

Variance estimation for clustered recurrent event data with a small number of clusters

Douglas E. Schaebel^{*,†}

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.

SUMMARY

Often in biomedical studies, the event of interest is recurrent and within-subject events cannot usually be assumed independent. In semi-parametric estimation of the proportional rates model, a working independence assumption leads to an estimating equation for the regression parameter vector, with within-subject correlation accounted for through a robust (sandwich) variance estimator; these methods have been extended to the case of clustered subjects. We consider variance estimation in the setting where subjects are clustered and the study consists of a small number of moderate-to-large-sized clusters. We demonstrate through simulation that the robust estimator is quite inaccurate in this setting. We propose a corrected version of the robust variance estimator, as well as jackknife and bootstrap estimators. Simulation studies reveal that the corrected variance is considerably more accurate than the robust estimator, and slightly more accurate than the jackknife and bootstrap variance. The proposed methods are used to compare hospitalization rates between Canada and the U.S. in a multi-centre dialysis study. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: bootstrap; clustered failure time data; jackknife; proportional means model; recurrent events; robust variance; sandwich estimator; semi-parametric model; variance estimation

1. INTRODUCTION

In many clinical and observational studies, the event of interest is recurrent; i.e. may occur repeatedly for an individual. Examples include hospitalizations, infections, acute myocardial infarctions, wheezing episodes and bouts of depression. In biomedical studies, it is not usually reasonable to assume that events within a subject are independent. Since the within-subject dependence structure may be complicated to model or simply not of interest, marginal modelling is often indicated. Examples of existing failure time models for clustered failure time data include the marginal hazard models proposed by Wei *et al.* [1], the common baseline hazard model of Lee *et al.* [2], and the mixed baseline hazard models of Spiekerman and

*Correspondence to: Douglas E. Schaebel, Department of Biostatistics, University of Michigan, 1420 Washington Heights, M4140 SPH-2, Ann Arbor, MI 48109-2029, U.S.A.

†E-mail: deschau@umich.edu

Lin [3] and Clegg *et al.* [4]. For recurrent event data, the mean number of events and the event occurrence rate are better intuited than the hazard function and many authors have proposed regression methods based on the mean or rate function. For example, Pepe and Cai [5] developed semi-parametric methods for modelling the rate function, wherein each numbered recurrent event has a distinct baseline rate. Lawless and Nadeau [6] considered modelling the mean number of events, and developed pertinent theory for the discrete-time case. Subsequently, Lin *et al.* [7] developed the asymptotic theory through empirical processes for the continuous time setting. The marginal mean/rate methods of Pepe and Cai [5], Lawless and Nadeau [6] and Lin *et al.* [7] each involve estimating the regression parameter as the solution to an estimating equation derived through a working independence assumption, with a robust (sandwich) estimator employed for the variance. Underlying each of these methods is the assumption that subjects are independent.

Cases often arise where, not only within-subject events, but study individuals are not independent. For example, in a retrospective study of hospitalization rates among renal failure patients, patients within the same geographic region may be correlated due to unmeasured patient characteristics which predispose them to poor health or propensity to utilize medical services. In an asthma study, children from the same neighbourhood may share certain environmental risk factors (e.g. air particulate levels) which may induce wheezing episodes. Or, in a multi-centre study of technique failures among patients on dialysis, patients from the same centre may be correlated due to centre-specific characteristics with respect to practice patterns. Since the model of Lin *et al.* [7] assumes that study subjects are independent, Schaubel and Cai [8] proposed two proportional rates models applicable when subjects are clustered; one model which had a common baseline rate across clusters, and a second which featured cluster-specific baseline rates.

The investigation which motivated the methods proposed in this report was a prospective international multi-centre study of peritoneal dialysis (PD) patients [9]. Patients who experience renal failure must undergo renal replacement therapy in order to remain alive. Although kidney transplantation is the preferred treatment method, patients usually begin on dialysis, and often remain on dialysis due to an ever-increasing shortfall in the availability of donor organs. There are two major forms of dialysis: haemodialysis (HD) and PD. In the context of publicly funded health care, as in Canada, PD is less costly; approximately 30 per cent of Canadian dialysis patients are put on PD. In the U.S., less than 10 per cent of dialysis patients receive PD. The discrepancy between Canada and the U.S. with respect to the percentage of patients receiving PD leads naturally to the question of whether patients in one country have better prognosis than the other. The Canada–U.S.A. (CANUSA [9]) PD study offers the opportunity to address this research question. Specifically, it is of interest to compare hospitalization rates between Canadian and American PD patients, adjusting for well-known prognostic factors.

The CANUSA study consists of 14 dialysis centres. Hospital admissions for a given patient cannot reasonably be assumed to be independent. Moreover, patients treated at the same centre cannot be assumed to be independent, since centre-specific practice patterns may differ markedly, and due to shared and unmeasured patient characteristics associated with living in the centre's catchment area. Thus, the independent units in the CANUSA study are the centres and, with only 14 centres, the robust variance estimator would not be expected to provide a reliable basis for inference. As such, an alternative method of estimating the variance is required.

We propose a corrected version of the robust variance estimator for use in analysing clustered recurrent event data with a small number of moderate-to-large-sized clusters of subjects. Through an extensive simulation study, we show that the robust variance estimator is very inaccurate with 15 or fewer clusters. The corrected variance estimator is shown to be significantly more accurate, yielding empirical coverage probabilities much closer to the nominal level. We also examine bootstrap and jackknife estimators, and find that the corrected robust variance estimator is more accurate than either the jackknife or bootstrap.

The remainder of this article is organized as follows. In Section 2, we formalize the problem and derive the proposed variance estimators. In Section 3, we describe a simulation study. Hospitalization data from the previously described CANUSA dialysis study are analysed in Section 4. Concluding remarks are contained in Section 5.

2. MODELS AND METHODS

In the development that follows, let n represent the number of clusters, with the number of subjects in cluster j denoted by n_j . Let $N_{ij}^*(t)$ denote the number of events experienced by the i th subject from the j th cluster as of time t . The covariate vector will be represented by $\mathbf{Z}_{ij}(s)$, wherein all time-dependent covariates are assumed to be ‘external’ in the sense of Kalbfleisch and Prentice [10]; i.e. not directly affected by the recurrent event process.

2.1. Rate/mean regression model

Lin *et al.* [7] proposed the following models for recurrent event data:

$$E[N_{ij}^*(t)|\mathbf{Z}_{ij}] = e^{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}} \mu_0(t) \quad (1)$$

$$E[dN_{ij}^*(t)|\mathbf{Z}_{ij}(t)] = e^{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(t)} d\mu_0(t) \quad (2)$$

the proportional means and rates models, respectively. Model (1) is a special case of (2), applicable when the covariate vector is not time-dependent. The event mean and rate functions are marginal, in the sense that they are not conditional on the event history, unlike the proportional intensity models (e.g. Reference [11]).

Assumptions underlying the above-listed models include independence of subjects. In situations subjects where clusters of subjects may be correlated, and when the within-subject and within-cluster event dependence structures are not of interest, marginal modelling is still indicated. Schaubel and Cai [8] recently proposed two proportional rates models for recurrent event data, given by

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s)] = e^{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)} d\mu_0(s) \quad (3)$$

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s)] = e^{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)} d\mu_{0j}(s) \quad (4)$$

where $d\mu_0(t)$ and $d\mu_{0j}(t)$ are unspecified baseline rate functions, $\boldsymbol{\beta}_0$ is a parameter vector, and $\mathbf{Z}_{ij}(s)$ is covariate vector. The choice between model (3) and (4) depends on the data structure and the aims of the investigator. For example, if cluster sizes are very small, model (4) is not practical. Even if cluster-sizes are moderate-to-large, other considerations may lead to preference of (3), particularly with regard to interpretation of the regression coefficient.

In both models, the parameter, $\boldsymbol{\beta}_0$, reflects comparisons between covariate levels; in (4), such comparisons are made among subjects within the same cluster, while the comparison is averaged over clusters in (3). For example, in the previously described CANUSA study, interest lies in comparing hospitalization rates between Canada and the U.S. Model (3) is the natural choice, since the effect of a cluster-level covariate cannot be addressed through model (4). That is, since all patients within a centre are necessarily from the same country, differences in hospital rates by country are not estimable through a model which stratifies by centre.

In practice, since the observation period of a study of finite duration, events for a subject may be censored. Let the observed number of events for subject i from cluster j be represented by $N_{ij}(t) = N_{ij}^*(t \wedge C_{ij})$, where $a \wedge b = \min(a, b)$ and C_{ij} denotes censoring time. It is assumed that C_{ij} is independent of $N_{ij}^*(\cdot)$, conditional on the covariate vector.

2.2. Results: large number of clusters

We begin by describing results applicable when the number of clusters is large. Schaubel and Cai [8] showed that a consistent estimator of $\boldsymbol{\beta}_0$ is given by $\hat{\boldsymbol{\beta}}$, the solution to the estimating equation $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^{\tau} \{\mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta})\} dN_{ij}(s) \quad (5)$$

with τ being a truncation point typically set to the maximum censoring time, $\mathbf{E}(s; \boldsymbol{\beta}) = \mathbf{S}^{(1)}(s; \boldsymbol{\beta})/S^{(0)}(s; \boldsymbol{\beta})$ and $\mathbf{S}^{(d)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} I(C_{ij} > s) \mathbf{Z}_{ij}(s)^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{ij}(s)}$ for $d = 0, 1, 2$, where, for a vector \mathbf{a} , $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. An estimator for the cumulative rate function is given by $\hat{\mu}_0(t; \hat{\boldsymbol{\beta}})$, where

$$\hat{\mu}_0(t; \boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t S^{(0)}(s; \boldsymbol{\beta})^{-1} dN_{ij}(s) \quad (6)$$

Under certain regularity conditions $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a Normal distribution with mean $\mathbf{0}$ and a covariance which can be consistently estimated by $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$, where

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} \hat{\boldsymbol{\Phi}}(\boldsymbol{\beta}) \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1} \quad (7)$$

with

$$\begin{aligned} \hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}) &= \int_0^{\tau} \mathbf{V}(s; \boldsymbol{\beta}) S^{(0)}(s; \boldsymbol{\beta}) d\hat{\mu}_0(s; \boldsymbol{\beta}) \\ \mathbf{V}(s; \boldsymbol{\beta}) &= \frac{\mathbf{S}^{(2)}(s; \boldsymbol{\beta})}{S^{(0)}(s; \boldsymbol{\beta})} - \mathbf{E}(s; \boldsymbol{\beta})^{\otimes 2} \\ \hat{\boldsymbol{\Phi}}(\boldsymbol{\beta}) &= n^{-1} \sum_{j=1}^n \hat{\boldsymbol{\Psi}}_j(\boldsymbol{\beta})^{\otimes 2} \\ \hat{\boldsymbol{\Psi}}_j(\boldsymbol{\beta}) &= \sum_{i=1}^{n_j} \int_0^{\tau} \{\mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta})\} d\hat{M}_{ij}(s; \boldsymbol{\beta}) \\ \hat{M}_{ij}(t; \boldsymbol{\beta}) &= N_{ij}(t) - \int_0^t I(C_{ij} > s) e^{\boldsymbol{\beta}^T \mathbf{Z}_{ij}(s)} d\hat{\mu}_0(s; \boldsymbol{\beta}) \end{aligned} \quad (8)$$

It is also shown that $n^{1/2}\{\hat{\mu}_0(t; \hat{\beta}) - \mu_0(t)\}$ converges weakly to a zero-mean Gaussian process with a covariance function which can be consistently estimated empirically.

The variance estimator in (7) converges almost surely to $\Sigma(\beta_0)$ as $n \rightarrow \infty$ and has been shown to be a good estimator of the variance of $\hat{\beta}$ when the number of clusters is at least 50. However, it may not be sufficiently accurate when the number of clusters is low, which could lead to confidence intervals for the elements of β_0 which are unrealistically narrow and, correspondingly, invalid inference. Since the CANUSA data set to be analysed is based on a sample of $n = 14$ centres, we cannot be confident of the accuracy of $\hat{\Sigma}(\hat{\beta})$.

In the next subsection, we propose three alternative methods to estimate the variance of $\hat{\beta}$ when fitting the proportional rates model to clustered recurrent event data with a small number of moderate-to-large-sized clusters.

2.3. Proposed variance estimators

It has been shown by Schaubel and Cai [8] that

$$n^{1/2}(\hat{\beta} - \beta_0) = \Omega(\beta_0)^{-1} n^{-1/2} \sum_{j=1}^n \Psi_j(\beta_0) + o_p(1) \tag{10}$$

where we define

$$\begin{aligned} \Omega(\beta) &= \int_0^\tau \mathbf{v}(s; \beta) \mathbf{s}^{(0)}(s; \beta) d\mu_0(s) \\ \Psi_j(\beta) &= \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \mathbf{e}(s; \beta)\} dM_{ij}(s; \beta) \end{aligned}$$

with $\mathbf{s}^{(d)}(s; \beta)$, $\mathbf{e}(s; \beta)$ and $\mathbf{v}(s; \beta)$ being the limiting values of $\mathbf{S}^{(d)}(s; \beta)$, $\mathbf{E}(s; \beta)$ and $\mathbf{V}(s; \beta)$, respectively. As stated, in (9), $\Psi_j(\beta_0)$ is estimated by $\hat{\Psi}_j(\hat{\beta})$, wherein $M_{ij}(t; \beta_0) = N_{ij}(t) - \int_0^t I(C_{ij} > s) e^{\beta_0^T \mathbf{Z}_{ij}(s)} d\mu_0(s; \beta_0)$ is replaced by $\hat{M}_{ij}(t; \hat{\beta}) = N_{ij}(t) - \int_0^t I(C_{ij} > s) e^{\hat{\beta}^T \mathbf{Z}_{ij}(s)} d\hat{\mu}_0(s; \hat{\beta})$. A concern is that the magnitude of the residual-type term $\hat{M}_{ij}(t; \hat{\beta})$ may significantly underestimate that of the true underlying error term, $M_{ij}(t; \beta_0)$. This is particularly a concern in small samples; or, more specifically, situations where there is a small number of independent units. In addition, under-coverage of the standard robust variance estimator may be exacerbated by the association among events within and among clustered subjects.

In an attempt to improve coverage probability (CP) of confidence intervals and the accuracy of hypothesis tests involving β_0 , we propose the following corrected version of the robust variance estimator:

$$\hat{\Sigma}^M(\hat{\beta}) = \hat{\Omega}(\hat{\beta})^{-1} \hat{\Phi}^M(\hat{\beta}) \hat{\Omega}(\hat{\beta})^{-1} \tag{11}$$

with $\hat{\Phi}^M(\beta) = n^{-1} \sum_{j=1}^n \{\hat{\Psi}_j^M(\beta)\}^{\otimes 2}$,

$$\begin{aligned} \hat{\Psi}_j^M(\beta) &= \left\{ \mathbf{I} + \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \mathbf{E}(s; \beta)\} d\hat{\mathbf{D}}_{ij}^T(s; \beta) \Omega(\beta)^{-1} \right\} \hat{\Psi}_j(\beta) \\ &\quad + n^{-1} \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \mathbf{E}(s; \beta)\} I(C_{ij} > s) e^{\beta_0^T \mathbf{Z}_{ij}(s)} \mathbf{S}^{(0)}(s; \beta)^{-1} dM_{\cdot j}(s; \beta) \end{aligned}$$

where \mathbf{I} is the identity matrix, $M_{\cdot j}(t; \boldsymbol{\beta}) = \sum_{i=1}^{n_j} M_{ij}(t; \boldsymbol{\beta})$ and

$$\hat{\mathbf{D}}_{ij}(t; \boldsymbol{\beta}) = -n \int_0^t I(C_{ij} > s) \{ \mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta}) \} e^{\boldsymbol{\beta}^T \mathbf{Z}_{ij}(s)} d\hat{\mu}_0(s; \boldsymbol{\beta}) \quad (12)$$

The corrected variance represents an attempt to improve upon the estimation of the error term, $M_{ij}(t; \boldsymbol{\beta}_0)$. The derivation is based on Taylor Series expansions and is outlined in Appendix A.

Natural choices of methods to generate small-sample estimators of the regression parameter include the jackknife and the bootstrap [12], which we now describe. In computing the jackknife estimator and its corresponding estimated variance, n estimators of $\boldsymbol{\beta}_0$ are computed, each with a different cluster deleted. We define $\hat{\boldsymbol{\beta}}_{(-j)}$ as the estimator based on the subsample which excluded cluster j . Pseudo-values are defined as $\hat{\boldsymbol{\xi}}_j = n\hat{\boldsymbol{\beta}} - (n-1)\hat{\boldsymbol{\beta}}_{(-j)}$, where, as before, $\hat{\boldsymbol{\beta}}$ is the estimator based on all clusters. The jackknife estimator of $\boldsymbol{\beta}_0$ and its associated variance estimate are given by

$$\hat{\boldsymbol{\beta}}^J = n^{-1} \sum_{j=1}^n \hat{\boldsymbol{\xi}}_j \quad (13)$$

$$\hat{\boldsymbol{\Sigma}}^J = \{n(n-1)\}^{-1} \sum_{j=1}^n (\hat{\boldsymbol{\xi}}_j - \hat{\boldsymbol{\beta}}^J)^{\otimes 2} \quad (14)$$

For the bootstrap, B samples are selected, each containing n clusters selected with replacement. The bootstrap estimator of $\boldsymbol{\beta}_0$ and its estimated variance are given by

$$\hat{\boldsymbol{\beta}}^B = B^{-1} \sum_{b=1}^B \hat{\boldsymbol{\beta}}_b \quad (15)$$

$$\hat{\boldsymbol{\Sigma}}^B = (B-1)^{-1} \sum_{b=1}^B (\hat{\boldsymbol{\beta}}_b - \hat{\boldsymbol{\beta}}^B)^{\otimes 2} \quad (16)$$

where $\hat{\boldsymbol{\beta}}_b$ is the estimator computed from the b th resample.

3. SIMULATION STUDY

In order to examine the operating characteristics of the variance estimators discussed in the previous section, we performed a simulation study. Since interest lies specifically in the case of a very small number of moderate-to-large-sized clusters, we set $n = 15$ (clusters) and examined cluster sizes of $n_j = 25$ and 50. We also examined the case where cluster size varies greatly, setting $n_j = 5 \times j$ for $j = 1, \dots, 15$, which results in a minimum and maximum cluster sizes of 5 and 75, respectively, and an average cluster size of $\bar{n}_j = 40$.

In simulating the event times, we first generated a latent (frailty) variate for each cluster, $Q_j \sim \text{Gamma}(\sigma_Q^{-2}, \sigma_Q^{-2})$, to induce positive correlation among subjects within a cluster. A subject-specific frailty, $R_i \sim \text{Gamma}(\sigma_R^{-2}, \sigma_R^{-2})$, was then generated in order to induce

correlation among within-subject event times. Note that $E[Q_j R_i] = 1$, while $\sigma_{QR}^2 \equiv V(Q_j R_i) = (\sigma_Q^2 + 1)(\sigma_R^2 + 1) - 1$, with Kendall's Tau rank correlation for any pair of events within-subject equal to $\sigma_{QR}^2(2 + \sigma_{QR}^2)^{-1}$. We chose $\sigma_Q^2 = 0.25, 0.5$ and $\sigma_R^2 = 0.5, 1.0$ such that, as in most practical situations, the correlation within-subject exceeds that within-cluster.

A covariate, Z_{ij} , was generated from the Bernoulli(0.5) distribution. For each subject, events were then simulated, with the k th event time for the i th subject within the j th cluster given by

$$T_{ijk} = T_{ijk-1} - \log\{1 - U_{ijk}\} \{Q_j R_i d\mu_0 e^{\beta_0 Z_{ij}}\}^{-1}$$

where the U_{ijk} 's are independent Uniform(0, 1) random variates and $T_{ij0} \equiv 0$. This set-up implies the following conditional and marginal models:

$$E[N_{ij}^*(t) | Z_{ij}, Q_j, R_j] = Q_j R_i d\mu_0 t e^{\beta_0 Z_{ij}}$$

$$E[N_{ij}^*(t) | Z_{ij}] = d\mu_0 t e^{\beta_0 Z_{ij}}$$

respectively. We set $\beta_0 = \log(2)$ and $d\mu_0 = 0.125, 0.25$ and 0.5 . Censoring times were generated as $C_{ij} \sim \text{Uniform}(0, 5)$, representing a study of length 5 time units with randomly staggered entry times and administrative censoring only.

Each data configuration was replicated 500 times. For each simulated data set, we estimated β_0 through (i) the solution to (5), denoted by $\hat{\beta}$ (ii) the jackknife (13), denoted $\hat{\beta}^J$ and (iii) the bootstrap (15), denoted $\hat{\beta}^B$. The standard error of $\hat{\beta}$ was estimated through the usual robust variance estimator given in (7), which is denoted $\hat{\Sigma}(\hat{\beta})$, as well as the corrected estimator in (A7), denoted $\hat{\Sigma}^M(\hat{\beta})$. Naturally, the variance of the jackknife and bootstrap estimators were estimated by (14) and (16), respectively.

Results of the simulation study are presented in Tables I and II. Note that Tables I and II are based on the same simulated data sets, to ensure comparability of the different estimators. In Table I, we list the bias and empirical standard deviation (ESD) of $\hat{\beta}$, as well as the average estimated standard error (ASE) and empirical 95 per cent CP based on the robust estimator and its corrected counterpart. The parameter estimator, $\hat{\beta}$, is approximately unbiased for β_0 . For almost all data configurations, the robust standard error greatly underestimates the ESD; correspondingly, empirical CPs are well below the nominal 95 per cent. The corrected variance estimator approximates the ESD with much greater accuracy, with empirical CP notably greater than the robust estimator. Empirical coverage averaged approximately 90 per cent for the robust variance and was less than 90 per cent for approximately half of the runs. CP averaged approximately 93 per cent for the corrected estimator. Even with the correction some under-coverage persisted, as CP was as low as 90 per cent for three of the 36 runs.

In Table II, we display simulation results pertaining to the jackknife and bootstrap estimators. Both $\hat{\beta}^J$ and $\hat{\beta}^B$ are approximately unbiased. For both methods, the accuracy of the estimated standard errors and 95 per cent empirical CP exceeds those of the robust variance estimator. However, the corrected variance is generally superior in both respects. For all configurations, CP associated with $\hat{\Sigma}^M$ is at least as great as the CP for either $\hat{\Sigma}^J$ or $\hat{\Sigma}^B$. For $n_j = 25$, the jackknife and bootstrap variance estimators perform virtually equal. Empirical coverage is slightly greater for the jackknife than the bootstrap for $n_j = 50$, although the opposite

Table I. Simulation results: robust variance estimator and proposed corrected version.

n_j	σ_Q^2	σ_R^2	$d\mu_0(t)$	$\hat{\beta}$		ASE		CP	
				Bias	ESD	$\hat{\Sigma}(\hat{\beta})$	$\hat{\Sigma}^M(\hat{\beta})$	$\hat{\Sigma}(\hat{\beta})$	$\hat{\Sigma}^M(\hat{\beta})$
25	0.25	0.5	0.125	0.008	0.194	0.182	0.196	0.92	0.94
			0.25	0.013	0.150	0.147	0.160	0.93	0.94
			0.5	0.006	0.134	0.125	0.141	0.91	0.94
		1.0	0.125	0.002	0.229	0.201	0.216	0.89	0.91
			0.25	-0.016	0.195	0.171	0.186	0.90	0.92
			0.5	-0.009	0.166	0.148	0.166	0.89	0.93
	0.5	0.5	0.125	-0.008	0.215	0.189	0.204	0.89	0.91
			0.25	0.004	0.175	0.153	0.169	0.88	0.90
			0.5	0.006	0.147	0.138	0.159	0.91	0.95
		1.0	0.125	-0.007	0.249	0.212	0.228	0.87	0.90
			0.25	0.006	0.196	0.177	0.193	0.89	0.93
			0.5	-0.018	0.186	0.159	0.180	0.89	0.93
50	0.25	0.5	0.125	0.002	0.137	0.130	0.141	0.92	0.94
			0.25	0.006	0.112	0.104	0.116	0.90	0.93
			0.5	0.000	0.099	0.089	0.105	0.90	0.94
		1.0	0.125	-0.001	0.157	0.141	0.152	0.89	0.92
			0.25	0.006	0.133	0.121	0.133	0.92	0.94
			0.5	0.002	0.119	0.108	0.126	0.92	0.94
	0.5	0.5	0.125	0.001	0.156	0.135	0.147	0.89	0.92
			0.25	0.007	0.127	0.112	0.127	0.89	0.94
			0.5	-0.003	0.111	0.097	0.121	0.90	0.94
		1.0	0.125	0.002	0.163	0.151	0.164	0.91	0.93
			0.25	0.005	0.142	0.128	0.143	0.91	0.93
			0.5	-0.008	0.136	0.116	0.139	0.88	0.94
$5 \times j$	0.25	0.5	0.125	-0.012	0.149	0.139	0.152	0.91	0.92
			0.25	-0.003	0.123	0.113	0.127	0.89	0.93
			0.5	-0.005	0.110	0.094	0.113	0.89	0.95
		1.0	0.125	-0.005	0.170	0.155	0.170	0.91	0.93
			0.25	0.004	0.140	0.131	0.147	0.91	0.94
			0.5	-0.003	0.133	0.116	0.134	0.89	0.93
	0.5	0.5	0.125	-0.005	0.156	0.143	0.158	0.92	0.94
			0.25	0.005	0.147	0.120	0.138	0.86	0.90
			0.5	0.002	0.125	0.104	0.129	0.86	0.95
		1.0	0.125	0.006	0.186	0.161	0.178	0.90	0.92
			0.25	-0.010	0.156	0.140	0.158	0.88	0.92
			0.5	-0.004	0.140	0.127	0.153	0.91	0.96

Based on $n = 15$ clusters; $n_j =$ cluster size, $\sigma_Q^2 =$ variance of cluster-specific frailty, $\sigma_R^2 =$ variance of subject-specific frailty, $d\mu_0(t) =$ baseline event rate, $\hat{\beta} =$ solution to (5), ASE = average estimated standard error, CP = empirical 95 per cent coverage probability, ESD = empirical standard deviation, $\hat{\Sigma}(\hat{\beta}) =$ uncorrected robust variance estimator, (7); $\hat{\Sigma}^M(\hat{\beta}) =$ corrected robust variance estimator, (A7); five hundred simulations were performed for each data configuration.

Table II. Simulation results: jackknife and bootstrap estimators.

n_j	σ_Q^2	σ_R^2	$d\mu_0(t)$	$\hat{\beta}_J$				$\hat{\beta}_B$			
				Bias	ESD	ASE	CP	Bias	ESD	ASE	CP
25	0.25	0.5	0.125	0.003	0.195	0.197	0.93	0.014	0.195	0.187	0.92
			0.25	0.011	0.152	0.159	0.94	0.016	0.149	0.150	0.94
			0.5	0.004	0.137	0.136	0.93	0.008	0.130	0.127	0.92
		1.0	0.125	-0.003	0.231	0.218	0.91	0.006	0.231	0.207	0.89
			0.25	-0.019	0.199	0.186	0.92	-0.014	0.192	0.177	0.91
			0.5	-0.011	0.170	0.161	0.91	-0.010	0.164	0.150	0.90
	0.5	0.5	0.125	-0.014	0.218	0.210	0.90	-0.003	0.215	0.197	0.89
			0.25	0.002	0.179	0.170	0.90	0.006	0.173	0.156	0.89
			0.5	0.004	0.154	0.154	0.93	0.007	0.142	0.141	0.92
		1.0	0.125	-0.015	0.254	0.237	0.90	0.000	0.250	0.221	0.89
			0.25	0.002	0.202	0.197	0.92	0.009	0.194	0.183	0.91
			0.5	-0.020	0.195	0.177	0.91	-0.018	0.182	0.163	0.90
50	0.25	0.5	0.125	0.000	0.139	0.139	0.94	0.004	0.138	0.132	0.93
			0.25	0.005	0.114	0.112	0.92	0.007	0.110	0.105	0.91
			0.5	-0.001	0.102	0.096	0.92	0.000	0.098	0.090	0.90
		1.0	0.125	-0.003	0.158	0.151	0.91	0.002	0.158	0.144	0.89
			0.25	0.005	0.135	0.130	0.93	0.006	0.131	0.122	0.93
			0.5	0.001	0.122	0.117	0.93	0.003	0.117	0.110	0.93
	0.5	0.5	0.125	-0.001	0.158	0.148	0.91	0.004	0.155	0.140	0.91
			0.25	0.006	0.131	0.124	0.92	0.008	0.124	0.115	0.92
			0.5	-0.004	0.117	0.108	0.92	-0.002	0.107	0.100	0.92
		1.0	0.125	-0.001	0.168	0.166	0.93	0.004	0.160	0.155	0.93
			0.25	0.004	0.146	0.142	0.92	0.004	0.140	0.131	0.92
			0.5	-0.009	0.142	0.129	0.91	-0.007	0.131	0.119	0.90
$5 \times j$	0.25	0.5	0.125	-0.015	0.151	0.152	0.92	-0.010	0.148	0.145	0.92
			0.25	-0.006	0.125	0.125	0.92	-0.001	0.122	0.117	0.92
			0.5	-0.007	0.114	0.104	0.90	-0.005	0.108	0.099	0.92
		1.0	0.125	-0.009	0.172	0.172	0.93	-0.001	0.170	0.162	0.92
			0.25	0.001	0.142	0.145	0.92	0.006	0.139	0.137	0.92
			0.5	-0.004	0.136	0.128	0.91	-0.003	0.131	0.121	0.93
	0.5	0.5	0.125	-0.008	0.162	0.162	0.93	-0.002	0.155	0.152	0.93
			0.25	0.003	0.154	0.137	0.89	0.007	0.143	0.127	0.90
			0.5	0.003	0.132	0.118	0.89	0.002	0.123	0.110	0.89
		1.0	0.125	0.001	0.192	0.184	0.92	0.009	0.183	0.171	0.92
			0.25	-0.011	0.163	0.160	0.91	-0.009	0.153	0.148	0.91
			0.5	-0.005	0.147	0.145	0.93	-0.003	0.137	0.136	0.94

Based on $n = 15$ clusters; n_j = cluster size, σ_Q^2 = variance of cluster-specific frailty, σ_R^2 = variance of subject-specific frailty, $d\mu_0(t)$ = baseline event rate, $\hat{\beta}$ = solution to (5), ASE = average estimated standard error, CP = empirical 95 per cent coverage probability, ESD = empirical standard deviation, $\hat{\beta}^J$ = jackknife estimator, $\hat{\beta}^B$ = bootstrap estimator, five hundred simulations were preformed for each data configuration.

holds for the variable cluster size setting. Across all configurations, CPs were approximately 92 per cent for the jackknife and 91 per cent for the bootstrap variance estimators. Note that results were very similar to those presented in Tables I and II when the null case, $\beta_0 = 0$, was examined.

4. APPLICATION

We applied the proposed methods to the analysis of hospitalization rates in the previously described Canada–U.S.A. PD (CANUSA) Study. Our objectives were to compare Canadian and American PD patients with respect to hospitalization rates, adjusting for known prognostic factors, and to examine the various methods of variance estimation in real-data setting.

CANUSA was a prospective cohort study of 680 PD patients from 14 renal centres in the U.S. and Canada, with cluster sizes (patients per centre) ranging from 11 to 95. Patients were enrolled between September 1, 1990 and December 31, 1992, with follow-up until December 31, 1993. Patients were followed from the time they began PD until the earliest of death, receipt of a kidney transplant, technique failure or withdrawal from dialysis. A previous mortality analysis of the CANUSA data revealed that U.S. patients experienced significantly higher covariate-adjusted mortality rates, relative to Canadian patients [13]. In addition to mortality, the quality of life of these patients while they are alive and receiving PD is also of great interest. The number of hospitalizations is a well-accepted measure of patient morbidity, due to its concreteness, lack of subjectivity and ease of interpretation.

In total, there were 1157 hospitalizations among the CANUSA cohort. Since our objective was to analyse hospitalization patterns among patients receiving PD, we fitted the following conditional rate model:

$$E[dN_{ij}^*(t) | \mathbf{Z}_{ij}(t), PD_{ij}(t) = 1] = e^{\beta_0^\top \mathbf{Z}_{ij}(t)} d\mu_0(t)$$

where $PD_{ij}(t) = 1$ if the i th patient from centre j is receiving PD at time t and 0 otherwise. Note that use of a conditional model avoids issues pertaining to the dependent censoring due to death (90 patients), transplantation (137 patients), technique failure (118 patients) or withdrawal (18 patients). Methods for the proportional rates model described in Section 2 can be applied upon re-defining the at-risk process as $I(C_{ij} \geq t)PD_{ij}(t)$.

Results of our analysis are displayed in Table III, where the estimated regression coefficient, $\hat{\beta}$, represents the covariate-adjusted log rate ratio for patients in the U.S. relative to Canada. The corresponding hospitalization rate ratio (U.S.:Canada) is, therefore, estimated at $e^{\hat{\beta}}$. Note that p -values are based on the Wald test. We first computed $\hat{\beta}$ as the solution to (5), which yields $\hat{\beta} = 0.236$ and an estimated rate ratio of $e^{0.236} = 1.27$. Several estimators of the variance of $\hat{\beta}$ were then employed, the first of which is given by $\hat{\Omega}(\hat{\beta})^{-1}$, with $\hat{\Omega}(\hat{\beta})$ as defined in (8). This is the variance estimator derived by Andersen and Gill [11] for the proportional intensity model under the assumption of conditionally independent increments. Based on (8), the standard error is estimated at $\widehat{SE}(\hat{\beta}) = 0.085$, which, given the stringent underlying independence assumptions, we expect to greatly underestimate the true variability in $\hat{\beta}$. The Wald test based on $\hat{\Omega}(\hat{\beta})^{-1}$ implies that hospitalization rates are significantly elevated among U.S. patients, relative to those in Canada, with $p = 0.006$.

Table III. Analysis of Canada–U.S.A. (CANUSA) multi-centre PD study: parameter estimator and standard error estimates for log hospitalization rate ratio (U.S.A.:Canada).

Assumed independence	β_0 estimator (eq.)	$V(\hat{\beta})$ estimator (eq.)	β_0 estimate	\widehat{SE}	p
w/in cluster, w/in subject	$\hat{\beta}$ (5)	$\hat{\Omega}(\hat{\beta})^{-1}$ (8)	0.236	0.085	0.006
w/in cluster, w/in subject	$\hat{\beta}$ (5)	(17)	0.236	0.107	0.028
w/in cluster, w/in subject	$\hat{\beta}$ (5)	(18)	0.236	0.123	0.055
w/in cluster, w/in subject	$\hat{\beta}$ (5)	$\hat{\Sigma}(\hat{\beta})$ (7)	0.236	0.141	0.094
w/in cluster, w/in subject	$\hat{\beta}$ (5)	$\hat{\Sigma}^M(\hat{\beta})$ (A7)	0.236	0.246	0.337
w/in cluster, w/in subject	$\hat{\beta}^J$ (13)	$\hat{\Sigma}^J(\hat{\beta})$ (14)	0.243	0.195	0.213
w/in cluster, w/in subject	$\hat{\beta}^B$ (15)	$\hat{\Sigma}^B(\hat{\beta})$ (16)	0.238	0.183	0.192

$\hat{\beta}$ = solution to (5), $\hat{\beta}^J$ = jackknife estimator, $\hat{\beta}^B$ = bootstrap estimator, $\hat{\Omega}(\beta)^{-1}$ = variance estimator based on Martingale theory, $\hat{\Sigma}(\beta)$ = uncorrected robust variance estimator, (7); $\hat{\Sigma}^M(\hat{\beta})$ = corrected robust variance estimator, (A7); $\hat{\Sigma}^J(\beta)^M$ = jackknife variance estimator, $\hat{\Sigma}^B(\beta)^M$ = bootstrap variance estimator, \widehat{SE} = estimated standard error.

We next estimated the variance as

$$\hat{\Omega}(\hat{\beta})^{-1} \left(\sum_{j=1}^n \sum_{i=1}^{n_j} N_{ij}(\tau) \right)^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \{ \mathbf{Z}_{ij} - \mathbf{E}(s; \hat{\beta}) \} d\hat{M}_{ij}(s; \hat{\beta}) \}^{\otimes 2} \hat{\Omega}(\hat{\beta})^{-1} \tag{17}$$

derived under the proportional rates model under the assumption that within-patient and within-centre hospitalizations are independent. Based on (17), $\widehat{SE}(\hat{\beta}) = 0.107$, larger than the estimated SE based on (8), but still expected the greatly understate the true SE. From the Wald test based on (17), hospitalization rates among U.S. patients still appear to be significantly increased, with $p = 0.028$.

We then employed a variance estimator which accounts for the dependence among events within-patient, but assumes the independence of patients treated at the same centre:

$$\hat{\Omega}(\hat{\beta})^{-1} \left(\sum_{j=1}^n n_j \right)^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \left\{ \int_0^\tau \{ \mathbf{Z}_{ij} - \mathbf{E}(s; \hat{\beta}) \} d\hat{M}_{ij}(s; \hat{\beta}) \right\}^{\otimes 2} \hat{\Omega}(\hat{\beta})^{-1} \tag{18}$$

This is the variance estimator proposed by Lin *et al.* [7] for recurrent event data, but with independent subjects. With $\widehat{SE}(\hat{\beta}) = 0.123$, the Wald test yielded a borderline-significant result ($p = 0.055$).

We next employed the variance estimator proposed by Schaubel and Cai [8] and listed in (7), which accounts for the dependence among hospitalizations within a patient and the clustering of patients within centres. The corresponding standard error is estimated at $\widehat{SE}(\hat{\beta}) = 0.141$, which is larger than previous estimates which assumed within-centre and/or within-patient independence, but still yields a moderately interesting p -value ($p < 0.10$). Since the CANUSA study consisted of only $n = 14$ centres, (7), which was only advocated for $n \geq 50$, may still underestimate $V(\hat{\beta})$.

When the proposed corrected version of the robust variance estimator, given in (11), was used, the estimated variance is notably increased, with $\widehat{SE}(\hat{\beta}) = 0.246$, leading to a highly non-significant Wald test ($p = 0.337$).

The jackknife and bootstrap variance estimators, given by (14) and (16), respectively, far exceed the robust variance estimator, but are less than the corrected estimator, with $\widehat{SE}^J(\hat{\beta}^J) = 0.213 > \widehat{SE}^B(\hat{\beta}^B) = 0.192$. These results are quite consistent with the patterns observed in the simulation study, with respect to the rank ordering of the variance estimators. The magnitude of the discrepancy between the standard robust variance estimator and the jackknife, bootstrap and (particularly) the corrected estimator exceeds that typically observed in the simulation study. Under-coverage of the robust standard estimator is due to underestimation of the 'score' process variance, which results from estimating β_0 as well as the inter-event correlations among and within subjects. It is difficult to discern exactly which of these factors is most at work, particularly since methods for measuring correlation among recurrent event sequences are not well-developed.

Therefore, after accounting for the lack of independence among hospitalizations within-patient and patients within-centre, and accounting for the small number of clusters, it appears that the previously reported 27 per cent increase in the hospitalization rate for U.S. (*versus* Canadian) PD patients may be attributable to chance alone.

5. DISCUSSION

The robust (sandwich) variance estimator has been shown to be inaccurate in the recurrent event data setting with a small number of independent units. This was demonstrated in Reference [14], who studied the setting wherein subjects are independent and may experience recurrent events of multiple types; no solution to the problem of under-coverage was proposed. The degree of underestimation in the large-sample standard errors has been shown to increase with the degree of within-subject correlation, and the under-estimation is more pronounced when subjects are clustered (i.e. when adding a second level of dependence to the data structure). Underestimation of the variance could lead to confidence intervals with true coverage notably less than the nominal level, in addition to invalid inference.

We propose a corrected version of the robust variance estimator for the regression coefficient in the proportional rates model, with application to studies involving a small number of moderate-to-large-sized clusters. Simulation studies reveal that the proposed variance estimator leads to much more accurate standard error estimates and empirical coverage probabilities, in

situations where the robust estimator is shown to be quite inaccurate, although the correction factor does not completely eliminate the under-coverage. Jackknife and bootstrap estimators are also examined, with the jackknife appearing to be slightly more accurate than the bootstrap, but not as accurate as the corrected robust estimator. With respect to computation, the bootstrap involved routine application of SAS macros, while the jackknife and corrected robust estimator also involved SAS-IML (SAS Institute; Cary, NC).

When we applied the proposed and other methods to the analysis of hospitalization rates among dialysis patients, inference was found to depend greatly on the nature of the variance estimator. Results ranged from highly significant, under the assumption that within-patient hospitalizations are independent, to highly non-significant, after accounting for the within-centre clustering of patients. Our analysis offers no evidence that hospitalization rates are unequal between PD patients in Canada and the U.S., despite the fact that the percentage of dialysis patients receiving PD is much higher in Canada.

In the real-data application, a conditional rate function was modelled, $E[dN_{ij}^*(t)|PD_{ij}(t) = 1]$; i.e. the hospitalization rate while patients were receiving PD and, hence, prior to death, transplantation and technique failure. The objective of the analysis was to compare hospitalization rates by country, specifically among PD patients. The experience of patients following transplantation or technique failure is not of interest, here. The specification of the conditional rate function is somewhat unorthodox. However, in the presence of death, to jointly describe the (mortality, event) experience of the cohort, one could model death through standard univariate survival analysis, then model the event rate among patients while they survive. The mortality component of the analysis having already been conducted [13], we turn our attention to the hospitalization rate, conditional on survival. The conditional rate function modelled is analogous to the cause-specific hazard function widely used in competing risks analysis, as mentioned in Reference [7]. Since the rate model explicitly conditions on the non-occurrence, at time t , of the terminal events, censoring via terminal event does not constitute dependent censoring. A drawback to the conditional approach is the interpretation. However, the primary objective of the current article is to address the issue of clustered subjects, not terminating events. An alternative would be a marginal model for the event rate, as proposed by Ghosh and Lin [15]. The analysis of recurrent events in the presence of death and other terminating events is discussed in detail by Cook and Lawless [16].

To the best of our knowledge, no corrected version of the robust variance estimator has been proposed for use in recurrent event data. Mancl and DeRouen [17] developed a corrected variance for use in generalized estimating equation modelling, while Fay and Graubard [18] proposed an adjusted sandwich estimator for use in the Cox regression model. Both of these methods involve expanding residual-term elements to account for the use of estimated, as opposed to known, parameters.

APPENDIX A

In deriving a corrected version of $\hat{\Psi}_f(\boldsymbol{\beta})$, we begin by noting that

$$M_{ij}(t; \boldsymbol{\beta}_0) = \hat{M}_{ij}(t; \hat{\boldsymbol{\beta}}) - \{\hat{M}_{ij}(t; \hat{\boldsymbol{\beta}}) - \hat{M}_{ij}(t; \boldsymbol{\beta}_0)\} \quad (\text{A1})$$

$$- \{\hat{M}_{ij}(t; \boldsymbol{\beta}_0) - M_{ij}(t; \boldsymbol{\beta}_0)\} \quad (\text{A2})$$

Naturally, we can compute $\hat{M}_{ij}(t; \hat{\boldsymbol{\beta}})$ after obtaining $\hat{\boldsymbol{\beta}}$ and $\hat{\mu}_0(t; \hat{\boldsymbol{\beta}})$; but, it is necessary to estimate (A1) and (A2). Beginning with (A1), using a Taylor series expansion

$$-\{\hat{M}_{ij}(t; \hat{\boldsymbol{\beta}}) - \hat{M}_{ij}(t; \boldsymbol{\beta}_0)\} = \hat{\mathbf{D}}_{ij}^{\text{T}}(s; \boldsymbol{\beta}_*) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \quad (\text{A3})$$

where $\hat{\mathbf{D}}_{ij}(s; \boldsymbol{\beta}_*) = -\partial \hat{M}_{ij}(t; \boldsymbol{\beta}) / \partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_*}$ and $\boldsymbol{\beta}_*$ lies on the line segment joining $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. The derivative is given by

$$\frac{\partial \hat{M}_{ij}(t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = n \int_0^t I(C_{ij} > s) \{ \mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta}) \} e^{\boldsymbol{\beta}^{\text{T}} \mathbf{Z}_{ij}(s)} d\hat{\mu}_0(s; \boldsymbol{\beta}) \quad (\text{A4})$$

Applying (10), in addition to the strong convergence of $\hat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_0$, we have

$$-\{\hat{M}_{ij}(t; \hat{\boldsymbol{\beta}}) - \hat{M}_{ij}(t; \boldsymbol{\beta}_0)\} = \hat{\mathbf{D}}_{ij}^{\text{T}}(s; \boldsymbol{\beta}_0) \boldsymbol{\Omega}(\boldsymbol{\beta}_0)^{-1} \sum_{j=1}^n \boldsymbol{\Psi}_j(\boldsymbol{\beta}_0) \quad (\text{A5})$$

In considering a representation for (A2),

$$\begin{aligned} -\{\hat{M}_{ij}(t; \boldsymbol{\beta}_0) - M_{ij}(t; \boldsymbol{\beta}_0)\} &= \int_0^t I(C_{ij} > s) e^{\boldsymbol{\beta}^{\text{T}} \mathbf{Z}_{ij}(s)} d\{\hat{\mu}_0(s; \boldsymbol{\beta}) - \mu_0(s)\} \\ &= n^{-1} \int_0^t I(C_{ij} > s) e^{\boldsymbol{\beta}^{\text{T}} \mathbf{Z}_{ij}(s)} n^{-1} S^{(0)}(s; \boldsymbol{\beta}_0)^{-1} dM(s; \boldsymbol{\beta}_0) \end{aligned} \quad (\text{A6})$$

where $M(t; \boldsymbol{\beta}) = \sum_{j=1}^n \sum_{i=1}^{n_j} M_{ij}(t; \boldsymbol{\beta})$. Combining (A5) and (A6), then eliminating mean-zero cross-product terms involving uncorrelated subjects from different clusters, we generate a corrected version of $\hat{\boldsymbol{\Psi}}_j(\hat{\boldsymbol{\beta}})$, which we denote by $\hat{\boldsymbol{\Psi}}_j^M(\hat{\boldsymbol{\beta}})$, where

$$\begin{aligned} \hat{\boldsymbol{\Psi}}_j^M(\boldsymbol{\beta}) &= \left\{ \mathbf{I} + \sum_{i=1}^{n_j} \int_0^{\tau} \{ \mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta}) \} d\hat{\mathbf{D}}_{ij}^{\text{T}}(s; \boldsymbol{\beta}) \boldsymbol{\Omega}(\boldsymbol{\beta})^{-1} \right\} \hat{\boldsymbol{\Psi}}_j(\boldsymbol{\beta}) \\ &\quad + n^{-1} \sum_{i=1}^{n_j} \int_0^{\tau} \{ \mathbf{Z}_{ij}(s) - \mathbf{E}(s; \boldsymbol{\beta}) \} I(C_{ij} > s) e^{\boldsymbol{\beta}_0^{\text{T}} \mathbf{Z}_{ij}(s)} S^{(0)}(s; \boldsymbol{\beta})^{-1} dM_{.j}(s; \boldsymbol{\beta}) \end{aligned}$$

where \mathbf{I} is the identity matrix, $M_{.j}(t; \boldsymbol{\beta}) = \sum_{i=1}^{n_j} M_{ij}(t; \boldsymbol{\beta})$.

Thus, a corrected variance estimator for $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\Sigma}}^M(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\beta}})^{-1} \hat{\boldsymbol{\Phi}}^M(\hat{\boldsymbol{\beta}}) \hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\beta}})^{-1} \quad (\text{A7})$$

where $\hat{\boldsymbol{\Phi}}^M(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n \{ \hat{\boldsymbol{\Psi}}_j^M(\boldsymbol{\beta}) \}^{\otimes 2}$.

ACKNOWLEDGEMENTS

The author thanks David Churchill and Kevin Thorpe of McMaster University in Hamilton, Ontario for providing access to the CANUSA database and to the reviewers for their very constructive comments on the original version of the article.

REFERENCES

1. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* 1989; **84**:1065–1073.
2. Lee EW, Wei LJ, Amato DA. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, Klein JP, Goel PK (eds). Kluwer: Dordrecht, 1992.
3. Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* 1998; **93**:1164–1175.
4. Clegg LX, Cai J, Sen PK. A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* 1999; **55**:805–812.
5. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* 1993; **88**:811–820.
6. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**:158–168.
7. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society B* 2000; **62**:711–730.
8. Schaubel DE, Cai J. Proportional rates/means models for clustered recurrent event data, submitted.
9. Canada–U.S.A. Peritoneal Dialysis Study Group. Adequacy of dialysis and nutrition in continuous peritoneal dialysis: association with clinical outcomes. *Journal of the American Society of Nephrology* 1996; **7**:198–207.
10. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: Hoboken, 2002.
11. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1100–1120.
12. Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM: Philadelphia, PA, 1982.
13. Churchill DN, Thorpe KE, Vonesh EF, Keshaviah PR. Lower probability of patient survival with continuous peritoneal dialysis in the United States compared with Canada. *Journal of the American Society of Nephrology* 1997; **8**:965–971.
14. Cai J, Schaubel DE. Marginal means and rates models for multiple-type recurrent event data. *Lifetime Data Analysis* 2004; **10**:121–138.
15. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica* 2002; **12**:663–688.
16. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 1997; **16**:911–924.
17. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**:126–134.
18. Fay MP, Graubard BI. Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics* 2001; **57**:1198–1206.