Jurnal
Teknologi

Full paper

Proportional Odds Model Affiliate with Random-Effect Longitudinal Model

Chin Wan Yoke[a], Zarina Mohd Khalid[a*]

[a] *Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

*Corresponding author: zarinamkhalid@utm.my

Article history

Received :21 January 2013
Received in revised form :
7 May 2013
Accepted :25 June 2013

Graphical abstract

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Abstract

Joint survival-longitudinal analysis gains popularity in recent clinical studies. A proportional hazards (PH) model in survival sub-model is commonly an alternative path to simplify a complex covariates hazard model into a regression model. The PH model however closed only to the Weibull distribution, brought about inappropriate application for the log-logistic observations. Proportional odds (PO) model in that case raised forward to perform similarly with the PH model. The subsequent modelling study is therefore producing a joint PO-longitudinal analysis rather than a widely applicable joint PH-longitudinal analysis. Latent parameters is introduced as a linkage technique between the two sub-models. Investigation in this study relies on the simulation statistics in which the survival time-to-event data and longitudinal measurements are both influenced by a covariate effects. The repeatedly measures data additionally allow for different kind of missingness mechanisms. Maximum likelihood estimation method is applied to the joint model parameters estimation. The performance of the joint model and separated sub-models are then be compared. The illustrated results contributed better estimators on the joint model instead of separated model.

Keywords: Joint model; longitudinal model; maximum likelihood estimation; proportional odds (PO) model

Abstrak

Analisis bersama survival-longitudinal menguntungkan populariti dalam kajian klinikal baru-baru ini. Model bahaya berkadaran (PH) dalam survival sub-model biasanya merupakan laluan alternatif untuk memudahkan model bahaya covariat yang rumit ke dalam model regresi. Model PH bagaimanapun dianggar hanya dengan taburan Weibull, membawa permohonan yang tidak sesuai untuk pendapatan log-logistik. Model kemungkinan berkadar (PO) dalam hal ini dibangkitkan untuk melaksanakan tugas yang lebih kurang sama dengan model PH. Kajian model yang berikutnya menghasilkan analisis bersama PO-longitudinal dan bukannya analisis bersama PH-longitudinal yang dipohon secara meluas. Parameter pendam diperkenalkan sebagai satu teknik hubungan antara kedua-dua sub-model. Penyiasatan dalam kajian ini bergantung kepada statistik simulasi di mana survival data masa-ke-kejadian dan pengukuran longitudinal kedua-duanya dipengaruhi oleh kesan covariate. Data pengukuran berulang tambahan pula membenarkan pelbagai jenis mekanisme missingness. Kaedah penganggaran kebolehjadian maksimum digunakan untuk anggaran parameter model bersama. Prestasi model bersama dan sub-model yang dipisahkan dibandingkan seterusnya. Keputusan yang ditunjukkan menyumbang penganggar yang lebih baik pada model bersama dan bukannya model yang dipisahkan.

Kata kunci: Model bersama; model longitudinal; anggaran kebolehjadian maksimum; model kemungkinan berkadar (PO)

■1.0 INTRODUCTION

Joint model – a full model defined by the joint distribution between the separated models. It is frequently applying in the recent clinical studies of which the effect of the time-dependent covariates to the hazard of an event is of interest. The corresponding application study is subsequently contributes a widespread joint proportional hazards (PH) and longitudinal modeling in survival analysis. PH model in survival study has the ability to simplify the complex model into a regression model. Therefore, it is suit to the covariates-effects hazard model. According to Cox [1], the full model of the covariates hazard function can then easily be estimated by maximizing the partial likelihood.

PH model in addition closely related to a specific fundamental lifetime distributions, for instance Weibull function in such a way that the characteristics of Weibull distribution is closed under the PH model [2]. Unfortunately, this profit is failed to deal with the log-logistic survival outcomes. Proportional odds (PO) model acts as an alternative path to achieve analogous ability with PH model, takes the advantage to model the covariates-effects odds model for the log-logistic distribution that is closed under the PO model. Collett [3] explained that the PO model is unpopular be applied due to the reasons: 1) PO model performed similarly to the PH model in the presence of time-dependent variable to produce a non-proportional hazards model, and 2) PO model is lacked of the computer software to carry out the analysis.

The entire review studies are only focused on the joint PH-longitudinal model application (Wulfsohn and Tsiatis [4], Henderson *et al*. [5], Guo and Carlin [6], Tsiatis and Davidian [7] and Ibrahim *et al*. [8]). The main difference between Wulfsohn and Tsiatis [4] and Henderson *et al*. [5] is in terms of the linkage method. Former paper associated the survival and longitudinal model by means of a conditional score approach whilst the latter paper introduced a latent random-effects parameter to connect the two sub-models. Henderson *et al*. [5] and Guo and Carlin [6] are respectively differed from the estimation approach in such a way that Henderson *et al*. [5] proposed a two-stage model in fitting the joint model and Guo and Carlin [6] evaluated the parameters from the Bayesian approach. Tsiatis and Davidian [7] and Ibrahim *et al*. [8] are the review paper explained based on the PH-longitudinal model.

From the review study, limited learning is assessed on the joint PO-longitudinal model. As a result, the principle of this paper is to develop a unity model that form by the survival PO model and longitudinal model. The conjunction criteria between the two sub-models is therefore interplay by each other via the latent linkage parameters. The proposing joint model performances are followed to compare with the separated model.

The upper limitation is only valid for the joint model, however there is still a number of independent PO studies have been revised by the previous researcher, especially from the estimation approaches. The available estimation techniques are such as modified maximum likelihood approach [9,10,11], ranks based approach [12,13,14], likelihood sampler method [15], estimating equation approach [16], sieve maximum likelihood estimation [17] and Minorization-Maximization (MM) Algorithm [18].

On the other hand, Laird-Ware [19] proposed that the repeated measurements collected against time usually formularized in form of a random-effects model. The researchers hereby classified the random-effects model into a two-stage model in order to proceed with the estimation procedure. There is another form of classification provided by Goldstein [20], named as multilevel model. In the corresponding paper, the multilevel model hierarchically split the random-effects model into three simple linear regression models, which enclosed by one within-subjects model and two between-subjects models. The multilevel model is estimated by the iterative generalized least squares estimation that is proved to have similar performance with the maximum likelihood estimation (MLE).

In a real system, each longitudinal collections is sick of the missing datum. They are either restricted to the condition of missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR) [21]. The probability of MCAR missingness is independent on both of the observed and missing values against to the probability of MNAR missingness mechanisms, which relied heavily on the observed and missing data evaluation, while the probability of MAR only depends on the observed variable.

2.0 JOINT MODELING

This section discusses the mathematical modeling of the joint survival model and the longitudinal model. The longitudinal model is formularized by a random-effects model at the same time as the survival model represented by a proportional odds (PO) model.

2.1 Random-Effects Longitudinal Sub-Model

Let the repeated measurement for *i* subject of interests and *k* number of visited time-points represented as

$$y_{ik} = \mu_i(t_{ik}) + W_{1i}(t_{ik}) + \varepsilon_{ik} \,. \tag{1}$$

Model (1) has approximately the same symbol notation with the Laird-Ware random-effects model [19],

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i \,. \tag{2}$$

Comparing model structure (1) with model structure (2), the mean response and latent variable features are respectively written in term of $\mu_i(t_{ik}) = \mathbf{X}'_{1i}\mathbf{b}_1$ and $W_{1i}(t_{ik}) = \mathbf{Z}'_{1i}\mathbf{u}_i$ throughout the joint modeling process. $\mathbf{b}_1$ and $\mathbf{u}_i$ are the regression coefficients corresponding to $\mathbf{X}_{1i}$ explanatory variables and $\mathbf{Z}_{1i}$ matrix design. The error term whereas Normally distributed with mean zero and variance sigma square, $\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2_\varepsilon)$. The subject-specific random-effects (u_{0i}, u_{1i}) are then followed a zero-mean bivariate Gaussian distribution $\mathbf{u}_i \sim N(0, \Sigma)$ in which $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

2.2 Proportional Odds (PO) Survival Sub-Model

This subdivision is to explain the development of a joint modeling from survival approach. The basic survival odds model is usually given by

$$odds_i(t) = \frac{1 - S(t \mid x)}{S(t \mid x)} = \frac{1 - S_0(t)}{S_0(t)} \exp(x'b) \,, \tag{3}$$

where $S(t \mid x)$ is the conditional survivor function of an individual with covariate *x*, $\frac{1 - S_0(t)}{S_0(t)}$ is the baseline odds function given that $S_0(t)$ is the baseline survivor function, and $\exp(x'b)$ is a non-negative function with covariate *x* along with the corresponding regression coefficients *b*.

In constructing a joint model, $\exp(x'b)$ is essential to comprise of a subject-specific random-effects component, it is therefore re-formularized into $\exp[\mathbf{X}'_{2i}\mathbf{b}_2 + W_{2i}(t)]$. By definition, $\mathbf{X}'_{2i}\mathbf{b}_2$ may or may not have the element in common with $\mathbf{X}'_{1i}\mathbf{b}_1$ from longitudinal model. According to Henderson *et al.* [5], $W_i(t) = \{W_{1i}(t), W_{2i}(t)\}$ are the latent association which introduced to perform the stochastic dependence between the two sub-models. They are proportionally assuming in such a way that $W_{2i}(t) \propto W_{1i}(t)$. Therefore, the modify PO sub-model is defined as

$$odds_i(t) = \frac{1 - S(t \mid x)}{S(t \mid x)} = \frac{1 - S_0(t)}{S_0(t)} \exp[\mathbf{X}'_{2i}\mathbf{b}_2 + W_{2i}(t)]\,. \tag{4}$$

Taking the natural logarithm to the Eq. (4), a log-odds model is then displayed as

$$\log odds_i(t) = \log\left(\frac{1 - S_0(t)}{S_0(t)}\right) + \mathbf{X}'_{2i}\mathbf{b}_2 + W_{2i}(t)\,. \tag{5}$$

2.3 Simulation Procedure and Estimation

This study is concentrated on the computer simulation analysis. Therefore, a fully simulation procedures are specify in this paper. Goldstein's [20] multilevel model is suitable for simulating the repeated measurements. The simulation algorithm is as reported in the paper of Chin *et al.* [22] but with some correlation improvement between the two time-points [23]. Specify the full random effects longitudinal model with single covariate information as below:

$$y_{ik} = b_{10} + b_{11}t_{ik} + b_{12}x_i + b_{13}(x_i t_{ik}) + u_{0i} + u_{1i}t_{ik} + \varepsilon_{ik}\,. \tag{6}$$

Model (6) is split into a within-subjects model (Eq. 7) and two between-subjects models (Eq. 8) by the multilevel model. Along the simulation process stated in the unpublished paper, correlation between the predictors and dependent variables are additionally counted in the entire model (7) and (8).

$$y_{ik} = v_{0i} + v_{1i}t_{ik} + \varepsilon_{ik}\,, \tag{7}$$

$$v_{0i} = b_{10} + b_{12}x_i + u_{0i} \text{ and } v_{1i} = b_{11} + b_{13}x_i + u_{1i}\,. \tag{8}$$

Next, for the simulation study of the missing covariate *X*, adjusted logistic regression had been made from Mohamad [24] and Ho *et al*. [25]. The modified logistic regression in this study is therefore only assessed on the missing completely at random (MCAR) and missing at random (MAR) missingness mechanism. Given that

$$P(X \text{ is missing} \mid \text{data}) = \frac{\exp(a + bY + cX + dif)}{1 + \exp(a + bY + cX + dif)}\,, \tag{9}$$

where $a = \ln\left(\frac{q}{1-q}\right) - b\overline{Y} - c\overline{X}$ is a shifted logistic regression line which act as an intercept point to control the missing *X* values at the *q* proportions. Let $tu \sim Uniform(0,1)$. The consequent shifted $dif = \ln\left(\frac{tu}{1-tu}\right) - \frac{\exp(a + bY + cX)}{1 + \exp(a + bY + cX)}$ improves the logistic regression model to generate exact proportions of missing data. Set the letter $b = c = 0$ if the probability of model (9) lead to a MCAR model; in contrast, set $b = 1$, $c = 0$ to produce a MAR missing mechanism. The *q* percentage of the highest observations in model (9) is defined to be the missing values. The estimation of the separated longitudinal model is thus approximated using the complete-case analysis in which the entire subject *i* with missing values are forced to eliminate from the study.

Simulation avenue currently turns into the survival approach. The main observations is the events time which follow a log-logistic distribution influenced by the covariate effects. Given a set of random variable *T* can be generated from an inverse cumulative distribution function. The interested covariate log-logistic outcomes with the presence of linkage parameter is generated by

$$T = \alpha\left[\left(\frac{U}{1-U}\right)\exp[-(\mathbf{X}'_{2i}\mathbf{b}_2 + W_{2i}(T))]\right]^{1/\beta}. \tag{10}$$

In the computer simulation process, define $\mathbf{X}'_{2i}\mathbf{b}_2 = b_{21}x_i$ and $W_{2i}(T) = W_{2i} = \gamma_1 u_{0i} + \gamma_2 u_{1i}$. (γ_1, γ_2) hence executes as the latent association between the separated measurements and time-to-event developments.

The special properties of the proportional hazards (PH) model and proportional odds (PO) model at the same time allowed them to closed under the accelerated failure-time model. As a result, the corresponding log time is assumed to have a linear model [2]. Transform the simulated failure time, *T* from Eq. (10) into a modified log time, *Y*, the model is assumed to has

$$Y = \ln T = \mu + \boldsymbol{\delta}'\mathbf{X} + W_{2i}(Y) + \sigma\omega\,, \tag{11}$$

where ω has a standard logistic distribution in this study. Notice that the model (5) has a similar form as the model (11). Therefore, the parameters $(\mu, \sigma, \boldsymbol{\delta})$ are estimated through a maximum likelihood estimation method. As for the log-logistic distribution, the log-odds function (5) is defined as below:

$$\log odds_i(T) = \log(T/\alpha)^{\beta} + \mathbf{X}'_{2i}\mathbf{b}_2 + W_{2i}(T)\,. \tag{12}$$

The log-logistic parameters $(\alpha, \beta, \mathbf{b}_2, \boldsymbol{\gamma})$ are then related to $(\mu, \sigma, \boldsymbol{\delta}, \boldsymbol{\lambda})$ by taking

$$\begin{aligned} \hat{\beta} &= 1/\hat{\sigma}\,, \\ \hat{\alpha} &= [\exp(\hat{\mu}/\hat{\sigma})]^{1/\hat{\beta}}\,, \\ \hat{\mathbf{b}}_2 &= -\hat{\boldsymbol{\delta}}/\hat{\sigma}\,, \text{ and} \\ \boldsymbol{\gamma} &= -\hat{\boldsymbol{\lambda}}/\hat{\sigma}\,. \end{aligned} \tag{13}$$

3.0 ANALYSIS OF SIMULATION STUDIES

In this study, a series of pre-setting value is chosen arbitrarily for the entire estimators embedded in the survival model and longitudinal model. The true value of the parameters are assumed as follow: (1) time: [0,1,2], (2) covariate: $x \sim N(30, 4^2)$, (3) longitudinal element: $b_{10} = 2.5$, $b_{11} = -0.25$, $\rho_{xy} = 0.5$, $\sigma_{v0} = 1$, $\sigma_{v1} = 2$, $\rho = 0.5$, $\sigma_y = 1$, and (4) survival components: $\alpha = 1$, $\beta = 8$, $b_{21} = 0.25$, $\gamma_0 = -1.5$, $\gamma_1 = -2$. The remaining true values in the longitudinal model are evaluated by a mathematical relation such as

$$\begin{aligned} b_{12} &= \rho_{xy}\sigma_{v0}/\sigma_x\,, \\ b_{13} &= \rho_{xy}\sigma_{v1}/\sigma_x\,, \\ \sigma_1^2 &= \sigma_{v0}^2 - b_{12}^2\sigma_x^2\,, \\ \sigma_2^2 &= \sigma_{v1}^2 - b_{13}^2\sigma_x^2\,, \\ \sigma_\varepsilon^2 &= \sigma_y^2(1 - \rho^2)\,. \end{aligned}$$

100 samples are generated for each of the simulation iterations. Assuming θ as the true parameter, the estimator is then

evaluated by the $\bar{\hat{\theta}} = \sum_{i=1}^{M} \hat{\theta} \big/ M$ where M is the total number of simulations performed.

Table 1 Comparison of separated model and joint model with the complete observations, 15% and 50% X missingness for 100 simulations

Parameter	**True Value**	**Estimator (s.e.)**					
		Complete Observations		**MCAR**		**MAR**	
		Separated	**Joint**	**Separated**	**Joint**	**Separated**	**Joint**
				15% missing covariate *X*		**15% missing covariate *X***	
<u>*Longitudinal*</u>							
b_{10}	2.50	0.8351 (0.0299)	2.5108 (0.0699)	0.8288 (0.0331)	2.4969 (0.0680)	0.8791 (0.0302)	2.5656 (0.0654)
b_{11}	-0.25	1.1575 (0.1101)	-0.1763 (0.1307)	1.1560 (0.1140)	-0.1742 (0.1344)	1.2696 (0.1136)	-0.0623 (0.1296)
b_{12}	0.125	0.1663 (0.0026)	0.1250 (0.0023)	0.1663 (0.0028)	0.1253 (0.0022)	0.1584 (0.0027)	0.1201 (0.0021)
b_{13}	0.25	0.1033 (0.0017)	0.2486 (0.0042)	0.1032 (0.0017)	0.2482 (0.0044)	0.0998 (0.0017)	0.2407 (0.0042)
σ_1^2	0.75	1.4253 (0.0197)	0.8031 (0.0456)	1.4072 (0.0219)	0.8033 (0.0511)	1.2877 (0.0208)	0.7525 (0.0481)
σ_2^2	3	3.6666 (0.0406)	3.0406 (0.0848)	3.6531 (0.0524)	2.9931 (0.0965)	3.5988 (0.0470)	2.9052 (0.0913)
ρ	0.75	0.2317 (0.0094)	0.6679 (0.0055)	0.2270 (0.0103)	0.6644 (0.0057)	0.2090 (0.0099)	0.6551 (0.0058)
σ_ε^2	0.75	0.2581 (0.0034)	0.8697 (0.0217)	0.2584 (0.0036)	0.8857 (0.0252)	0.2579 (0.0037)	0.8772 (0.0256)
<u>*Survival*</u>							
α	1	1.1345 (0.0523)	0.9366 (0.0085)	1.2183 (0.0520)	0.9206 (0.0094)	1.2414 (0.0503)	0.9034 (0.0070)
β	8	2.9564 (0.0200)	8.1464 (0.0764)	2.9743 (0.0228)	8.2264 (0.0797)	3.0245 (0.0234)	8.2242 (0.0812)
b_{21}	0.25	0.0955 (0.0040)	0.2363 (0.0032)	0.1035 (0.0040)	0.2334 (0.0034)	0.1127 (0.0038)	0.2284 (0.0031)
γ_0	-1.5	-	-1.5664 (0.0335)	-	-1.5948 (0.0376)	-	-1.5938 (0.0370)
γ_1	-2	-	-2.0196 (0.0233)	-	-2.0397 (0.0249)	-	-2.0382 (0.0243)
				50% missing covariate *X*		**50% missing covariate *X***	
<u>*Longitudinal*</u>							
b_{10}	2.50			0.8329 (0.0428)	2.4838 (0.0802)	0.8597 (0.0363)	2.6502 (0.0973)
b_{11}	-0.25			1.1151 (0.1406)	-0.2170 (0.1559)	1.3171 (0.1531)	0.0512 (0.1778)
b_{12}	0.125			0.1665 (0.0034)	0.1253 (0.0025)	0.1503 (0.0036)	0.1113 (0.0035)
b_{13}	0.25			0.1034 (0.0021)	0.2490 (0.0051)	0.0949 (0.0024)	0.2291 (0.0061)
σ_1^2	0.75			1.3982 (0.0288)	0.8016 (0.0689)	1.2069 (0.0257)	0.7998 (0.0943)
σ_2^2	3			3.6693 (0.0659)	2.8971 (0.1229)	3.5542 (0.0700)	2.7971 (0.1340)
ρ	0.75			0.2126 (0.0123)	0.6676 (0.0067)	0.1756 (0.0145)	0.6421 (0.0084)
σ_ε^2	0.75			0.2565 (0.0049)	0.9840 (0.0394)	0.2552 (0.0049)	0.9899 (0.0554)
<u>*Survival*</u>							
α	1			1.1581 (0.0549)	0.9290 (0.0090)	1.2724 (0.0702)	0.8838 (0.0087)
β	8			2.9857 (0.0268)	8.4699 (0.1027)	3.1002 (0.0314)	8.4274 (0.1016)
b_{21}	0.25			0.0972 (0.0047)	0.2438 (0.0041)	0.1218 (0.0055)	0.2276 (0.0040)
γ_0	-1.5			-	-1.6883 (0.0528)	-	-1.6614 (0.0496)
γ_1	-2			-	-2.1025 (0.0332)	-	-2.0961 (0.0322)

Table 2 Comparison of separated model and joint model with the complete observations, 15% and 50% *X* missingness for 500 simulations

Parameter	True Value	Estimator (s.e.)					
		Complete Observations		MCAR		MAR	
		Separated	Joint	Separated	Joint	Separated	Joint
				15% missing covariate *X*		**15% missing covariate *X***	
Longitudinal							
b_{10}	2.50	0.8552 (0.0141)	2.5483 (0.0313)	0.8575 (0.0155)	2.5421 (0.0335)	0.9073 (0.0147)	2.6256 (0.0335)
b_{11}	-0.25	1.1793 (0.0523)	-0.1296 (0.0606)	1.1929 (0.0558)	-0.1160 (0.0641)	1.2974 (0.0556)	-0.0120 (0.0641)
b_{12}	0.125	0.1646 (0.0013)	0.1234 (0.0010)	0.1643 (0.0014)	0.1236 (0.0011)	0.1567 (0.0014)	0.1181 (0.0011)
b_{13}	0.25	0.1027 (0.0008)	0.2465 (0.0020)	0.1024 (0.0009)	0.2460 (0.0021)	0.0993 (0.0009)	0.2390 (0.0022)
σ_1^2	0.75	1.4038 (0.0092)	0.8147 (0.0194)	1.3962 (0.0100)	0.8090 (0.0203)	1.2900 (0.0097)	0.7807 (0.0206)
σ_2^2	3	3.6935 (0.0216)	2.9828 (0.0363)	3.6890 (0.0242)	2.9643 (0.0398)	3.6508 (0.0238)	2.9104 (0.0388)
ρ	0.75	0.2290 (0.0042)	0.6680 (0.0026)	0.2273 (0.0046)	0.6676 (0.0028)	0.2142 (0.0046)	0.6589 (0.0029)
σ_ε^2	0.75	0.2529 (0.0015)	0.8544 (0.0092)	0.2521 (0.0017)	0.8638 (0.0095)	0.2517 (0.0017)	0.8543 (0.0101)
Survival							
α	1	1.1785 (0.0249)	0.9264 (0.0047)	1.1829 (0.0248)	0.9264 (0.0047)	1.2058 (0.0253)	0.9008 (0.0034)
β	8	2.9644 (0.0095)	8.1772 (0.0314)	2.9753 (0.0107)	8.2290 (0.0334)	3.0139 (0.0107)	8.2154 (0.0333)
b_{21}	0.25	0.0996 (0.0019)	0.2336 (0.0016)	0.1000 (0.0020)	0.2351 (0.0017)	0.1080 (0.0019)	0.2276 (0.0014)
γ_0	-1.5	-	-1.5395 (0.0138)	-	-1.5564 (0.0148)	-	-1.5553 (0.0151)
γ_1	-2	-	-2.0368 (0.0102)	-	-2.0475 (0.0109)	-	-2.0446 (0.0109)
				50% missing covariate *X*		**50% missing covariate *X***	
Longitudinal							
b_{10}	2.50			0.8585 (0.0206)	2.5350 (0.0500)	0.8854 (0.0179)	2.5825 (0.0439)
b_{11}	-0.25			1.1651 (0.0727)	-0.1434 (0.0853)	1.2384 (0.0725)	-0.0524 (0.0840)
b_{12}	0.125			0.1647 (0.0018)	0.1239 (0.0016)	0.1507 (0.0018)	0.1136 (0.0015)
b_{13}	0.25			0.1028 (0.0011)	0.2469 (0.0028)	0.0965 (0.0011)	0.2329 (0.0028)
σ_1^2	0.75			1.3822 (0.0133)	0.9377 (0.0415)	1.1966 (0.0111)	0.8596 (0.0367)
σ_2^2	3			3.7133 (0.0321)	2.9166 (0.0497)	3.6272 (0.0316)	2.7847 (0.0507)
ρ	0.75			0.2189 (0.0059)	0.6637 (0.0047)	0.1894 (0.0061)	0.6452 (0.0049)
σ_ε^2	0.75			0.2533 (0.0022)	0.9960 (0.0452)	0.2502 (0.0022)	0.9485 (0.0205)
Survival							
α	1			1.2399 (0.0380)	0.9236 (0.0049)	1.2220 (0.0347)	0.8912 (0.0040)
β	8			3.0161 (0.0146)	8.3808 (0.0446)	3.1008 (0.0148)	8.4230 (0.0447)
b_{21}	0.25			0.0975 (0.0028)	0.2386 (0.0020)	0.1147 (0.0027)	0.2302 (0.0018)
γ_0	-1.5			-	-1.5744 (0.0213)	-	-1.5774 (0.0225)
γ_1	-2			-	-2.0881 (0.0142)	-	-2.0998 (0.0151)

Table 1 and Table 2 show the results of the simulation study. The estimators' performances are compared in term of accuracy and efficiency. Accuracy of an estimator is evaluated based on an unbiasedness whilst efficiency is determined by a standard error (s.e.) value. The bias is obtained significantly by the subtraction between the estimator value and the respective parameter true value, $\text{Bias}(\bar{\hat{\theta}}) = \bar{\hat{\theta}} - \theta$. The standard error on the other hand acquired by the division between the standard deviation and the simulation sample sizes, *M*, $SE(\hat{\theta}) = \sigma_\theta / M$.

The first objective in this study is to compare the performance of a joint model to a classical separated model of longitudinal and survival model. In general, a joint model achievement is much more superior to a separated model. Referring to Table 1 and Table 2, the standard errors are small enough to identify the goodness of fit of a model. The joint model estimators are concurrent relatively closed to the true values, indicated that the biases are comparatively small to it. A contradicted appearance occurred to the separated model, the bias of the estimators are very large. A supportive fact from Table 1 and Table 2, the performance of a $\bar{\hat{b}}_{11}$ value converges to a positive value rather than the negative true value. The performance of the joint model is therefore concluded to be better than a separated model in such a way that the survival and longitudinal processes in the joint model is influencing and improving each other via a latent association of (W_1, W_2).

Next, this study also aims to compare the effect of the missing covariate bring to the longitudinal process, survival process and joint modeling. The simulation study is examined for the complete observations (0% missing data *X*), missing completely at random (MCAR) and missing at random (MAR). The estimator's with complete observations always demonstrated the best performance to the analysis. All the estimated values are closed to the true values with small bias values except for the $\bar{\hat{b}}_{11}$ experienced an undesirable result in the simulations study. However, it does not affect the others parameters to converge.

This study also compare the percentage of missingness effect to the model. Although the estimators are converged approximately between the 15% and 50% missingness, the standard errors of the 50% missingness are higher than the 15% missingness.

Besides, the study on both of the 50% MCAR and 50% MAR cases in additional shows a better performance to the MCAR than MAR. The 100 simulations joint model outcome corresponds to a 50% MAR missingness mechanism seems to be unstable in such a way that the $\bar{\hat{b}}_{11}$ value has changed the direction of the slope from negative to positive. It is stable when it is undergone a 15% missingness or when the number of simulations is increased to 500 iterations. Eventually, as the number of simulation increases, the parameters estimators indicate improvement to the estimated values and/or the standard errors. The results are as displayed in Table 1 (100 simulations) and Table 2 (500 simulations).

■4.0 CONCLUSION

In conclusion, this study is performed better under the joint model rather than the separated model of survival and longitudinal model. The presence of a latent linkage association (W_1, W_2) in the joint model does improve the analysis by offering a good estimation for the joint model in this study. It is contrasting to the estimators obtaining from the separated model in such a way that the estimated parameters are seriously being underestimated or overestimated.

Furthermore, the performance of the simulation study is enhanced under the complete observations due to the deletion of 15% or 50% missing covariate X in either MCAR or MAR has critically reduced the total number of subjects being analyzed in the study. This is due to the complete-case analysis or the listwise deletion in which for all the subjects that carried with missing data are being removed from the analysis. If according to Allison [26], complete-case analysis only validated for the covariate's missingness that is independent of the dependent variable. Therefore, the method is fitting appropriately under the complete observations and MCAR missingness mechanism. As the degree of missing values increases, the accuracy of the estimators is further dropped.

Next, the increases number of simulations also influenced the accuracy of an estimator. The estimators approaches to the true values as the number of simulation study increases, it is as well as contribute a reduction to the standard error values. Therefore, the estimated parameters provide a better outcome to a large simulation studies. The same condition is applied to the model with a certain proportions of missing covariate.

Acknowledgement

This paper was supported in part by Fundamental Research Grant Scheme (FRGS) grant 4F004. The authors would also like to thank the Malaysian Ministry of Higher Education and Universiti Teknologi Malaysia for their financial funding through FRGS grant.

References

[1] D. R. Cox. 1972. Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*. 34(2): 187–220.

[2] J. P. Klein, M. L. Moeschberger. 2003. *Survival Analysis: Technique for Censored and Truncated Data*. 2nd ed. New York: Springer Science + Business Media, LLC.

[3] D. Collett. 2003. *Modelling Survival Data in Medical Research*. 2nd ed. United States of America: Chapman & Hall/CRC. 223–227.

[4] M. S. Wulfsohn, A. A. Tsiatis. 1997. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*. 53: 330–339.

[5] R. Henderson, P. Diggle, A. Dobson. 2000. Joint Modelling of Longitudinal Measurement and Event Time Data. *Biostatistics*. 1(4): 465–480.

[6] Guo, X. and B. P. Carlin. 2004. Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*. 58(1): 1–9.

[7] A. A. Tsiatis, M. Davidian. 2004. Joint Modeling of Longitudinal and Time-To-Event Data: An Overview. *Statistica Sinica*. 14: 809–834.

[8] J. G. Ibrahim, H. Chu, L. M. Chen. 2010. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *Journal of Clinical Oncology*. 28(16): 2796–2801.

[9] S. Bennett. 1983a. Log-logistic Regression Model for Survival Data. *Appl. Statist.* 32(2): 165–171.

[10] S. Bennett. 1983b. Analysis of Survival Data by the Proportional Odds Model. *Statistics in Medicine*. 2: 273–277.

[11] S. A.Murphy, A.J. Rossini, A. W. Van Der Vaart. 1997. Maximum Likelihood Estimation in the Proportional Odds Model. *Journal of the American Statistical Association*. 92(439): 968–976.

[12] A. N. Pettitt. 1982. Inference for the Linear Model Using a Likelihood Based On Ranks. *J. R. Statist. Soc.* B. 44(2): 234–243.

[13] A. N. Pettitt. 1983. Approximate Methods Using Ranks for Regression With Censored Data. *Biometrika*. 70(1): 121–132.

[14] A. N. Pettitt. 1984. Proportional Odds Models for Survival Data and Estimates Using Ranks. *Appl. Statist.* 32(2): 169–175.

[15] D. B. Dabrowska, K. A. Doksum. 1988. Estimation and Testing in a Two-Sample Generalized Odds-Rate Model. *Journal of the American Statistical Association*. 83(403): 744–749.

[16] S. C. Cheng, L. J. Wei, Z. Ying. 1995. Analysis of Transformation Models with Censored Data. *Biometrika*. 82(4): 835–845.

[17] X. Shen. 1998. Proportional Odds Regression and Sieve Maximum Likelihood Estimation. *Biometrika*. 85(1): 165–177.

[18] D. R. Hunter, K. Lange. 2002. Computing Estimates in the Proportional Odds Model. *Ann. Inst. Statist. Math.* 54(1): 155–168.

[19] N. M. Laird, J. H. Ware. 1982. Random-Effects for Longitudinal Data. *Biometrics*. 38: 963–974.

[20] H. Goldstein. 1986. Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Square. *Biometrika*. 73: 43–56.

[21] P. Diggle, M. G. Kenward. 1994. Informative Dropout in Longitudinal Data Analysis (with discussion). *Applied Statistics*. 43: 49–94.

[22] W. Y. Chin, Z. M. Khalid, M. K., Ho. 2011. Analysis of Repeated Measures via Simulation. *Simposium Kebangsaan Sains Matematik (SKSM 19)*. 9–11 November (unpublished).

[23] P. J. Diggle. 1988. An Approach to the Analysis of Repeated Measurements. *Biometrics*. 44(4): 959–971.

[24] I. Mohamad. 2003. Data Analysis in the Presence of Missing Data. PhD Thesis: Lanchester University.

[25] M. K. Ho, F. Yusof, I. Mohamad. 2012. Imputation of Missing Data with Different Mechanism. *Jurnal Teknologi*. 57 (Sciences & Engineering) Suppl 1: 57–67.

[26] P. D. Allison. *The SAGE Handbook of Quantitative Methods in Psychology*. 2009. London: SAGE Publications Ltd. Chapter 4, *Missing Data*. 72–89.