Published in final edited form as:

Lifetime Data Anal. 2020 October ; 26(4): 820–832. doi:10.1007/s10985-020-09500-6.

Statistical analysis of clustered mixed recurrent-event data with application to a cancer survivor study

Liang Zhu,

Biostatistics and Epidemiology Research and Design, Division of Clinical and Translational Sciences, The University of Texas Health Science Center at Houston, TX, 77030, USA

Sangbum Choi,

Department of Statistics, Korea University, Seoul, 02841, South Korea

Yimei Li,

Department of Biostatistics, St. Jude Children's Research Hospital, TN, 38105, USA

Xuelin Huang,

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, TX, 77030, USA

Jianguo Sun,

Department of Statistics, University of Missouri, MO, 65211, USA

Leslie L. Robison

Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, TN 38105, USA

Abstract

In long-term follow-up studies on recurrent events, the observation patterns may not be consistent over time. During some observation periods, subjects may be monitored continuously so that each event occurrence time is known. While during the other observation periods, subjects may be monitored discretely so that only the number of events in each period is known. This results in mixed recurrent-event and panel-count data. In these data, there is dependence among within-subject events. Furthermore, if the data are collected from multiple centers, then there is another level of dependence among within-center subjects. Literature exists for clustered recurrent-event data, but not for clustered mixed recurrent-event and panel-count data. Ignoring the cluster effect may lead to less efficient analysis. In this paper, we present a marginal modeling approach to take into account the cluster effect and provide asymptotic distributions of the resulting regression parameters. Our simulation study demonstrates that this approach works well for practical situations. It was applied to a study comparing the hospitalization rates between childhood cancer survivors and healthy controls, with data collected from 26 medical institutions across North America during more than 20 years of follow-up.

Yimei Li, Tel.: +1-9015956735. Fax: +1-9015958843. yimei.li@stjude.org.

Conflict of interest

The authors have declared no conflict of interest.

Keywords

cluster effect; marginal model; recurrent events; regression analysis

1 Introduction

Recurrent event or event history studies occur in many fields. In recurrent event studies, the event of interest may occur more than once per subject. Examples include tumors, hospitalizations, infections, epilepsies, and acute myocardial infarctions. Two types of data concerning recurrent events have been extensively discussed in the literature. One is recurrent-event data which arise when study subjects are monitored continuously and the times of all occurrences of the recurrent events are recorded (Cook and Lawless, 2007). The other is panel-count data, which occur if the subjects are observed periodically and only the number of events between two subsequent observation times are available (Sun and Zhao, 2013). These data are often seen in large observational databases derived from long-term follow-up studies. Such databases could provide very useful and comprehensive information on event processes, however, they may introduce complicated data structures. For example, when the observation types or questionnaire designs are not consistent during the follow-up period, the observed outcomes could be a mixture of recurrent-event and panel-count data, or in short, mixed recurrent-event data, though the underlying recurrent-event process stays the same. Furthermore, in long-term familial, environmental, or multi-center studies, subjects could be correlated or clustered, resulting in clustered mixed recurrent-event data.

The analysis that motivated our current work considers the hospitalization experience among the participants in the Childhood Cancer Survivor Study (CCSS). It is the largest multicenter long-term follow-up cohort study of childhood cancer survivors in North America. The original cohort follows over 14,000 childhood cancer survivors who were diagnosed in 1970s and have survived at least 5 years since their diagnosis. The primary aim is to study the late effects of the childhood cancer and related treatment. Hospitalization is one of the most important late effects, and each survivor may have multiple hospital admissions during their life. Due to the inconsistent design of questionnaires, panel-count data were collected at the baseline and recurrent-event data were collected at each follow-up after baseline. Moreover, participants in the CCSS came from 26 institutions across North America. It is likely that patients treated at the same institution are correlated, due to similarities in practice patterns of clinicians from the same institution and patient socioeconomic and clinical characteristics. Naturally, the analysis must account for the correlation among hospitalizations within the same patient and the same institution. Ignoring such dependencies could result in biased variance estimation and hence lead to erroneous conclusions.

For unmixed recurrent-event data, Cook and Lawless (2007) gave a relatively thorough review of the literature on the analysis of recurrent event data. Other authors who have investigated recurrent-event data include Cai and Schaubel (2004), Lawless and Nadeau (1995), Lin et al. (2000), Pepe and Cai (1993), and Wang and Chen (2000). For unmixed recurrent-event data with clusters, Schaubel and Cai (2005a, 2005b) extended the

semiparametric proportional rates models for recurrent-event data proposed by Lin et al. (2000) to account for within-cluster correlation, with common or cluster-specific baseline rates. Liu et al. (2012, 2014) additionally discussed the estimation of cluster effects based on Schaubel and Cai (2005b) with proportional mean/rate model, while He et al. (2017) proposed to estimate the cluster effects with additive rates models.

For mixed recurrent-event data, Zhu et al. (2013, 2014, 2015, 2017) discussed regression analysis with proportional mean/rate model or additive rate model based on estimating equation or maximum likelihood methods, respectively. However, all these methods assume independence among individuals, and, therefore, cannot be readily applied to studies with clustered subjects. In this article, we propose a marginal method for fitting proportional mean models to mixed recurrent-event data wherein study subjects are clustered. The proposed model is semiparametric in that a functional form is assumed for multiplicative covariate effects, while the baseline rate and the dependence structure among correlated events and subjects are left unspecified, as they are not of direct interest.

The remainder of the article is organized as follows. In Section 2, we will discuss regression analysis of clustered mixed recurrent-event data. A marginal model will be presented and an estimating equation-based approach will be developed for estimation of regression parameters. The asymptotic distributions of the presented estimators are given. Section 3 will present results obtained from a simulation study conducted to assess the finite sample properties of the proposed estimators. In Section 4, we will apply the proposed methods to the CCSS data discussed above and Section 5 contains some concluding remarks.

2 Models and methods

Consider a recurrent event study involving n independent subjects. Let $N_i^*(t)$ be the cumulative number of events up to time t for the i th individual, $i = 1, \dots, n$. Let k denote the k th cluster with a cluster size of n_k for $k = 1, \dots, K$. Let G_i be the index of the cluster for the i th subject. Let C_i denote the right censoring time and let τ denote the largest observation time. Also, define $N_i(t) = N_i^*(t \wedge C_i)$ as the observed recurrent-event process of interest. For subject i , assume that there exist a p -dimensional vector of covariates denoted by Z_i and a sequence of time points $T_{i,0} = 0 < T_{i,1} < \dots < T_{i,J_i}$ such that the subject is observed either continuously or discretely over each interval $(T_{i,j-1}, T_{i,j}]$, with J_i being the total number of observation intervals for the i th subject. Define $O_i(t) = \sum_{j=1}^{J_i} I(T_{i,j} \leq \min(t, C_i))$ as the observation process. Corresponding to the time points above, define the indicator function $r_i(t) = 1$ for $t \in (T_{i,j-1}, T_{i,j}]$ if the i th subject is observed continuously within the interval and $r_i(t) = 0$ otherwise. Note that if $r_i(t) = 1$ for all i and t , the observed data reduce to recurrent-event data, and if $r_i(t) = 0$ for all i and t , the observed data reduce to panel-count data.

We then specify the model for subject i from cluster k as

$$E\{dN_i^*(t) \mid Z_i, G_i = k\} = d\mu_{0k}(t)e^{\beta_0^T Z_i}, \quad (1)$$

where $\mu_{0k}(t)$ is the baseline rate function for the k th cluster, β_0 is the vector of regression parameters of interest. Model (1) is often referred to as the proportional rates model and commonly used in regression analysis of recurrent-event data (Lin et al., 2000) and panel-count data (Cheng and Wei, 2004). Let $G_{ik} = \mathbb{I}(G_i = k)$, and define $dN_{ik}^*(t) = G_{ik}dN_i^*(t)$, then model (1) can be written as

$$E\{dN_{ik}^*(t) \mid Z_i\} = d\mu_{0k}(t)e^{\beta_0^T Z_i}. \quad (2)$$

In the following, we will assume that the main goal is to develop an estimating equation-based approach to estimate β_0 .

For the i th subject in the k th cluster, define $N_{ik}(t) = G_{ik}N_i(t)$, representing the cumulative number of events occurring up to time t , and the at risk indicator functions $Y_{ik}(t) = G_{ik}\mathbb{I}(t \leq C_i)$ for recurrent-event records and $\Delta_{ik}(t) = G_{ik}\mathbb{I}(t \leq T_i, J_i)$ for panel-count records. Also define the observation process $O_{ik}(t) = G_{ik}O_i(t)$, $\tilde{N}_{ik}(t) = \int_0^t \Delta_{ik}(s)N_{ik}^*(s)dO_{ik}(s)$, and $\Gamma_{0k}(t) = \int_0^t \mu_{0k}(s)dE\{O_{ik}(s)\}$ for panel-count records. Furthermore define

$$M_{ik}(t; \beta_0, \mu_{0k}, \Gamma_{0k}) = r_i(t)M_{ik}^r(t; \beta_0, \mu_{0k}) + (1 - r_i(t))M_{ik}^p(t; \beta_0, \Gamma_{0k}),$$

where

$$M_{ik}^r(t; \beta_0, \mu_{0k}) = N_{ik}(t) - \int_0^t Y_{ik}(s)e^{\beta_0^T Z_i} d\mu_{0k}(s),$$

and

$$M_{ik}^p(t; \beta_0, \Gamma_{0k}) = \tilde{N}_{ik}(t) - \int_0^t \Delta_{ik}(s)e^{\beta_0^T Z_i} d\Gamma_{0k}(s).$$

Then it can be easily verified that $EM_{ik}(t; \beta_0, \mu_{0k}, \Gamma_{0k}) = 0$. That is, $M_{ik}(t; \beta_0, \mu_{0k}, \Gamma_{0k})$ is a zero-mean process. This motivates the following estimating equations for k th cluster, $k = 1, \dots, K$,

$$\sum_{i=1}^n r_i(t)dM_{ik}^r(t; \beta_0, \mu_{0k}) = 0, \quad \sum_{i=1}^n (1 - r_i(t))dM_{ik}^p(t; \beta_0, \Gamma_{0k}) = 0, \quad (3)$$

and

$$\sum_{i=1}^n \int_0^\tau Z_i dM_{ik}(t; \beta_0, \mu_{0k}, \Gamma_{0k}) = 0. \quad (4)$$

Given β_0 , equation (3) suggests that $\mu_{0k}(t)$ and $\Gamma_{0k}(t)$ can be estimated by

$$\hat{\mu}_{0k}(t; \beta_0) = \int_0^t \frac{\sum_{i=1}^n r_i(s) dN_{ik}(s)}{S_{rk}^{(0)}(\beta_0, s)}, \quad \hat{\Gamma}_{0k}(t; \beta_0) = \int_0^t \frac{\sum_{i=1}^n (1 - r_i(s)) d\tilde{N}_{ik}(s)}{S_{pk}^{(0)}(\beta_0, s)}, \quad (5)$$

respectively, where for $l = 0, 1$ and $k = 1, \dots, K$, $S_{rk}^{(l)}(\beta, t) = \sum_{i=1}^n r_i(t) Y_{ik}(t) e^{\beta^T Z_i Z_i^{\otimes l}}$ and $S_{pk}^{(l)}(\beta, t) = \sum_{i=1}^n (1 - r_i(t)) \Delta_{ik}(t) e^{\beta^T Z_i Z_i^{\otimes l}}$ with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa'$ for a vector a . By plugging the above estimates into equation (4), we obtain the following estimating equation

$$U(\beta) = \sum_{i=1}^n \sum_{k=1}^K \left[\int_0^\tau r_i(t) Y_{ik}(t) \{Z_i - \bar{Z}_{rk}(\beta, t)\} dN_{ik}(t) + \int_0^\tau \{1 - r_i(t)\} \Delta_{ik}(t) \{Z_i - \bar{Z}_{pk}(\beta, t)\} d\tilde{N}_{ik}(t) \right] = 0, \quad (6)$$

where $\bar{Z}_{rk}(\beta, t) = S_{rk}^{(1)}(\beta, t) / S_{rk}^{(0)}(\beta, t)$ and $\bar{Z}_{pk}(\beta, t) = S_{pk}^{(1)}(\beta, t) / S_{pk}^{(0)}(\beta, t)$.

Let $\hat{\beta}$ denote the estimator of β_0 given by the solution to equation (6) and define $\hat{M}_{ik}^r(t) = M_{ik}^r(t; \hat{\beta}, \hat{\mu}_{0k})$ and $\hat{M}_{ik}^p(t) = M_{ik}^p(t; \hat{\beta}, \hat{\Gamma}_{0k})$ for $i = 1, \dots, n$ and $k = 1, \dots, K$. Then it can be shown that $\hat{\beta}$ is consistent and one can asymptotically approximate the distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ by the multivariate normal distribution with mean zero and the covariance matrix $\hat{\Omega}^{-1} \hat{\Sigma} \hat{\Omega}^{-1}$, where

$$\hat{\Omega}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \left[\int_0^\tau r_i(t) \{Z_i - \bar{Z}_{rk}(\hat{\beta}, t)\}^{\otimes 2} dN_{ik}(t) + \{1 - r_i(t)\} \{Z_i - \bar{Z}_{pk}(\hat{\beta}, t)\}^{\otimes 2} d\tilde{N}_{ik}(t) \right],$$

and

$$\hat{\Sigma}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \left[\sum_{k=1}^K \int_0^\tau r_i(t) \{Z_i - \bar{Z}_{rk}(\hat{\beta}, t)\} d\hat{M}_{ik}^r(t) + \{1 - r_i(t)\} \{Z_i - \bar{Z}_{pk}(\hat{\beta}, t)\} d\hat{M}_{ik}^p(t) \right]^{\otimes 2}.$$

The proof is sketched in the Supplementary Material. Furthermore, one can estimate $\mu_{0k}(t)$ and $\Gamma_{0k}(t)$ by $\hat{\mu}_{0k}(t; \hat{\beta})$ and $\hat{\Gamma}_{0k}(t; \hat{\beta})$, respectively.

3 A simulation study

An extensive simulation study was conducted to assess the finite sample behaviors of the proposed estimates for mixed recurrent-event data with clusters. In the study, we considered a 3-dimensional covariate Z with Z_1 following the Bernoulli distribution with the probability of success 0.5, Z_2 following the standard normal distribution, and Z_3 following the uniform distribution $U(0, 1)$. The independent censoring time was generated from the uniform distribution $U(\tau/2, \tau)$ with $\tau = 1$. The underlying recurrent-event process $N_i^*(t)$ was simulated from the following Poisson model

$$E\{dN_i^*(t) \mid G_i = k, Z_i\} = \theta_k d\mu_0(t) e^{\beta_0^T Z_i}$$

with $\mu_0(t) = 3t$. For the cluster effect θ_k , $k = 1, \dots, K$, we considered two scenarios: 1) θ_k alternated among 0.1, 1, and 1.9, 2) θ_k followed a gamma distribution with mean 1 and variance 2. By multiplying θ_k with $\mu_0(t)$, we obtained cluster-specific baseline function. We considered $K = 10$ centers, with the number of subjects n_k ranging from 50 to 250.

We also considered the case when the Poisson assumption did not hold by generating the underlying recurrent-event process $N_i^*(t)$ from a mixed Poisson process

$$E\{dN_i^*(t) \mid G_i = k, Z_i, W_i\} = W_i \theta_k d\mu_0(t) e^{\beta_0^T Z_i},$$

where W_i followed the gamma distribution with mean 1 and variance 0.5. To generate mixed recurrent-event data, for each subject, we first generated a sequence of observation time points $t_{i,j}$'s from the Poisson process with the mean function $\mu_0(t) = 3t$. Then within each observation period $(t_{i,j-1}, t_{i,j}]$, the indicator of data type $r_i(t)$ was generated from the Bernoulli distribution with the probability of success p_r as before, $j = 1, \dots, J_i$, $i = 1, \dots, n$. If $r_i(t) = 1$, then the exact times of the events that occur during the period $(t_{i,j-1}, t_{i,j}]$, generated from above recurrent-event processes, were assumed to be known and recorded. Otherwise, the number of the events that occur during the period $(t_{i,j-1}, t_{i,j}]$ was counted and recorded. The results given below are based on 1,000 replications.

Tables 1 to 4 present the results on estimation of β_0 based on the simulated mixed recurrent-event data from either the Poisson process or the mixed-effects Poisson process under the scenario 1 or 2 with the true value of β_0 being (0,0,0) or (-0.5,0,0.5) and $p_r = 0.2, 0.5$ or 0.8. The results include the average of the point estimates (Estimate), the sample standard error of the point estimates (SSE), the average of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). Note that here the 95% CP is calculated as the proportion of times in which the obtained confidence interval (point estimate ± 1.96 estimated standard error) contains the true value. These results indicate that the proposed estimator seems to be unbiased and the estimated standard errors are close to the sample standard errors. The normal approximation to the distribution of the proposed estimate seems to perform well.

Also it can be seen from all tables that as the probability p_r of observing recurrent-event data increased, as expected, the estimators become more accurate as more information is available. In each table, for comparison, we also provide the results from the method of Zhu et al. (2013), which does not consider the cluster effect. The results show that considering the cluster effect can give smaller standard errors, and hence the estimate is more efficient.

4 An Application

Now we apply the methodology proposed in the Section 2 to a set of mixed recurrent-event data with clusters arising from the CCSS. The group we will study includes non-Hodgkin

lymphoma (NHL) survivors who were diagnosed and treated for their cancer during their childhood. Because of major advances in treatment, most children treated for lymphoma live into adulthood, so their quality of life as they get older has come into focus in recent years. Here we will study the recurrence of hospital visits in these cancer survivors. We hypothesize that NHL survivors have a higher hospitalization rate compared to healthy controls. Among others, one usage of such information is to help to potentially impact specific healthcare planning for these patients in terms of screening and management of chronic health conditions that lead to hospitalization.

Cancer survivors in the CCSS came from 26 institutions across North America. Healthy controls are recruited from the same geographic areas of the 26 institutions. These participants were distributed a baseline questionnaire in 1996 and some follow-up questionnaires periodically thereafter. The baseline questionnaire asked the participants “how many times were you admitted to any hospital”, while follow-up questionnaires not only asked the numbers of hospitalizations, but also the details (reason, procedure, hospital name and date of each hospitalization). As a result, we have a mixed recurrent-event and panel-count data with 26 clusters.

There are 630 NHL survivors and 1,860 healthy controls who have entered adulthood at the baseline follow-up. We use Z to denote the group with 1 for NHL survivors and 0 for controls. The mean number of hospitalization for the NHL survivor group and control group are 1.111 and 0.572, respectively. The average cluster size across 26 institutions is 96. The mean number of hospitalizations in each institution ranges from 0 to 3.83 for NHL survivors and 0.14 to 0.82 for controls.

We applied both the proposed method and the method of Zhu et al. (2013) which did not consider the cluster effect to this data. The estimated β from the proposed method is 0.560 with a standard error of 0.130, and the estimate from the method of Zhu et al. (2013) is 0.576 with a standard error of 0.144. The standard error based on the proposed method is smaller than the method of Zhu et al. (2013) and both results suggest that NHL survivors had a higher hospitalization rate compared to those healthy controls.

5 Discussion

This article considered regression analysis of clustered mixed recurrent-event data. Such data arise when study subjects are observed continuously over some periods while monitored discretely over other periods. As discussed above, the mixed recurrent-event data can occur in many fields. While some work exists for analyzing mixed recurrent-event data, little work has been done for such data with clusters. For the problem, we proposed a proportional rate model with cluster-specific baseline rate function for recurrent events, and the performed numerical study suggested that the proposed methods performed well in practical situations. It also indicated that the use of the cluster-specific rate function could lead to more efficient estimators compared to the common rate function when the cluster effect exists. The same was reflected in the real data analysis. The simulation code is given as supplementary material.

As pointed out by a reviewer, one basic assumption behind the proposed method is the proportional assumption on the covariate effect in model (1) or (2). Although it is the most commonly used model for regression analysis of event history data, it is easy to see that it may not hold sometimes and thus it would be important and useful to develop some test procedures for the proportionality. On the other hand, this may not be easy for the situation discussed above as it does not seem to exist such methods even for mixed recurrent-event data without clusters. To address this, an alternative is to generalize the proposed method to the case with time-varying covariate effects or with both time-dependent covariates and time-varying covariate effects. Note that several authors have discussed these issues for regression analysis of either recurrent-event data (Sun et al., 2009; Li et al., 2016; Yu et al., 2013) or panel-count data (Fang et al., 2017; Wang and Yu, 2019). In particular, Sun et al. (2009) considered multivariate recurrent-event data and developed some estimating equation-based inference procedures, which are similar to the methods given above. Furthermore, they proposed a Kolmogorov-Smirnov type statistic and a Cramér-von Mises type statistic for testing whether the covariate coefficients are time-varying. It would be useful to generalize these methods to the clustered mixed recurrent-event data discussed here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The work was partly supported by NIH Grant (R21CA198641; R03CA219450) to Zhu.

References

1. Cai J, Schaubel DE. Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* 2004; 10, 121–138. [PubMed: 15293628]
2. Cheng SC, Wei LJ. Inferences for a semiparametric model with panel data. *Biometrika*. 2000 3 1;87(1):89–97.
3. Cook RJ, Lawless JF. *The analysis of recurrent event data* 2007, Springer-Verlag:New York.
4. Fang S, Zhang HX, Sun LQ, Wang DH. Analysis of panel count data with time-dependent covariates and informative observation process. *Acta Mathematicae Applicatae Sinica, English Series*. 2017 Feb 1;33(1):147–56.
5. He H, Pan D, Sun L, Li Y, Robison LL, Song X. Analysis of a fixed center effect additive rates model for recurrent event data. *Computational Statistics and Data Analysis* 2017; 112: 186197.
6. Lawless JF and Nadeau C Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; 37, 158–168.
7. Li S, Sun Y, Huang CY, Follmann DA, Krause R. Recurrent event data analysis with intermittently observed timevarying covariates. *Statistics in medicine*. 2016 Aug 15;35(18):3049–65. [PubMed: 26887664]
8. Lin DY, Wei LJ, Yang I, Ying Z Semiparametric regression for the mean and rate function of recurrent events. *Journal of Royal Statistical Society, Series B* 2000; 69, 711–730.
9. Liu D, Kalbfleisch JD, Schaubel DE. Methods for estimating center effects on recurrent events. *Statistics in Biosciences* 2012; 6(1): 19–37.
10. Liu D, Kalbfleisch JD, Schaubel DE. Methods for estimating center effects on recurrent events. *Statistics in Biosciences* 2014; 6(1): 19–37. [PubMed: 24976871]

11. Pepe MS, Cai J Some graphic displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* 1993; 88, 811–820.
12. Pollard D Empirical processes: theory and applications InNSF-CBMS regional conference series in probability and statistics 1990 Jan 1 (pp. i–86). Institute of Mathematical Statistics and the American Statistical Association.
13. Schaubel DE, Cai J. Analysis of clustered recurrent-event data with application to hospitalization rates among renal failure patients. *Biostatistics* 2005a; 6: 404–419. [PubMed: 15831581]
14. Schaubel DE, Cai J. Semiparametric methods for clustered recurrent event data. *Lifetime Data Analysis* 2005b; 11 (3): 405–425. [PubMed: 16133887]
15. Sun J, Zhao X. *The Statistical Analysis of Panel Count Data*. 2013; New York: Springer.
16. Sun L, Zhu L, Sun J. Regression analysis of multivariate recurrent event data with timevarying covariate effects. *Journal of multivariate analysis*. 2009 11 1;100(10):2214–23.
17. Wang MC, Chen YQ. Nonparametric and semiparametric trend analysis of stratified recurrent time data. *Biometrics* 2000; 56, 789–794. [PubMed: 10985217]
18. Wang Y, Yu Z. A kernel regression model for panel count data with time-varying coefficients arXiv preprint arXiv:1903.10233. 2019 3 25.
19. Yu Z, Liu L, Bravata DM, Williams LS, Tepper RS. A semiparametric recurrent eventsmodel with timevarying coefficients. *Statistics in medicine*. 2013 Mar 15;32(6):1016–26. [PubMed: 22903343]
20. Zhu L, Zhao H, Tong X, Sun J, Srivastava DK, Leisenring W, Robison LL. Statistical analysis of mixed recurrent event data with application to cancer survivor study. *Statistics in Medicine* 2013; 32(11): 1954–63. [PubMed: 23139023]
21. Zhu L, Tong X, Sun J, Chen M, Srivastava DK, Leisenring W, Robison LL. Regression analysis of mixed recurrent-event and panel-count data. *Biostatistics* 2014; 15(3):555–568. [PubMed: 24648408]
22. Zhu L, Zhao H, Sun J, Leisenring W, Robison LL. Regression analysis of mixed recurrent-event and panel-count data with additive rate models. *Biometrics* 2015; 71(1): 71–79. [PubMed: 25345405]
23. Zhu L, Zhang Y, Li Y, Sun J, Robison LL. A semiparametric likelihood-based method for regression analysis of mixed panel-count data. *Biometrics* 2017; 74, 488–497. [PubMed: 28914961]

Table 1

Summary statistics for the simulation studies: Poisson with scenario 1

p_r	β	Proposed				Zhu et al. (2013)			
		BIAS	SSE	ESE	CP	BIAS	SSE	ESE	CP
0.2	$\beta_1=0$	-0.001	0.075	0.070	0.929	0.000	0.086	0.083	0.940
	$\beta_2=0$	0.000	0.037	0.035	0.928	0.000	0.042	0.041	0.946
	$\beta_3=0$	-0.005	0.121	0.122	0.951	-0.006	0.139	0.143	0.952
0.5	$\beta_1=0$	-0.002	0.069	0.067	0.943	0.001	0.076	0.080	0.954
	$\beta_2=0$	-0.001	0.034	0.033	0.943	0.000	0.040	0.040	0.949
	$\beta_3=0$	0.004	0.118	0.115	0.942	0.002	0.138	0.138	0.955
0.8	$\beta_1=0$	-0.004	0.061	0.057	0.935	-0.002	0.075	0.073	0.945
	$\beta_2=0$	0.001	0.029	0.028	0.937	0.002	0.037	0.037	0.948
	$\beta_3=0$	-0.002	0.103	0.098	0.938	-0.002	0.131	0.127	0.939
0.2	$\beta_1=-0.5$	0.000	0.071	0.071	0.954	0.001	0.084	0.083	0.955
	$\beta_2=0$	0.001	0.036	0.035	0.931	0.002	0.042	0.042	0.950
	$\beta_3=0.5$	-0.002	0.125	0.122	0.941	-0.001	0.145	0.145	0.946
0.5	$\beta_1=-0.5$	-0.003	0.068	0.067	0.942	-0.002	0.080	0.081	0.944
	$\beta_2=0$	0.001	0.035	0.033	0.939	0.001	0.042	0.040	0.943
	$\beta_3=0.5$	0.002	0.121	0.116	0.937	0.002	0.140	0.141	0.958
0.8	$\beta_1=-0.5$	-0.001	0.064	0.059	0.937	-0.002	0.078	0.076	0.941
	$\beta_2=0$	-0.003	0.030	0.029	0.949	-0.002	0.040	0.038	0.930
	$\beta_3=0.5$	-0.008	0.112	0.101	0.921	-0.005	0.137	0.131	0.945

Table 2

Summary statistics for the simulation studies: Poisson with scenario 2

p_r	β	Proposed				Zhu et al. (2013)			
		BIAS	SSE	ESE	CP	BIAS	SSE	ESE	CP
0.2	$\beta_1=0$	0.001	0.086	0.079	0.928	0.003	0.103	0.101	0.945
	$\beta_2=0$	0.001	0.041	0.039	0.942	0.002	0.051	0.050	0.950
	$\beta_3=0$	0.006	0.143	0.137	0.946	0.004	0.176	0.174	0.950
0.5	$\beta_1=0$	0.000	0.078	0.074	0.937	-0.001	0.099	0.097	0.939
	$\beta_2=0$	-0.001	0.040	0.037	0.929	0.000	0.050	0.048	0.947
	$\beta_3=0$	0.006	0.135	0.128	0.930	0.001	0.173	0.168	0.951
0.8	$\beta_1=0$	0.003	0.067	0.062	0.935	0.005	0.100	0.091	0.935
	$\beta_2=0$	0.001	0.034	0.031	0.938	0.003	0.046	0.045	0.938
	$\beta_3=0$	-0.005	0.117	0.107	0.922	0.002	0.170	0.157	0.937
0.2	$\beta_1=-0.5$	-0.008	0.085	0.080	0.936	-0.006	0.106	0.102	0.948
	$\beta_2=0$	0.001	0.043	0.040	0.930	0.001	0.054	0.052	0.940
	$\beta_3=0.5$	0.009	0.147	0.139	0.933	0.001	0.186	0.179	0.943
0.5	$\beta_1=-0.5$	0.002	0.081	0.075	0.927	-0.004	0.100	0.099	0.950
	$\beta_2=0$	0.000	0.040	0.037	0.936	0.000	0.050	0.050	0.940
	$\beta_3=0.5$	-0.002	0.141	0.130	0.921	0.000	0.175	0.173	0.948
0.8	$\beta_1=-0.5$	-0.004	0.067	0.063	0.926	-0.004	0.093	0.092	0.947
	$\beta_2=0$	0.002	0.033	0.031	0.934	0.001	0.046	0.046	0.952
	$\beta_3=0.5$	0.001	0.113	0.107	0.938	-0.004	0.160	0.160	0.962

Table 3

Summary statistics for the simulation studies: mixed Poisson with scenario 1

p_r	β	Proposed				Zhu et al. (2013)			
		BIAS	SSE	ESE	CP	BIAS	SSE	ESE	CP
0.2	$\beta_1=0$	-0.002	0.092	0.090	0.944	-0.005	0.103	0.101	0.938
	$\beta_2=0$	-0.002	0.049	0.045	0.930	-0.003	0.052	0.050	0.935
	$\beta_3=0$	-0.004	0.159	0.156	0.939	0.000	0.177	0.174	0.945
0.5	$\beta_1=0$	0.000	0.092	0.086	0.941	-0.002	0.099	0.098	0.943
	$\beta_2=0$	-0.001	0.045	0.043	0.936	-0.002	0.048	0.049	0.956
	$\beta_3=0$	-0.012	0.152	0.149	0.946	-0.009	0.171	0.169	0.948
0.8	$\beta_1=0$	-0.005	0.080	0.077	0.940	-0.001	0.092	0.091	0.952
	$\beta_2=0$	0.002	0.039	0.038	0.951	0.001	0.045	0.045	0.958
	$\beta_3=0$	-0.002	0.142	0.133	0.940	0.003	0.160	0.157	0.947
0.2	$\beta_1=-0.5$	0.003	0.093	0.091	0.949	0.002	0.103	0.101	0.953
	$\beta_2=0$	0.002	0.048	0.046	0.932	0.003	0.052	0.051	0.937
	$\beta_3=0.5$	-0.009	0.166	0.158	0.939	-0.012	0.182	0.177	0.948
0.5	$\beta_1=-0.5$	-0.002	0.093	0.087	0.941	0.001	0.104	0.099	0.939
	$\beta_2=0$	0.002	0.046	0.044	0.943	0.000	0.051	0.050	0.944
	$\beta_3=0.5$	0.005	0.157	0.152	0.945	-0.002	0.172	0.173	0.944
0.8	$\beta_1=-0.5$	0.001	0.081	0.078	0.936	-0.001	0.091	0.092	0.955
	$\beta_2=0$	0.002	0.042	0.039	0.931	0.003	0.047	0.046	0.940
	$\beta_3=0.5$	0.003	0.146	0.135	0.934	0.002	0.171	0.161	0.932

Table 4

Summary statistics for the simulation studies: mixed Poisson with scenario 2

p_r	β	Proposed				Zhu et al. (2013)			
		BIAS	SSE	ESE	CP	BIAS	SSE	ESE	CP
0.2	$\beta_1=0$	-0.001	0.111	0.104	0.940	0.001	0.128	0.122	0.927
	$\beta_2=0$	-0.001	0.055	0.052	0.937	0.000	0.064	0.061	0.942
	$\beta_3=0$	-0.005	0.194	0.180	0.951	-0.007	0.214	0.210	0.948
0.5	$\beta_1=0$	0.000	0.105	0.098	0.941	-0.003	0.121	0.117	0.951
	$\beta_2=0$	0.002	0.053	0.049	0.934	0.002	0.063	0.058	0.943
	$\beta_3=0$	0.000	0.179	0.170	0.948	-0.006	0.206	0.203	0.949
0.8	$\beta_1=0$	0.000	0.092	0.088	0.945	0.000	0.112	0.111	0.949
	$\beta_2=0$	0.001	0.048	0.044	0.929	0.001	0.057	0.055	0.949
	$\beta_3=0$	-0.009	0.164	0.151	0.934	-0.009	0.204	0.191	0.939
0.2	$\beta_1=-0.5$	-0.005	0.107	0.105	0.946	0.000	0.121	0.123	0.952
	$\beta_2=0$	0.000	0.054	0.053	0.943	0.000	0.065	0.062	0.948
	$\beta_3=0.5$	0.005	0.190	0.183	0.946	-0.004	0.227	0.216	0.941
0.5	$\beta_1=-0.5$	-0.006	0.108	0.100	0.934	-0.004	0.129	0.120	0.936
	$\beta_2=0$	0.000	0.052	0.050	0.936	0.001	0.060	0.060	0.952
	$\beta_3=0.5$	0.006	0.190	0.173	0.939	0.012	0.220	0.210	0.930
0.8	$\beta_1=-0.5$	-0.001	0.096	0.089	0.938	-0.001	0.119	0.112	0.932
	$\beta_2=0$	-0.002	0.048	0.044	0.934	-0.001	0.057	0.056	0.947
	$\beta_3=0.5$	-0.002	0.162	0.154	0.932	-0.008	0.195	0.196	0.958