Communications for Statistical Applications and Methods
2014, Vol. 21, No. 5, 387–393

DOI: http://dx.doi.org/10.5351/CSAM.2014.21.5.387
Print ISSN 2287-7843 / Online ISSN 2383-4757

Analysis of Recurrent Gap Time Data with a Binary Time-Varying Covariate

Yang-Jin Kim[1,a]

[a]Department of Statistics, Sookmyung Women's University, Korea

Abstract

Recurrent gap times are analyzed with diverse methods under several assumptions such as a marginal model or a frailty model. Several resampling techniques have been recently suggested to estimate the covariate effect; however, these approaches can be applied with a time-fixed covariate. According to simulation results, these methods result in biased estimates for a time-varying covariate which is often observed in a longitudinal study. In this paper, we extend a resampling method by incorporating new weights and sampling scheme. Simulation studies are performed to compare the suggested method with previous resampling methods. The proposed method is applied to estimate the effect of an educational program on traffic conviction data where a program participation occurs in the middle of the study.

Keywords: Gap times, recurrent event data, resampling method, time-varying covariate, WCR, YTOP.

1. Introduction

In a longitudinal study that involves n independent subjects, each subject experiences same type of event repeatedly (Cook and Lawless, 2007). There are two approaches to analyze recurrent event data: a total time scale and a gap time scale that are incorporated depending on the intrinsic attribute of an event process. Recurrent gap times have been widely applied under a renewal process. Several estimating procedures have been suggested with an adjustment of the correlation among gap times. Wang and Chang (1999) and Lin and Ying (2001) suggested nonparametric estimators and for regression problem, Cai and Schaubel (2004) and Huang and Chen (2003) applied a proportional hazard model and Sun *et al.* (2006) assumed an additive model, respectively. Furthermore, the association among gap times was also estimated through a frailty effect. Recently, as an alternative approach, Luo and Huang (2011) suggested a modified within-cluster resampling(MWCR) as the extension of a within cluster resampling(WCR) method which is known as very effective for a cluster data with an informative cluster size (Williamson *et al.*, 2008). That is, similar with a clustered data, in a recurrent event, a greater number of observed gap times is associated with a longer risk of an event. Darlington and Dixon (2013) extended their idea and considered a weighted partial likelihood to overcome the computational burden of MWCR.

In this article, our interest is to estimate the effect of a time- varying covariate which often occurs at a longitudinal study. When a recurrent event data are analyzed with a total time scale, the effect of a time-varying covariate is estimated very naturally with a counting process. Huang *et al.*(2010) used a pairwise pseudo-likelihood which is extended to multivariate recurrent data by Zhao *et al.*

This Research was supported by the Sookmyung Women's University Research Grants 2013.

[1] Department of Statistics, Sookmyung Women's University, Chungpa-Dong, Yongsan-Gu, Seoul 140-742, Korea.
E-mail: yjin@sookmyung.ac.kr

Published 30 September 2014 / journal homepage: http://csam.or.kr

(2012). However, for a time-varying covariate, more caution is needed with a resampling technique at recurrent gap time. As Darlington and Dixon (2013) remarked, these two methods cannot be applied to a time-varying covariate. In this paper, we adjust a resampling technique to estimate the effect of a binary time-varying covariate on recurrent gap times.

In Section 2, we present a brief review of two resampling methods and describe a suggested method. In Section 3, a Monte Carlo simulation study is performed to compare the methods and to validate the performance of a proposed method. A real data from a traffic conviction is applied as an example in Section 4. We conclude with a general discussion in Section 5.

2. Covariate Weighted Analysis

Let t_{ij} be the gap time between the $(j-1)^{\text{th}}$ and the j^{th} events $(j = 1, \ldots, n_i)$ of an individual $i(= 1, \ldots, n)$. For individuals with $n_i > 1$, a gap time censoring indicator is defined as $\delta_{ij} = 1$ for $j = 1, \ldots, n_{i-1}$ and $\delta_{i_{n_i}} = 0$. Also, denote s_{ij-1} and s_{ij} as the $(j-1)^{\text{th}}$ and the j^{th} observed event times, respectively and then define $t_{ij} = s_{ij} - s_{ij-1}$. A censoring time, τ_i and gap times, $t'_{ij}s$ are assumed to be independent. They also have a following relation that $s_{in_i} = \sum_{j=1}^{n_i} t_{ij} > \tau_i$ and $s_{in_{i-1}} = \sum_{j=1}^{n_{i-1}} t_{ij} < \tau_i$. Let $X_i(\cdot)$ be a binary time-varying covariate. Let Z_i^* be the occurrence time of an event changing the status of a time-varying covariate. That is, a time-varying covariate is defined as $X_i(s) = I(Z_i^* \leq s)$. Here the recurrent gap times are assumed to be independent with Z_i^*. With a definition $x_{ij} = X_i(s)$ where $s = \sum_{l=1}^{j} t_{il}$, a gap time $t_{ij} = t$ is assumed to follow a proportional hazard model,

$$\lambda_i(t|x_{ij}) = \lambda_0(t)\exp\left(\beta' x_{ij}\right), \tag{2.1}$$

where β is a vector of the regression parameters and $\lambda_0(s)$ is a baseline hazard function with a cumulative hazard function, $\Lambda_0(s) = \int_0^s \lambda_0(u)du$. In this paper, our interest is to adjust a resampling technique for analyzing a recurrent gap time data. Before presenting the suggested method, the methods proposed by Luo and Huang (2011) and Darlington and Dixon (2013) are briefly summarized.

2.1. Weighted resampling method

(i) Risk weighted resampling: MWCR

By extending the WCR method to recurrent gap times, a longer risk effect on recurrent gap times is adjusted. Define $x_{ij} = x_i(s)$, where $s = \sum_{l=1}^{j} t_{il}$. Denote $\mathbf{y}_{b,i}^* = \{(t_{b,i}^*, \delta_{b,i}^*, x_{b,i}^*), i = 1, \ldots, n\}$ as the $b(= 1, \ldots, B)^{\text{th}}$ resampling data selected randomly from $\mathbf{y}_{ij} = \{(t_{ij}, \delta_{ij}, x_{ij}), i = 1, \ldots, n, j = 1, \ldots, n_{i-1}\}$ and define $\tilde{\beta}_b$ as the estimates obtained from the bth resampled data. Then the MWCR estimator is derived with B's estimates as follows,

$$\tilde{\beta} = \frac{1}{B}\sum_{b=1}^{B}\tilde{\beta}_b \tag{2.2}$$

and the corresponding variance is estimated

$$\mathrm{Var}\left(\tilde{\beta}\right) = \frac{1}{B}\sum_{b=1}^{B}\tilde{\Sigma}_b - \frac{1}{B-1}\sum_{b=1}^{B}\left(\tilde{\beta}_b - \tilde{\beta}\right)^{\otimes 2}, \tag{2.3}$$

where $\tilde{\Sigma}_b$ is an estimator of the variance-covariance matrix of the maximum partial likelihood estimator, $\tilde{\beta}_b$.

(ii) Event-weighted PH: EW

While a MWCR is understood as a method for a weighted risk set, Darlington and Dixon (2013) considered a same problem by directly assigning weights to the events. By extending the method of Williamson *et al.* (2008), such intension is applied to the partial likelihood,

$$L_p(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left[\frac{\exp\left(\beta' x_{ij}\right)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} Y_{kl}(t_{ij}) \exp\left(\beta' x_{ij}\right)} \right]^{w_i \delta_{ij}},$$

where $w_i = 1/[\sum_{j=} \delta_{ij}]$ and a following log partial likelihood function is derived,

$$U_{\mathrm{EW}}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} w_i \delta_{ij} \left[x_i - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_i} x_k Y_{kl}(t_{ij}) \exp(x_k \beta)}{\sum_{k=1}^{n} \sum_{l=1}^{n_i} Y_{kl}(t_{ij}) \exp(x_k \beta)} \right], \tag{2.4}$$

where $Y_{kl}(t) = I(t_{kl} > t)$ is an inverse of the number of observed gap times. Darlington and Dixon (2013) showed two methods have similar result by simulation with a time-fixed covariate.

2.2. Covariate weighted within cluster resampling: CwWCR

With a longitudinal data, covariates can be changed over time. In this paper, we consider a covariate changing a value in a middle of study. This results in a binary time-varying covariate. At next section, two methods explained in Section 2.1 do not provide desirable results with a time-varying covariate. A covariate weighted WCR(CwWCR) method modified an EW method to estimate the effect of a binary time-varying covariate on recurrent gap times. We assume that a covariate can change a value in the middle of a study. The $b(= 1, \ldots, B)^{\mathrm{th}}$ resampling procedure is as follows,

(step 1) δ_i^x is defined as either 1 or 0, Whether a covariate value is changed. In detail, if the ith subject has a time-varying covariate, then $\delta_i^x = 1$ and $= 0$ otherwise.

(step 2) For a subject with $\delta_i^x = 0$, randomly select one from $\mathbf{y}_i = (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{im_i})$, $m_i = n_i - 1$ and define a weight, $w_{i1} = 1$.

(step 3) For a subject with $\delta_i^x = 1$, divide $\mathbf{y}_i$ into two sets according to covariate values. That is, the first set is $g_{i1} = (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{ik_i})$ and the second set $g_{i2} = (\mathbf{y}_{ik_i+1}, \ldots, \mathbf{y}_{in_i-1})$, respectively, where

$$k_i = \arg\max \left\{ j : \sum_{l=1}^{j} t_{il} < Z_i^* \right\}.$$

Then randomly select one sample from the each set and assign the following weights,

$$w_{i1} = \frac{k_i}{m_i} \quad \text{and} \quad w_{i2} = 1 - \frac{k_i}{m_i},$$

respectively.

(step 4) With data sampled at (step 2) and (step 3), estimate β by using the following estimating function,

$$U_{\mathrm{CW}}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{\delta_i^x + 1} w_{ij} \delta_{ij} \left[x_{ij} - \frac{\sum_{k=1}^{n} \sum_{l=1}^{\delta_k^x + 1} x_{kl} Y_{kl}(t_{ij}) \exp(x'_{kl} \beta)}{\sum_{k=1}^{n} \sum_{l=1}^{\delta_k^x + 1} Y_{kl}(t_{ij}) \exp(x'_{kl} \beta)} \right]$$

Table 1: Comparison of simulation results with $\beta = 0.5$ and $\rho = 1$

p_x		$n = 100$				$n = 200$			
		$\hat{\beta}$	SSE	SEE	CP	$\hat{\beta}$	SSE	SEE	CP
0.4	MWCR	0.712	0.110	0.119	0.556	0.697	0.071	0.082	0.364
	EW	0.245	0.274	0.082	1.000	0.245	0.192	0.059	0.979
	CwWCR	0.480	0.189	0.198	0.984	0.487	0.127	0.140	0.986
0.6	MWCR	0.723	0.096	0.101	0.393	0.709	0.069	0.070	0.169
	EW	0.263	0.238	0.071	1.000	0.264	0.168	0.052	0.958
	CwWCR	0.488	0.170	0.198	0.970	0.486	0.119	0.139	0.975
0.8	MWCR	0.743	0.104	0.096	0.300	0.722	0.067	0.067	0.096
	EW	0.286	0.222	0.073	0.996	0.286	0.154	0.050	0.974
	CwWCR	0.494	0.161	0.217	0.990	0.485	0.112	0.153	0.997

and denote $\hat{\beta}^b_{cw}$ as the estimate from the b^{th} resampling.

(step 5) Repeat (step 1)–(step 4) B times and then obtain $(\hat{\beta}^1_{\text{cw}}, \ldots, \hat{\beta}^B_{\text{cw}})$ and $(\hat{\Sigma}^1_{\text{cw}}, \ldots, \hat{\Sigma}^B_{\text{cw}})$ and a final estimate is calculated as

$$\hat{\beta}_{\text{cw}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^b_{\text{cw}}.$$

Then corresponding variance is estimated with,

$$\hat{\Sigma}_{\text{cw}} = \text{Var}\left(\hat{\beta}_{\text{cw}}\right) = \frac{1}{B}\sum_{b=1}^{B}\hat{\Sigma}^b_{\text{cw}} - \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}^b_{\text{cw}} - \hat{\beta}_{\text{cw}}\right)^{\otimes 2}.$$

3. Simulation

To evaluate the performance of a suggested method for a binary time-varying covariate, a Monte Carlo simulation is conducted. Sample size n is set to be 100 and 200. Censoring times $\tau = (\tau_1, \ldots, \tau_n)$ are generated from a uniform distribution, $U(2.5, 3.5)$. For all resampling methods, the number of resampling is set to be 500. Each dataset is generated with the following schemes,

(i) Set $x_i(t) = 0$ at $t = 0$. Then generate u_i from a Poisson process with a following intensity function, $0.5\exp(\beta x_i(t))\rho_i$, where $\beta = 0.5$ or -0.5. For the correlation among gap times within a subject, a frailty term is incorporated and two cases are considered: (a) independence($\rho_i = 1$) and (b) $\rho \sim (\alpha, \alpha^{-1})$ with $\alpha = 0.1$. Define a recurrent event time as $s_{ij} = s_{ij-1} + u_i$ with $s_{i0} = 0$.

(ii) To determine whether the i^{th} subject has a time- varying covariate, generate an indicator, $\delta^x_i \sim$ Bernoulli(p_x). Three values(= 0.4, 0.6, 0.8) for p_x are applied to evaluate the effect of a proportion of time-varying covariates among n samples.

(iii) For a subject with $\delta^x_i = 1$, generate the occurrence time of the event changing covariate value, Z^*_i from $U(0, \tau_i - 0.1)$. Then $x_i(t)$ changes from 0 to 1 if $t > s_{ij}$.

For each scenario, we generate 500 datasets. Table 1 and Table 2 show the simulation results obtained from three methods under an independent assumption($\rho_i = 1$) and Table 3 and 4 present the results for correlated gap times. In each table, SSE is the sample standard deviation of 500 estimates and SEE is the mean of se($\hat{\beta}$)'s. For independent gap times, a MWCR method gives over estimation for both $\beta = 0.5$ and $\beta = -0.5$. The biases also increase as p_x increases. A EW method shows

Table 2: Comparison of simulation results with $\beta = -0.5$ and $\rho = 1$

p_x		$n = 100$				$n = 200$			
		$\hat{\beta}$	SSE	SEE	CP	$\hat{\beta}$	SSE	SEE	CP
0.4	MWCR	0.029	0.174	0.178	0.199	0.016	0.113	0.134	0.026
	EW	−0.421	0.333	0.140	0.999	−0.417	0.232	0.096	1.000
	CwWCR	−0.484	0.195	0.193	0.943	−0.486	0.145	0.142	0.946
0.6	MWCR	0.037	0.136	0.152	0.044	0.032	0.096	0.113	0.002
	EW	−0.444	0.278	0.119	1.000	−0.448	0.195	0.085	1.000
	CwWCR	−0.521	0.194	0.195	0.943	−0.508	0.127	0.138	0.958
0.8	MWCR	0.072	0.128	0.136	0.026	0.053	0.084	0.100	0.000
	EW	−0.475	0.247	0.110	0.100	−0.480	0.175	0.078	1.000
	CwWCR	−0.485	0.195	0.194	0.943	−0.480	0.132	0.137	0.950

Table 3: Comparison of simulation results with $\beta = 0.5$ and $\rho \sim G(0.1, 0.1)$

p_x		$n = 100$				$n = 200$			
		$\hat{\beta}$	SSE	SEE	CP	$\hat{\beta}$	SSE	SEE	CP
0.4	MWCR	0.696	0.278	0.293	0.892	0.702	0.086	0.087	0.359
	EW	0.102	0.253	0.381	0.988	0.232	0.060	0.192	0.970
	CwWCR	0.509	0.192	0.201	0.928	0.476	0.140	0.142	0.948
0.6	MWCR	0.706	0.278	0.293	0.892	0.715	0.075	0.077	0.229
	EW	0.271	0.253	0.381	0.996	0.257	0.051	0.168	0.960
	CwWCR	0.510	0.162	0.192	0.978	0.502	0.119	0.137	0.974
0.8	MWCR	0.757	0.225	0.230	0.800	0.742	0.070	0.069	0.068
	EW	0.339	0.236	0.527	0.992	0.283	0.052	0.157	0.954
	CwWCR	0.530	0.162	0.215	0.990	0.509	0.108	0.155	1.000

Table 4: Comparison of simulation results with $\beta = -0.5$ and $\rho \sim G(0.1, 0.1)$

p_x		$n = 100$				$n = 200$			
		$\hat{\beta}$	SSE	SEE	CP	$\hat{\beta}$	SSE	SEE	CP
0.4	MWCR	0.111	0.127	0.142	0.002	0.059	0.112	0.141	0.022
	EW	−0.410	0.135	0.330	1.000	−0.398	0.096	0.230	1.000
	CwWCR	−0.490	0.280	0.238	0.910	−0.505	0.183	0.168	0.934
0.6	MWCR	0.080	0.140	0.160	0.041	0.085	0.1096	0.118	0.000
	EW	−0.430	0.112	0.276	1.000	−0.429	0.079	0.194	1.000
	CwWCR	−0.483	0.237	0.224	0.936	−0.452	0.158	0.157	0.920
0.8	MWCR	0.111	0.126	0.142	0.002	0.104	0.089	0.105	0.000
	EW	−0.463	0.107	0.247	1.000	−0.465	0.072	0.174	1.000
	CwWCR	−0.488	0.210	0.215	0.937	−0.489	0.150	0.140	0.915

under estimated one for $\beta = 0.5$ and over estimated for $\beta = -0.5$. In particular, as sample size increases and correlation exist, an EW method brings worse results. However, as p_x increases, the biases decrease which is different with a MWCR's result. This result can be explained that an EW method uses all recurrent gap times from each subject while a MWCR method needs only one sample. Therefore, a MWCR method could not have a chance to reflect the change of covariate value. The intension to present both SSE and SEE is to validate the estimated standard error of $\hat{\beta}$ (*i.e.*, SEE) because it is also important to maintain a coverage probabilities. According to the result of tables, a EW method show SSE and SEE are so different and give unsuitable coverage probabilities. Relatively, a suggested approach (CwWCR) satisfies 95 coverage probabilities. The CwWCR method still shows very desirable results for dependence cases (Table 3 and Table 4).

Table 5: Application of resampling methods to YTOP data

	MWCR			EW			CwWCR		
	Est	SE	*P*-value	Est	SE	*P*-value	Est	SE	*P*-value
YTOP	0.0002	0.028	1.000	−0.698	0.222	0.002	−0.631	0.160	< 0.001
Gender(Male = 1)	0.0000	0.037	1.000	0.163	0.161	0.310	0.218	0.130	0.094

4. Data Analysis

In this section, we applied the suggested method to a Young Traffic Offenders Program(YTOP) data (Sun *et al.*, 2001). The data includes 192 drivers' conviction record since they got a driver license. Main interest is to evaluate the effect of an educational program which is experienced in the middle of a study. That is, every subject had included in a Non-YTOP group at start of study and then among 192 drivers, 98 drivers became program participants. We applied a suggested method to estimate the effect of two covariates; a time-varying covariate, YTOP participation($I(t >$participation time)) and a time independent covariate, gender(Male = 1, Female = 0). Table 5 shows the result of three methods. For a MWCR method, YTOP and gender have no significant effects. However, a EW method provides similar result with a CwWCR which is found at simulation study with a negative β. A YTOP group has a negative effect which means that the education program was effective to detain the recurrence of a conviction. For a gender, a positive coefficient means that male drivers had faster conviction times.

5. Concluding Remarks

Recurrent gap times are analyzed with several approaches. A resampling technique has been applied to consider an informative risk set. The purpose of this study is to extend a resampling method to estimate a time-varying covariate which often occurs in a longitudinal data. A simulation study was performed using a R-package in order to compare previous two methods which do not consider a time-varying covariate. According to results, these two resampling methods give biased estimates for a time-varying covariate while the suggested method results in unbiased at several situations. However, our suggestion is applied only to binary time dependent covariates. The extension to time-varying continuous covariate would be dealt with in a future study. We consider a weighted approach using a kernel method as a weighting procedure often applied to a missing covariate problem (Zhou and Wang, 2000).

References

Cai, J. and Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event types, *Lifetime Data Analysis*, **10**, 121–138.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*, Springer, New York.

Darlington, G. A. and Dixon, S. N. (2013). Event-weighted proportional hazards modeling for recurrent gap time data, *Statistics in Medicine*, **32**, 124–130.

Huang, Y. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events, *Lifetime Data Analysis*, **9**, 293–303.

Huang, C. Y., Qin, J. and Wang, M. C. (2010). Semiparametric analysis for recurrent event data with timedependent covariates and informative censoring, *Biometrics*, **66**, 39–49.

Lin, D. Y. and Ying, Z. (2001). Nonparametric tests for the gap time distribution of serial events based on censored data, *Biometrics*, **57**, 369–375.

Luo, X. and Huang, C. Y. (2011). Analysis of recurrent gap time data using the weighted risk-set

method and the modified within-cluster resampling method, *Statistics in Medicine*, **30**, 301–311.

Sun, L. Q., Park, D. H. and Sun, J. G. (2006). The additive hazard model for recurrent gap times, *Statistica Sinica*, **16**, 919–932.

Sun, J., Kim, Y., Hewett, J., Johnson, J. C., Farmer, J. and Gibler, M. (2001). Evaluation of traffic injury prevention programs using counting process approaches, *Journal of American Statistical Association*, **96**, 469–475.

Wang, M. C. and Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function, *Journal of the American Statistical Association*, **94**, 146–153.

Williamson, J. M., Kim, H. Y., Manatunga, A. and Addiss, D. G. (2008). Modeling survival data with informative cluster size, *Statistics in Medicine*, **27**, 543–555.

Zhao, X., Liu, L., Liu, Y. and Xu, W. (2012). Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring, *Biometrical Journal*, **54**, 585–599.

Zhou, H. and Wang, C. Y. (2000). Failure time regression with continuous covariate measured with error, *Journal of Royal Statistical Society B*, **62**, 657–665.

Received May 18, 2014; Revised July 5, 2014; Accepted August 21, 2014