	2	0	6	10	0	4	1
7	1	0	0	14	0	1	1
8	1	0	0	18	0	1	1
9	1	1	0	5	0	1	3
9	2	0	5	18	0	1	3
10	1	1	0	12	0	1	1
10	2	1	12	16	0	1	1
10	3	0	16	18	0	1	1
11	1	0	0	23	0	3	3
12	1	1	0	10	0	1	3
12	2	1	10	15	0	1	3
12	3	0	15	23	0	1	3

id	int_count	Event	Start	Stop	tx	num	size
13	1	1	0	3	0	1	1
13	2	1	3	16	0	1	1
13	3	1	16	23	0	1	1
14	1	1	0	3	0	3	1
14	2	1	3	9	0	3	1
14	3	1	9	21	0	3	1
14	4	0	21	23	0	3	1
15	1	1	0	7	0	2	3
15	2	1	7	10	0	2	3
15	3	1	10	16	0	2	3
15	4	1	16	24	0	2	3
16	1	1	0	3	0	1	1
16	2	1	3	15	0	1	1
16	3	1	15	25	0	1	1
17	1	0	0	26	0	1	2
18	1	1	0	1	0	8	1
18	2	0	1	26	0	8	1
19	1	1	0	2	0	1	4
19	2	1	2	26	0	1	4
20	1	1	0	25	0	1	2
20	2	0	25	28	0	1	2
21	1	0	0	29	0	1	4
22	1	0	0	29	0	1	2
23	1	0	0	29	0	4	1
24	1	1	0	28	0	1	6
24	2	1	28	30	0	1	6
25	1	1	0	2	0	1	5
25	2	1	2	17	0	1	5
25	3	1	17	22	0	1	5
25	4	0	22	30	0	1	5
26	1	1	0	3	0	2	1
26	2	1	3	6	0	2	1
26	3	1	6	8	0	2	1
26	4	1	8	12	0	2	1
26	5	0	12	30	0	2	1
27	1	1	0	5	0	1	3
27	2	1	5	18	0	1	3
28	1	1	0	3	0	2	1
28	2	0	3	10	0	2	1
28	3	0	10	22	0	2	1
29	1	0	0	18	0	1	4
30	1	1	0	10	0	1	2
30	2	1	10	20	0	1	2
31	1	1	0	25	0	3	1
32	1	0	0	12	0	1	1
33	1	1	0	3	0	1	5
33	2	1	3	14	0	1	5
33	3	0	14	20	0	1	5
34	1	1	0	13	0	3	1
35	1	1	0	8	0	1	1
35	2	1	8	26	0	1	1
36	1	0	0	4	0	2	1
36	2	0	4	12	0	2	1
37	1	0	0	11	0	1	3
37	2	1	11	14	0	1	3
37	3	1	14	26	0	1	3

id	int_count	Event	Start	Stop	tx	num	size
38	1	1	0	25	0	1	1
39	1	1	0	9	0	2	1
39	2	1	9	20	0	2	1
40	1	0	0	1	0	1	2
40	2	1	1	11	0	1	2
40	3	1	11	26	0	1	2

Note: Subjects with no event have a single observation. for instance, subject id from 1 to 5,7,8,11,17,21,22,23,29,31,32,34, and 38 with start time equals to 0 and stops equal to follow-up time, whiles subjects with at least one event have two or more rows (ids 6, ,9,10,12,13,14, 15,16,18,19,20,24,25, 26,27, 28,30,33,35,36,37,39 and 40)

The patient with id 2 has a censored time at a month (Stop=1). The patient with id number 6 on the other hand had an event at time t = 6 (event = 1) and censored at time t=10 (event=0).
Similarly, the subject with id number 16 had an event on the 3rd month and repeated on the 15th month and censored at the 25th month.

Variables in solution

- Id = Patient Identity which identifies the patient;
- Intcount= Number of observations;
- Num = initial number of tumors;
- Size =initial size of tumors in centimeters;
- Event= event status (0=censored, 1=event);
- Tx =Treatment (0=placebo, 1=thiotepa);
- Start = Beginning of an interval where patient is at risk for an event (in months);
- Stop = End of interval due to an event (1) or censoring (0) (in months); and
- The Start, Stop is used to define the time interval of risks

3 Results and Discussion

A pictorial view of the recurrent events is presented (Figure 1). The output is a plot of recurrent data. A bullet (•) indicates a recurrent event while a circle(o) indicates censoring. Each time interval (in months) starts at zero and ends at a length of time until the next event.

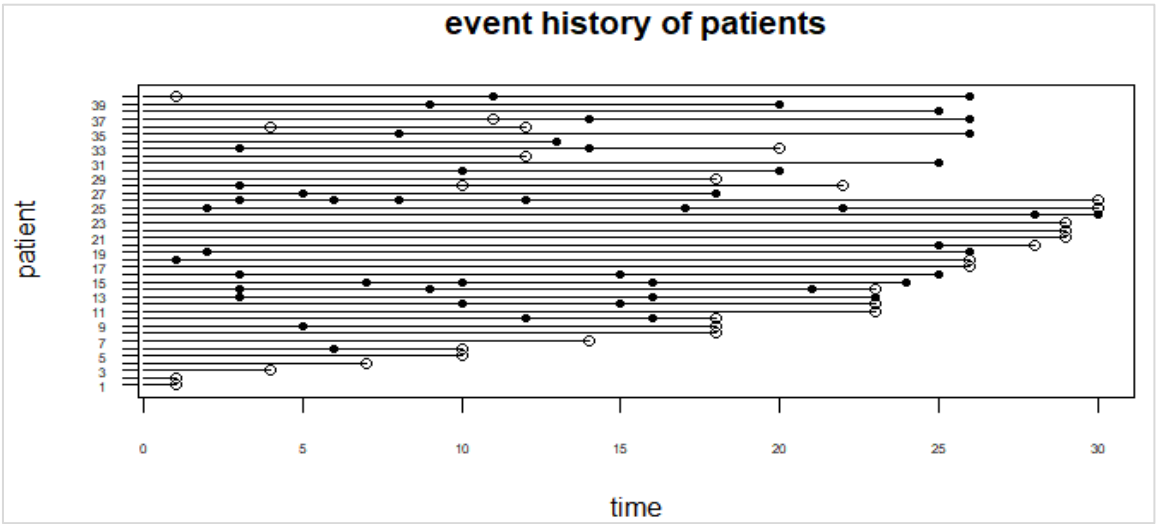

Fig. 1. Output of recurrent events of all 40 subjects

We note that the robust standard error for the initial number of tumors (row 2 for each model, column 5) for AG; PWP-TT; PWP-GT; and Marginal models were respectively: 0.08; 0.07; 0.06; and 0.07, the standard errors were approximately the same for all the four models. In respect of the initial size of tumor (row 3 for each model, column 5) for AG; PWP-TT; PWP-GT; and Marginal models, the robust standard errors were respectively: 0.09; 0.10; 0.08; and 0.10, we note again that the standard errors did not differ considerably from each other. The hazard ratios for the initial number of tumors and size of tumor (row 2 for each model, column 8) are interpreted model by model as follows: For AG, the hazard ratio for the initial number of tumors indicates that for every unit increase in number of tumors, the rate of bladder cancer episode increases by 10%, the hazard ratio for initial size of tumor indicates that the subjects in the control group experience bladder cancer episodes at a rate which is 104% higher than the treatment group. For PWP-TT, the hazard ratio estimates for initial number of tumors indicate that for every unit increase in number of tumors, the rate of bladder cancer episode increases by 6%, the hazard ratio for initial size of tumor indicates that the subjects in the control group experience bladder cancer episodes at a rate which is 101% higher than the treatment group. For PWP-GT, the hazard ratio estimates for initial number of tumors indicate that for every unit increase in number of tumors, the rate of bladder cancer episode increases by 10%, the hazard ratio for initial size of tumor indicates that the subjects in the control group experience bladder cancer episodes at a rate which is 102% higher than the treatment group. For the Marginal model, the hazard ratio estimates for initial number of tumors indicate that for every unit increase in number of tumors, the rate of bladder cancer episode increases by 6%, the hazard ratio for initial size of tumor indicates that the subjects in the control group experience bladder cancer episodes at a rate which is 101% higher than the treatment group. (Table 3A for AG; PWP-TT; PWP-GT and Marginal) and (Table 3B for Frailty models).

Table 3A. Results generated from the R platform for four models for analyzing recurrent events

	Coef.	**exp(coef)**	**se(coef)**	**robust se**	**z**	**Pr(>|z|)**	**Exp (-coef)**	**lower .95**	**upper .95**
The Andersen and Gill model (Counting Process) AG									
tx	N/A	N/A	0.00	0.00	N/A	N/A	N/A	N/A	N/A
num	-.09	0.91	0.12	0.08	-1.12	0.26	1.10	0.77	1.07
size	-.04	0.96	0.10	0.09	-0.41	0.68	1.04	0.81	1.15
Prentice, Williams and Peterson models1981 (Conditional 1) PWP- Total Time									
tx	N/A	N/A	0.00	0.00	N/A	N/A	N/A	N/A	N/A
num	-.06	0.94	0.11	0.07	-0.79	0.43	1.06	0.82	1.09
size	-.01	0.99	0.11	0.10	0.49	0.96	1.01	0.81	1.22
Prentice, Williams and Peterson models1981 (Conditional 2) PWP- Gap Time									
tx	N/A	N/A	0.00	0.00	N/A	N/A	N/A	N/A	N/A
num	-.09	0.91	0.11	0.06	-1.42	0.16	1.10	0.80	1.04
size	-.02	0.98	0.11	0.08	-0.23	0.82	1.02	0.83	1.15
The Marginal Means/Rates Model (Wei, Lin and Weissferd, 1989) WLM									
tx	N/A	N/A	0.00	0.00	N/A	N/A	N/A	N/A	N/A
num	-.06	0.94	0.11	0.07	-0.79	0.43	1.06	0.82	1.09
size	-.01	0.99	0.11	0.10	-0.45	0.96	1.01	0.81	1.22

Table 3B. Results generated for Frailty model for analyzing recurrent events

	The Frailty Model								
	Coef	**se(coef)**	**se2**	**Chisq**	**df**	**p**	**exp(-coef)**	**lower .95**	**upper .95**
tx	.00	0.00	1.00				N/A	N/A	N/A
num	-.09	0.12	0.12	0.67	1.0	0.41	1.10	0.73	1.14
size	-.04	0.10	0.10	0.12	1.0	0.73	1.04	0.79	1.18

We have the results from the R platform showing the various significant test results (Tables 4A and 4B). Of particular interest now is the results listed at row one of Table 4A, the standard error (se) of the models AG; PWP total time; PWP gap time and marginal are given respectively as: se = 0.047; se = 0.055; se = 0.045; se = 0.055. A small robust standard error means there is more variation within subjects than between subjects, while a big robust standard error means there is less variation within subjects than between subjects. From the simulated results, the robust standard errors of all four models were small regardless of whether events were

independent within subjects. On the basis of the ongoing analysis, we note that the results from the PWP-GT model happens to have the smallest standard error. We are of the candid opinion that if the within subject events are independent, then we need to use the PWP-GT model to analyze recurrent event data. As has been stated earlier, the choice of a model depends to a large extent on the research question and type of data available. The PWP-GT model could be used if the research question purports to find out whether the treatment was effective for the □□□ event since the previous event happened. However, if the research question is designed to find out whether treatment was effective for the □□□ event since the start of treatment, then we could use the PWP- TT. The AG model will be adequate if a common baseline hazard could be assumed, but the model lacks the details and versatility of the event-specific models. Applying a robust variance may not be adequate when there is within-subject correlation. The WLW model is very appropriate for data where there are different types of events for the same person, and the baseline hazard is potentially different for each type, such as multi-type event data. The WLW model overestimates treatment effect and is not recommended.

Table 4A. Concordance, Likelihood, Walt test, Score test and Robust Results

	AG	PWP- Total time	PWP- Gap time	Marginal
Concordance	0.5(se = 0.047)	0.44(se = 0.055)	0.44(se = 0.045)	0.44(se = 0.055)
Likelihood ratio test	0.75,2 df, p=0.7	0.32, 2 df, p=0.9	0.74, 2 df, p=0.7	0.32, 2 df, p=0.9
Wald test	1.25,2 df, p=0.5	0.65, 2 df, p=0.7	2.06, 2 df, p=0.4	0.65, 2 df, p=0.7
Score (logrank) test	0.68,2 df,p=0.7,	0.03,2 df, p=0.9	0.67,2 df, p=0.7	0.03, 2 df, p=0.9
Robust	0.97 p=0.6	0.48 p=0.8	1.28 p=0.5	0.48, p=0.8

Note: *For all Models: The likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not.*

Table 4B. Concordance, Likelihood Results generated from the for the Frailty Model

Variance of random effect	= 0.2007005 I-likelihood = -161.4
Degrees of freedom for terms	= NaN 1 1 0
Concordance	= 0.52 (se = 0.05)
Likelihood ratio test	= 0.76 on NaN df, p=NA

Iterations: 9 outer, 37 Newton-Raphson

Fig. 2 gives the survival experiences using the four different models, while Figs. 3 and 4 give pictorial views of the hazard's ratios of the first ten subjects from the bladder cancer data.

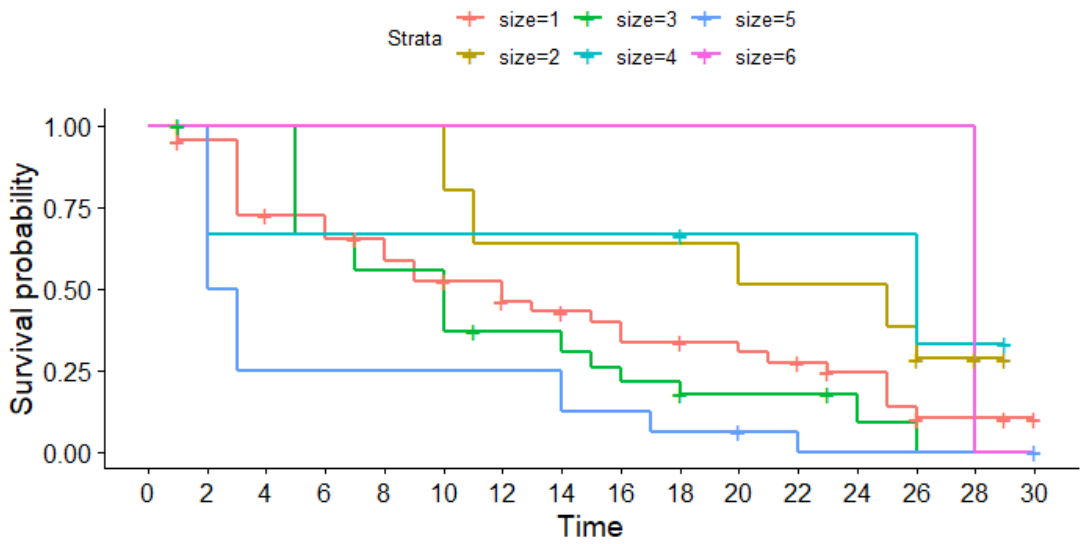

Fig. 2. Display of the survival probabilities of the four models for analyzing recurrent data

Hazard ratios estimates along with confidence intervals and *p-values* are plotted for each variable. The means are shown as squares and confidence interval estimates as lines. The righthand side shows the *p-values* for the corresponding regression coefficients which can be obtained from summary (AG_Model). The variable tx is located around 1 hence, its effect is marginal or minimal (See Fig. 3).

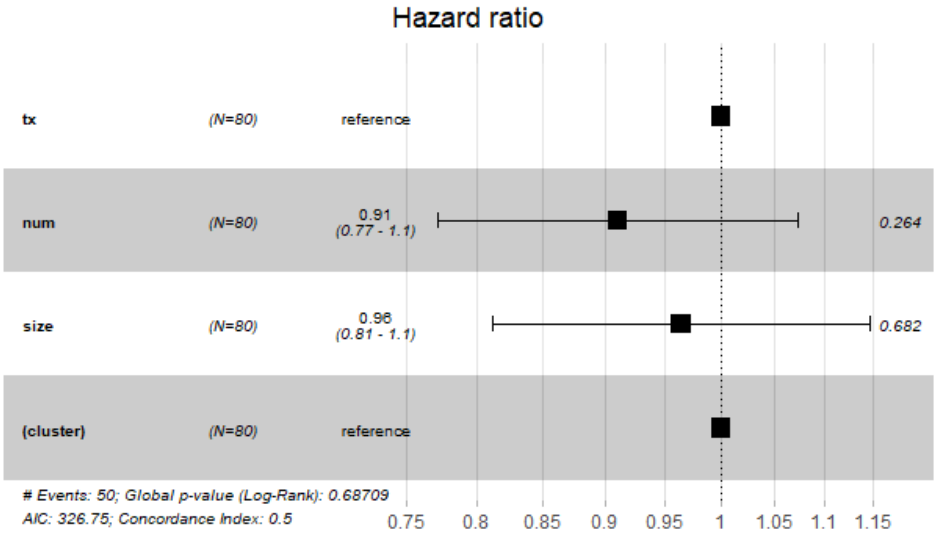

Fig. 3. Forest plot of hazard ratios for a multivariate CPHM. (AG Model); Summary of a multivariate Cox Proportional Hazard Model (AG Model) results shown using a forest plot

Similarly, hazard ratio estimates along with confidence intervals and *p-values* are plotted for each variable. The means are shown as squares and confidence interval estimates as lines. The righthand side shows the *p-values* for the corresponding regression coefficients which can be obtained from summary (Frailty Model). The variable tx is located around 1 hence, its effect is marginal (Figure 4).

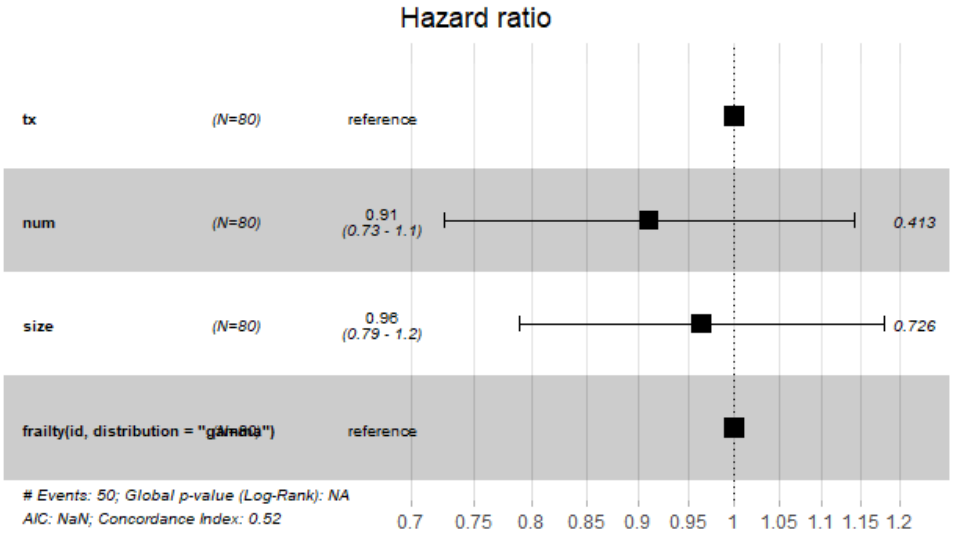

Fig. 4. Summary of a multivariate Cox Proportional Hazard Model (Frailty model, distribution = gamma) results shown using a forest plot

4 Conclusion

In conclusion, we join hands with earlier researchers and say that PWP-GT has proved to be the most useful model for analyzing recurrent event data, providing answers to slightly different research questions.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Nelson WB. Recurrent events data analysis for product repairs, disease recurrences, and other applications. Society for Industrial and Applied Mathematics; 2003.

[2] Li QH, Lagakos SW. Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. Statistics in medicine. 1997;16(8):925-940.

[3] Christer AH, Whitelaw J. An operational research approach to breakdown maintenance: Problem recognition. The Journal of the Operational Research Society. 1983;34(11):1041-1052.

[4] Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68:373-379.

[5] Prentice RL, Kalbfleisch JD. Aspects of the analysis of multivariative failure time data. SORT. 2003;27(1): January-June.

[6] Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. Ann Stat. 1982;10:1100-1120.

[7] Hosmer Jr, DW, Lemeshow S. Applied survival analysis: regression modelling of time to event data (1999). Eur Orthodontic Soc. 1999;561-2.

[8] Lin X, Carroll RJ. Semiparametric regression for clustered data using generalized estimating equations. Journal of the American Statistical Association. 2001;96(455):1045-1056.

[9] Velayutham B, Chadha VK, Singla N, Narang P, Gangadhar Rao V, Nair S, Natrajan M. Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: A multi-centric prospective study. PLoS One. 2018;13(7):e0200150.

[10] Duchateau L, Janssen P, Kezic I, Fortpied C. Evolution of recurrent asthma event rate over time in frailty models. J R Stat Soc Series C (Appl Stat). 2003;52:355–63.

[11] Kieser, AK, Rauch, G. A systematic comparison of recurrent event models for application to composite endpoints. BMC Med Res Methodol. 2018;18(2).
DOI:https://doi.org/10.1186/s12874-017-0462-x

[12] Jahn-Eimermacher A, Ingel K, Ozga AK, Preussler S, Binder H. Simulating recurrent event data with hazard functions defined on a total time scale. BMC Medical Research Methodology. 2015;15(1):1-9.

[13] Lu J, Shen D. Application of Survival Analysis in Multiple Events Using SAS, EP02_PharmaSUG2018_paper. 2018.

[14] Twisk JWR, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. Journal of Epidemiology & Community Health. 2005;59**:**706- 710.

DOI:https://doi.org/10.1093/ije/dyu222

[15] Ozga AK, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. BMC Med Res Methodol. 2018;18**:**2.
DOI:https://doi.org/10.1186/s12874-017-0462-x

[16] Cai J, Douglas E, Schaubel E. Analysis of recurrent event data, Handbook of Statistics, Elsevier. 2003;23: 603-623.
DOI:https://doi.org/10.1016/S0169- 7161(03)23034-0

[17] Kelley PJ, Lim LL-Y. Survival analysis for recurrent event data an application to childhood infectious diseases. Statistics in medicine. 2000;19:13-33.

[18] Amorim LDAF, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. International Journal of Epidemiology. 2015;44(1):324–333.

[19] Yadav CP, Sreenivas V, Khan MA, Pandey RM. An overview of statistical models for recurrent events analysis: A review. Epidemiology (Sunnyvale). 2018;8:354.
DOI: 10.4172/2327-4972.1000354.

[20] Kleinbaum DG, Klein M. Recurrent event survival analysis. In: Survival Analysis. Statistics for Biology and Health. Springer, New York, NY; 2012.
DOI:https://doi.org/10.1007/978-1-4419-6646-9_8

[21] Byer NE. Spontaneous regression and disappearance of subclinical rhegmatogenous retinal detachment. American Journal of Ophthalmology. 2001;131(2):269-270.

[22] Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc. 1989;84:1065-1073.

[23] Schaubel DE, Zeng D, Cai J. A semiparametric additive rates model for recurrent event data. Lifetime Data Analysis. 2006;12(4):389-406.

[24] Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons. 2011;360.

[25] Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. Springer Science & Business Media; 2012.

[26] Thenmozhi M, Jeyaseelan V, Jeyaseelan L, Isaac R, Vedantam R. Survival analysis in longitudinal studies for recurrent events: Applications and challenges. Clinical Epidemiology and Global Health. 2019;7(2):253-260.

[27] Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. British Journal of Sports Medicine. 2014;48(17):1287-1293.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/77313