Pak. J. Statist.
2018 Vol. 34(6), 459-478

LONG-TERM SURVIVOR MODEL WITH ALTERNATIVE LINKS IN THE CURE

Diego I. Gallardo[1§], **Heleno Bolfarine**[2] and **Antonio Carlos Pedroso-de-Lima**[2]

[1] Departamento de Matemáticas, Facultad de Ingenieria
Universidad de Atacama, Copiapó, Chile.
[2] Instituto de Matemática e Estatística, Universidade de São Paulo
São Paulo, Brasil. Email: hbolfar@ime.usp.br; acarlos@ime.usp.br
[§] Corresponding author Email: diego.gallardo@uda.cl

ABSTRACT

Most studies involving cure fraction models in survival analysis relate the probability of cure and possible prognostic factors considering a logistic link function. As is well known, the logistic density function is symmetric, preventing the possibility of asymmetry in the relationship between the cure probability and the covariates. The main purpose of this paper is to predict the chance (or probability) of cure for future patients, considering other links between the covariates with the cure fraction. Some links extensively considered in the literature for modeling the cure probabilities are the probit and the complementary loglog links (c-loglog). We evaluate the use of asymmetric links when the main goal is the estimation of survival probabilities. In addition, the patients are considered to be grouped in clinics and, for a given clinic, there may be a non-null association among the survival times, thus affecting the survival probabilities. An EM-type algorithm conjugated with REML estimation procedure in the M-step is considered for parameter estimation. Model comparisons are implemented using the cAIC and cBIC scores. A simulation study suggests that the proposed estimation procedure behaves well for moderate and large sample sizes. A real data illustration reveals also the good performance of the proposed model and estimation approach.

KEYWORDS

Mixture model, bivariate random effects, REML, cure rate.

1. INTRODUCTION

In recent decades, medical treatments have helped to improve the life expectancy of patients with many types of cancer. In some cases, patients submitted to similar medical procedures and sharing certain genetic as well as environmental characteristics may live longer than originally expected. Statistical models to deal with data sets generated by these scenarios have been developed by several authors, such as Chen et al. (1999b) and Berkson and Gage (1952). More recently, there have been a number of generalizations proposed in the literature (see, e.g., Rodrigues et al. (2009a, 2009b, 2012), Cancho et al. (2013), Barreto-Souza (2015) and Gallardo et al. (2016, 2018)). In this paper we consider the approach motivated by Berkson and Gage (1952), known as the mixture model approach, that assumes the existence of two subgroups of patients in the population, namely cured and uncured or susceptible patients. We also assume a parametric framework, where a family of distributions is assigned to the survival times related to the uncured individuals. The fraction of cured subjects is allowed to change according to a set of potential covariates through a conveniently chosen link function. In the context of mixture models, it is usual to consider a logistic link function for the cure fraction and an exponential model for the survival times of uncured patients, the so-called logistic-exponential model; however, other options have been described as the logistic-Weibull model (Farewell, 1982) and, in a semi parametric approach, Fang et al. (2005) and Lu (2008) assume a logistic link function for the probability of cure.

The extensive use of the logistic distribution (in a cure rate model and other contexts) can be explained by the ease of computational implementation and interpretation. For discussion about this topic, we suggest Chen et al. (1999a) and Bazán et al. (2006, 2010, 2014). However, it also imposes an important limitation, that of being symmetric about the origin. Consequently, if the objective of the study is to predict the survival time or the probability of cure for future patients, such approach may not be appropriate, particularly when trying to predict patients with unusual values for some covariates. Other link functions extensively studied in the literature in this situation are the probit link (normal cumulative distribution function) and the complementary loglog link (Gumbel cumulative distribution function). The first is symmetrical and the latter, asymmetrical. In addition to studying link functions other than the logistic, in this work we also consider that, due to the adopted experimental design, the sample units are grouped in clusters, so that for those subjects from a given cluster, non-observable characteristics may interfere with the outcomes. We consider, then, random effects models to take into account possible similarities of survival experience for those units. In this context, Yau and Ng (2001) proposed a mixture model incorporating random effects on the cure rate (that we will denote by V_j) and on the survival function of susceptible individuals (U_j) assuming that, for a given cluster j, U_j and V_j are independent and normally distributed; their work is based on the logit link for the cure probability. As the two effects are related to the same cluster, it is conceivable include some sort of dependency; Lai and Yau (2008) assumed that (U_j,V_j) followed a bivariate normal distribution, again under a logit link for the cure probability. In the same context, Lai and Yau (2010) considered a power family of functions to link the covariates related to the survival of susceptible individuals. Although we will base our development in the mixture model, it is worth mentioning that, for the cure model of Chen et al. (1999b), Lopes and Bolfarine (2012) incorporated random effects in the cure probability, whereas Gallardo et al. (2013) generalized a bit more, including random effects in both, the cure probability and the survival function of susceptible individuals, although assuming independence between the random effects.

In this paper, we propose a unified family of link functions, including the logit, probit and complementary loglog functions for the random effects mixture model, where the random effects follow a bivariate normal distribution. In Section 2 we present the ordinary mixture cure rate model. In Section 3 we extend the model of Section 2, incorporating random effects. In Section 4 we develop the classical procedure for parameter estimation, combining the EM algorithm and an approach based on Best Linear Unbiased Predictor (BLUP) and Restricted Maximum Likelihood (REML). In Section 5, the model is considered in a real life database. In Section 6 we present a simulation study to evaluate the performance of the estimation procedure presented in this paper. In Section 7, we conclude that the proposed estimation approach is satisfactory, being a viable alternative for real data when it seems adequate the use of an asymmetrical link for the probability of cure.

2. THE MIXTURE MODEL

Following Berkson and Gage (1952), we assume the target population contains two types of units, susceptible to the event of interest and non-susceptible or cured. It is not known whether a given unit is cured or not; so that we consider a non-observable (latent) random variable Y assuming value 1 if the unit is susceptible and 0 otherwise. For non-cured units, we denote $S(t|\lambda)$ as the survival function related to the time up to the event of interest is observed, where λ is a set of unknown parameters. If $P(Y = 1) = \pi$, the survival and hazard functions of the entire population are given, respectively, by

$$S_p(t|\lambda,\pi) = 1 - \pi + \pi S\ (t|\lambda) \text{ and } h_p(t|\lambda,\pi) = \frac{\pi f\ (t|\lambda)}{1-\pi+\pi S\ (t|\lambda)},$$

where $f(t|\lambda) = -\frac{dS(t|\lambda)}{dt}, t \geq 0$. Note this model does not belong to the family of proportional hazards models. Let T_i^* be the time to failure for the i-th experimental unit and C_i the corresponding censoring time. Considering right censoring, for each unit i, $i = 1, \dots, n$, the actually observed time is represented by random variables $T_i = \min\ (T_i^*, C_i)$ and $\delta_i = I(T_i^* \leq C_i)$. The available (observed) data will be denoted by $D_{obs} = (t,\delta)$, where $t = (t_1, \dots, t_n)$ and $\delta = (\delta_1, \dots, \delta_n)$. We also define $Y = (Y_1, \dots, Y_n)$ so that the complete data will be denoted by $D = (t,\delta,Y)$. It is easy to see that the complete likelihood function for (λ,π) is given by

$$L(\lambda,\pi|D) = \prod_{i \in \Delta_0} (1-\pi)^{1-Y_i} \times \prod_{i \in \Delta_1} \pi h(t_i|\lambda)^{\delta_i} S\ (t_i|\lambda),$$

where $h(t|\lambda) = f(t|\lambda)/S(t|\lambda)$ and Δ_0 and Δ_1 denotes the subset of indices for cured and susceptible individuals, respectively. Following Lu (2010), we have that

$$E(Y_i|D_{obs}) = \delta_i + (1-\delta_i)\frac{\pi S(t_i|\lambda)}{1-\pi+\pi S(t_i|\lambda)}, i = 1, \dots, n, \tag{1}$$

In order to allow the inclusion of covariates $x_i = (x_{i1}, \dots, x_{ir})^T$ among the susceptible items, assuming a proportional hazards model for T_i^*,

$$S(t|\xi_i,\lambda) = [S_0\ (t|\lambda)]^{exp\{\xi_i\}}, i = 1, \dots, n, \tag{2}$$

for $\xi_i = x_i^T\gamma$, where $S_0(t|\lambda)$ is known as the baseline survival function, γ is a r-vector of unknown parameters. Also, we consider covariates $z_i = (z_{i1}, \dots, z_{is})^T$ related to the cure fraction, i.e.,

$$\pi_i = \psi(z_i^T\beta), i = 1, \dots, n,$$

where β is a s-vector of (also unknown) parameters and $\psi\ (\cdot)$ is any of the link functions in the Table 1. As discussed in Gallardo et al. (2013), the vector γ cannot contain an intercept in order to avoid identifiability problems.

Table 1
Link Functions for the Cure Probability in the Mixture Model.
$\Phi(\cdot)$ is the Cumulative Distribution Function for the Standard Normal Model

Link	Logit	Probit	C-loglog
$\psi\,(u)$	$e^u/\,(1+e^u)$	$\Phi\,(u)$	$\exp\,(-\exp\,(u))$

3. THE MIXTURE MODEL WITH RANDOM EFFECTS

As discussed in Section 2, we assume that experimental units are grouped in clusters, so that dependence among lifetimes from a same cluster may arise. Assume there are J clusters with n_j units in cluster j, $j = 1, \dots, J$. Therefore, Y_{ij} is the unobserved binary random variable assuming value 1 for the i-th unit from cluster j is susceptible to the event of interest and 0 otherwise. For $j = 1, \dots, J$, let $Y_j = (Y_{1j}, Y_{2j}, \dots, Y_{n_j j})^T$. Also, let T_{ij}^* be the time to event for the i-th experimental unit in the j-th cluster and C_{ij} the corresponding censoring time, so that $T_{ij} = \min\,(T_{ij}^*, C_{ij})$ and $\delta_{ij} = I\big(T_{ij}^* \le C_{ij}\big)$. We consider the situation where $S\,(\cdot\,|\lambda)$ is as in (2). In addition, we also include a second index for the covariates, so that, for each unit we have information on x_{ij} and z_{ij}. To account for the effect of the j-th cluster, let $(U_j,\, V_j)$, $j = 1, \dots, J$, be random variables related to the survival time of susceptible units and to the cure fraction, so that

$$\xi_{ij} = x_{ij}^T\gamma + U_j \text{ and } \pi_{ij} = \psi\big(z_{ij}^T\beta + V_j\big),\, i = 1, \dots, n_j; j = 1, \dots, J. \tag{3}$$

The random variables U_{j}, $j = 1, \dots, J$ are usually called frailties. As in Lai and Yau [18], we assume that $\big(U_j, V_j\big) \sim N_2\,(0, \Sigma)$, with

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}, \tag{4}$$

where ρ is the correlation between U_j and V_j. Moreover, let $U = (U_1, \dots, U_J)$, $V = (V_1, \dots, V_J)$ and $b = (U, V)$, so that $b \sim N\,(0, \Omega)$, where $\Omega = \Sigma \otimes I_J$ and I_J is the identity matrix of dimension J. High values of U_j suggest that susceptible units in j-th cluster are at greater risk of suffering the event of interest; low values of V_j suggest units in$j$-th cluster have greater probability of being cured. As mentioned above, it is not possible to include an intercept in the model, so that the random effects U_j, $j = 1, \dots, J$ take this role. The complete BLUP log-likelihood function for the model is given by $\ell = \ell_1 + \ell_2$, where

$$\ell_1 = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\Big\{Y_{ij}\big[log\,\pi_{ij} + \xi_{ij}\,log\,S_0\big(t_{ij}\big|\lambda\big) + \delta_{ij}\big(log\,h_0\big(t_{ij}\big|\lambda\big) + \xi_{ij}\big)\big] \\ + \big(1 - Y_{ij}\big)\,log\,\,\big(1 - \pi_{ij}\big)\Big\} \tag{5}$$

and

$$\ell_2 = -\frac{1}{2(1-\rho^2)}\left(\frac{U^TU}{\sigma_u^2} + \frac{V^TV}{\sigma_v^2} - \frac{2\rho U^TV}{\sigma_u\sigma_v}\right) - \left(\frac{J}{2}\right) log\{4\pi^2\sigma_u^2\sigma_v^2\,(1-\rho^2)\}. \tag{6}$$

See Yau and Ng (2001) and Lai and Yau (2008) for more details about the construction of ℓ_1 and ℓ_2. In this work, we consider two scenarios for the baseline survival $S_0\,(\cdot\,|\lambda)$: a Weibull model, where

$$S_0(t|\lambda) = exp(-\lambda t^{\alpha}), t, \lambda, \alpha > 0, \tag{7}$$

and a piecewise exponential (PE) model (Friedman, 1987)

$$S_0(t|\lambda) = exp\left\{-\sum_{l=1}^{L} \lambda_l \nabla_l\,(t)\right\} t, \lambda_1, \dots, \lambda_L > 0, \tag{8}$$

where

$$\nabla_l(t) = \begin{cases} 0, if\ t < a_{l-1} \\ t - a_{l-1}, if\ a_{l-1} \le t < a_l,\ l = 1, \dots, L, \\ a_l - a_{l-1}, if\ t \ge a_l \end{cases} \tag{9}$$

with $a = (a_1, \dots, a_{L-1})$ a vector of known constants, $a_0 = 0$ and $a_L = 1$. Such models have the proportional risks property, which is reasonable in many biological contexts. For a more detailed discussion about the topic, see Chen et al. [1]. For the Weibull model, $\lambda = (\alpha, \lambda)$, and depending on values of α, the risk function may be monotonic increasing, decreasing or constant, whereas for the PE model, $\lambda = (\lambda_1, \dots, \lambda_L)$, and non-monotonic risk functions are accommodated as well. As a particular case, when $\rho = 0$ and $S_0\,(\cdot\,|\boldsymbol{\lambda})$ is the survival function in (7), we have the model discussed by Yau and Ng (2001). Also, unlike Lai and Yau (2008), we assume a parametric approach, which allows model comparison based on measures of relative goodness of fit such as cAIC and cBIC, conditional versions of AIC and BIC criteria, respectively, with a penalization that depends on the predicted random effects and their respective variances. For more details, see Donohue et al. (2011).

4. ESTIMATION

The estimation of parameters and prediction of random effects for the mixture model discussed in the previous section relies on expressions (5) and (6). At first, following Yau and Ng (2001), the baseline survival function $S_0\,(\cdot\,|\boldsymbol{\lambda})$ is considered as known but its estimation will be discussed later in this section.

In order to deal with the latent variables $Y_{ij},\ i = 1, \dots, n_j,\ j = 1, \dots, J,$ we have implemented an EM algorithm in the following way. This procedure has drawn attention in recent years in the context of cure rate models. For instance, Pal and Balakrishnan (2016, 2017), Balakrishnan and Pal (2016), Gallardo et al. (2017a, 2017b). In the E-step, with fixed values from a previous iteration (say, $k-1$) for $\varsigma = (\beta, \gamma, b)$ and $\tau = (\sigma_u^2, \sigma_v^2, \rho)$, using (1) we have

$$\begin{aligned} \tilde{y}_{ij} &= E\,(Y_{ij}|D_{obs}, \zeta^{(k-1)}, \tau^{(k-1)}, \lambda^{(k-1)}) \\ &= \delta_{ij} + (1 - \delta_{ij}) \frac{\pi_{ij}^{(k-1)} S_0\big(t_{ij}|\lambda^{(k-1)}\big)^{exp\,\xi_{ij}^{(k-1)}}}{1-\pi_{ij}^{(k-1)} + \pi_{ij}^{(k-1)} S_0\big(t_{ij}|\lambda^{(k-1)}\big)^{exp\,\xi_{ij}^{(k-1)}}}, k = 1,2, \dots,. \end{aligned} \tag{10}$$

Let $\tilde{y}^{(k)} = (\tilde{y}_{11}^{(k)}, \ldots, \tilde{y}_{Jn_J}^{(k)})$ be the resulting values in (10). In the M-step, we update the parameters and the random effects in $\varsigma^{(k)}$ considering the restricted maximum likelihood (REML) approach proposed by McGilchrist and Yau (1995). At this point, the value of $\tau^{(k)}$ is also taken as fixed, so that we may find the BLUE and BLUP of the elements in $\varsigma^{(k)}$ (Henderson, 1975) by the Newton-Raphson algorithm, that is,

$$\varsigma_{m+1}^{(k)} = \varsigma_m^{(k)} + \left\{B_m^{(k)}\right\}^{-1}\left(\frac{\partial \ell}{\partial \varsigma}\right)\Big|_{\varsigma=\varsigma_m^{(k)}}, m = 0,1,2, \ldots, \tag{11}$$

where $\varsigma_0^{(k)}$ is an initial value at step k. The matrix B is given by the negative of the second derivative of ℓ with respect to ς. Algebraic details are provided in Appendix A. Let $(\hat{\gamma}^{(k)}, \hat{\beta}^{(k)})$ and $\hat{b}^{(k)}$ be the BLUE and BLUP for (γ, β) and b at the k-th iteration, respectively. The matrix B evaluated at $\hat{\varsigma}^{(k)}$ and its inverse, according to the partition $\gamma|\beta|b$, are

$$\hat{B}^{(k)} = \begin{bmatrix} \hat{B}_{11}^{(k)} & \hat{B}_{12}^{(k)} & \hat{B}_{13}^{(k)} \\ \hat{B}_{21}^{(k)} & \hat{B}_{22}^{(k)} & \hat{B}_{23}^{(k)} \\ \hat{B}_{31}^{(k)} & \hat{B}_{32}^{(k)} & \hat{B}_{33}^{(k)} \end{bmatrix} \text{ and } \left\{\hat{B}^{(k)}\right\}^{-1} = \begin{bmatrix} \hat{A}_{11}^{(k)} & \hat{A}_{12}^{(k)} & \hat{A}_{13}^{(k)} \\ \hat{A}_{21}^{(k)} & \hat{A}_{22}^{(k)} & \hat{A}_{23}^{(k)} \\ \hat{A}_{31}^{(k)} & \hat{A}_{32}^{(k)} & \hat{A}_{33}^{(k)} \end{bmatrix}.$$

Defining the matrices

$$Q_1 = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}, Q_2 = \begin{pmatrix} 0 & I_J \\ I_J & 0 \end{pmatrix} \text{ and } Q_3 = \begin{pmatrix} 0 & 0 \\ 0 & I_J \end{pmatrix}. \tag{12}$$

We next update the value of $\tau^{(k)}$. As such, let $\hat{R}_1^{(k)} = tr\left(\hat{A}_{33}^{(k)} + \hat{b}^{(k)}\hat{b}^{(k)T}\right)Q_1$, $\hat{R}_2^{(k)} = tr\left(\hat{A}_{33}^{(k)} + \hat{b}^{(k)}\hat{b}^{(k)T}\right)Q_2/2$ and $\hat{R}_3^{(k)} = tr\left(\hat{A}_{33}^{(k)} + \hat{b}^{(k)}\hat{b}^{(k)T}\right)Q_3$. It follows that (see Lai and Yau, 2008) the REML estimators of $\tau^{(k)}$ at the k-th iteration may be expressed as

$$\{\hat{\sigma}_u^2\}^{(k)} = \hat{R}_1^{(k)}/J, \{\hat{\sigma}_v^2\}^{(k)} = \hat{R}_3^{(k)}/J \text{ and } \hat{\rho}^{(k)} = \hat{R}_2^{(k)}\Big/\sqrt{\hat{R}_1^{(k)}\hat{R}_3^{(k)}}. \tag{13}$$

The variances of these estimators are computed in Appendix B.

The values of $S_0\,(t|\lambda^{(k)})$ are updated considering the REML estimators for ς and τ obtained at the k-th iteration; a profile log-likelihood function (ℓ_p) is computed for λ at the k-th iteration assuming either a Weibull or a PE distribution for the baseline survival function. If $\lambda_0^{(k)}$ is an initial value, then the Newton-Raphson algorithm will result

$$\lambda_{m+1}^{(k)} = \lambda_m^{(k)} + \left\{C_m^{(k)}\right\}^{-1}\left(\frac{\partial \ell_p}{\partial \lambda}\right)\Big|_{\lambda=\lambda_m^{(k)}}, m = 0,1,2, \ldots, \tag{14}$$

where C is minus the second derivative of the profile log-likelihood function ℓ_p with respect to λ. Details are given in Appendix C. The standard error for the estimator of λ can be estimated by Jackknife (Miller, 1974).

The whole process to get estimates and predicted values for ς, τ and λ is based on iterative procedures based on equations (10)-(14). Initial values for fixed quantities are obtained through the fit of a linear model without random effects; the random effects U and V are initially taken as zero and, for the corresponding variances and correlation coefficient ρ, initial values are arbitrarily set (e.g., 0.5 and zero, respectively).

5. APPLICATION

Grande et al. (2000) conducted a follow-up study at the Dentistry School of the University of São Paulo, Brazil, to compare retention efficacy of two products: Delton (a standard sealant) and Optibond (a dual cure glass filled dental adhesive). The study involved 37 children with ages ranging from 11 to 17 years. Each individual received both sealants, randomly assigned to each side of the mouth. In total, there are 171 teeth receiving the sealants, related to the 37 children. All subjects were instructed with respect to oral hygiene during the trial period (30 months). Periodically, trained researchers evaluated the sealants condition and assigned scores 1 (perfect sealed), 2 (partial loss) or 3 (total loss). The response variable was defined as the time (in months) up to partial or total sealant loss.

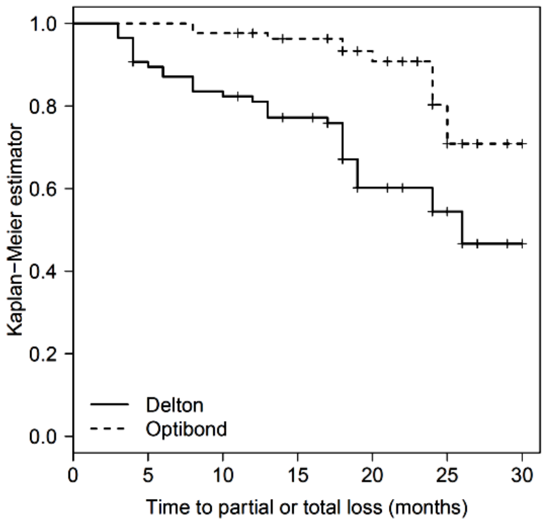

Figure 1: Kaplan-Meier Estimates by Sealant

Figure 1 shows Kaplan-Meier estimates for each sealant material, suggesting a higher retention for Optibond when compared to Delton. For this problem, it is appropriated to consider a survival model with cure rate since it is possible that the sealant remains in good conditions until adulthood. Here, we term as cure a tooth i for which the sealant remains in perfect conditions throughout the life of an individual j, with probability π_{ij}. The clustering, related to measurements made on the same individual, suggests the need to consider a mixed model. Therefore, we applied the models discussed in Section 3 based on Weibull as well as PE distributions with $L = 2$ and $L = 3$. We also adopted the link functions described in Table 1. In addition, for $j = 1, \dots, 37$ and $i = 1, \dots, n_j$, the following quantities were defined:

- $z_{ij} = (1, Delton_{ij}, Age^{*}_{ij})^T$. For the i-th tooth in the j-th individual, $Delton_{ij}$ indicates whether Delton ($= 1$) or Optibond ($= 0$) was applied, and Age^{*}_{ij} denotes the age centered at the its median value.
- $x_{ij} = Delton_{ij}$, with the same codification as before.
- $\beta = (\beta_{Intercept}, \beta_{Delton}, \beta_{Age^{*}})^T$, denoting the intercept and the effects of material and age on the cure probability.
- γ, denoting the effect of Delton (in comparison with Optibond) on the retention time for susceptible subjects.

Note that the age of subjects is considered only when modeling the cure probability.

Finally, we also consider the cases for bivariate ($\rho \neq 0$) random effects. Table 2 presents the values of cAIC and cBIC for the fitted models. Both criteria suggest that the best model is the mixture model based on the PE distribution with $L = 2$ for the non-cured individuals and c-loglog (Complementary) link function for the cure probability.

Table 2
cAIC and cBIC for Mixture Models with Bivariate Random Effects and Selected Link Functions

Model for S_0 $(\cdot\,|\boldsymbol{\lambda})$	**Criterion**	**Link**		
		Logit	**Probit**	**C-loglog**
Weibull	cAIC	615.47	616.80	619.40
	cBIC	762.41	763.86	765.61
PE ($L = 2$)	cAIC	570.24	567.70	**564.76**
	cBIC	717.93	715.54	**712.09**
PE ($L = 3$)	cAIC	568.92	568.37	565.61
	cBIC	716.44	716.04	712.94

In addition, Figure 2 depicts point estimates and confidence intervals based on the three link functions and the PE model for S_0 $(\cdot\,|\boldsymbol{\lambda})$. The random effect associated to π_{ij} is fixed at zero and the sealant is Optibond. Note that logit and probit links provide somewhat closer results when compared to the complementary log-log link. The confidence intervals suggest similar conclusions: around the median age, the interval for the c-loglog link has no intersection with the intervals for the other two links. Similar patterns were observed for different values of the random effects V_j's and Delton sealant.

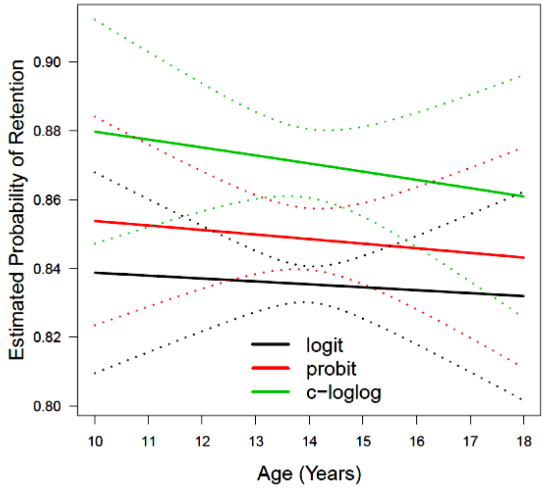

Figure 2: Estimated Probability of Retention and Confidence Intervals for Different Ages and Optibond Sealant. Solid lines are Point Estimates. Dashed lines are 95% confidence intervals.

Table 3 shows estimates for parameters in the selected model. We can see that type of sealant is an important predictor for both, survival of uncured individuals and probability of retention. The estimated variances for the random effects in the cure fraction are much larger than the variances for the frailties.

Table 3
REML Estimates for the Mixture Model with Bivariate Random Effects, PE Distribution (L = 2) for Non-Cured Individuals and Link c-loglog for Cured Individuals

Parameter	Estimates	s.e.
γ_{Delton}	2.77002	0.25164
$\beta_{Intercept}$	0.62461	0.14432
β_{Delton}	-0.80348	0.20413
β_{Age^*}	-0.00893	0.07102
σ_u^2	0.87742	0.07739
σ_v^2	0.06904	0.13149
ρ	0.99630	0.17341
λ_1	0.00168	0.00031
λ_2	0.04991	0.29198

The estimated hazard ratio for uncured individuals is $e^{\hat{\gamma}} = 15.96$ in favor of Optibond, that is, for a given susceptible subject, a partial or total loss of material in a tooth sealed with Delton has a risk approximately 15 times greater than for a similar tooth receiving Optibond. For two individuals of same age, given the negative value of β_{Delton}, the probability of non-susceptibility to sealant loss (cure) is smaller for Delton when compared to Optibond.

6. SIMULATION STUDY

In this section, we present results of two simulation studies conducted to assess the performance of the proposed methodology. The first one considers the model of Section 3 and the complementary log-log link function for the cure probability. Parameter recovery based on the estimation procedure of Section 4 is evaluated. The second study focuses on the evaluation of the cure rate estimation for different scenarios produced by the three link functions discussed earlier.

6.1 Parameter Recovery

We consider similar conditions as described in Yau and Ng (2001) and Lai and Yau (2008) but with the complementary log-log link (asymmetric) function instead of the (symmetric) logit link considered by them in the cure probability. The computational routines were developed in R, version 2.15.1 (2015).

Data was simulated considering $J = 10$ clinics with 15 patients each: 10 of them were randomly assigned to the treatment group ($x_{ij} = 1$) and the remaining 5 to the control group ($x_{ij} = 0$). Cure probability was based on covariates $z_{ij} = (1, x_{ij})$. Random effects (U_j, V_j) were simulated from a $N(0, \Sigma)$ distribution, with Σ as in (4). The values of Y_{ij} (susceptibility indicator) were simulated from a Bernoulli distribution, with mean π_{ij}. If $Y_{ij} = 0$, failure time was taken as 1; otherwise, it was simulated from the density $\mathrm{f_0}(\mathrm{t}|\gamma, \lambda) = \mathrm{h_0}(\mathrm{t}|\lambda) \exp\, (x_{ij}\gamma + U_j)\mathrm{S_0}(\mathrm{t}|\lambda)^{\exp\,(x_{ij}\gamma + U_j)}$. Failure times greater than the censoring time C (a fixed quantity) were considered as censored at C. We evaluated two cases: $\mathrm{S_0}(\mathrm{t}|\lambda)$ based on a Weibull distribution as in (7), with parameters $\lambda = 0.01$ and $\alpha = 1$ and $\mathrm{S_0}(\mathrm{t}|\lambda)$ related to a PE distribution with $L = 2$, as in (8), with parameters $\lambda_1 = 0.01$, $\lambda_2 = 0.02$ and constant $a = 150$. In both cases, the vector λ was taken as known. Such specifications imply a mean for the time-to-event of 100 and 88.7 for Weibull and PE models, respectively. The regression coefficients were fixed as $\gamma = -0.5$, $\beta_0 = 0.5$ and $\beta_1 = -0.5$. Such values imply a reduction of the risks in $e^{-0.5} \approx 0.6$ for individuals with covariate $x = 1$ compared with individuals with $x = 0$. Additionally, the specified cure rate is given by $e^{-1} \approx 0.37$ and $\exp\, (-e^{-0.5}) \approx 0.19$ for individuals with covariate $x = 0$ and $x = 1$, respectively. In addition, for each group we considered $\{\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, C = 1000\}$ and ρ was taken in the set $\{0.2, 0.4, 0.6, 0.8\}$ in order to allow us the assessment of the estimation procedure for different intensities of correlation. A total of 500 simulations was taken in each case.

It is to be expected that an increase in ρ should improve the performance of the estimators in general, since the inclusion of this parameter in the model is justified by the dependence of the random effects in a given clinic. Simulation results are presented in Tables 5 and 6 in Appendix D. As in Yau and Ng [17] we also consider SE_1 and SE_2, the mean of the standard errors and the standard deviations of the estimates for the 500 simulated samples, respectively. Comparison between these two quantities provides an idea whether there is over or underestimation of the standard errors of our estimators.

With respect to the estimation of γ, the bias seems to decrease as ρ increases and the variance of the corresponding estimator is slightly overestimated. For $\hat{\beta}_0$ and $\hat{\beta}_1$ the biases turned out to be very small for all considered cases, but the variance of both are also overestimated by a small quantity. Moreover, $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ have an acceptable bias, that is improved when the correlation ρ is larger. The variances of these estimators are slightly overestimated, except in Case 4, for the PE model, when it seems to be underestimated. Finally, the performance of $\hat{\rho}$ is quite reasonable when compared to the average bias, although the variance of the corresponding estimator seems to be more overestimated than the ones related to the other estimators; however, it seems to occur some improvement as ρ increases.

Based on both simulation studies, we may conclude that, at least in the cases considered here, the performance of the proposed estimators is acceptable in relation to bias and the corresponding estimated variances. The results show improvement as the correlation ρ becomes larger.

6.2 Assessing the Effect on the Cure Rate Estimation

We also simulated the model in Section 3 based on the logit, probit and complementary loglog link functions; for each case, we estimated the cure probability. The simulation procedure in the previous subsection was considered, except that we studied only $S\,(\cdot\,|\lambda)$ derived from a Weibull distribution. The simulation parameters were fixed as $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300, \sigma_u^2 = 0.5, \sigma_v^2 = 0.5$ and $\rho = 0.5$. Different values for σ_u^2 and σ_v^2 where taken, from 0.5 to 1.0 and values of ρ from 0.5 to 0.8. In addition, we considered three combinations for the number of clinics and the number patients, as follows:

- $J = 5, n_j = 6, j = 1, \ldots, 5$;
- $J = 5, n_j = 12, j = 1, \ldots, 5$, and
- $J = 10, n_j = 6, j = 1, \ldots, 10$.

In each clinic, one third of the patients were randomly assigned to the treatment group ($x_{ij} = 1$) and the remaining, to the control group ($x_{ij} = 0$). The covariates z_{ij} were drawn from the standard normal distribution truncated at the interval (-3,3). Also, two values were randomly imputed as -3 and two values as 3, in order to guarantee extreme values in the covariates, where the effect of an asymmetric link could be more advantageous when compared with a symmetric link. In each case, a total of 500 replicates were simulated. The results are presented in Table 7 in Appendix E. For the estimates of the cure probabilities we computed the average bias and the average of maximum bias for the 500 simulated samples (and the corresponding standard deviations).

In general terms, the choice of a wrong link seems to slightly increase the average bias of the cure probability estimates, but the major problem is that the maximum bias of cure rate estimates can increase considerably.

On the other hand, an increase on the number of patients in each clinic leads to a decrease in the maximum bias associated to the cure probabilities, whose estimates seems to become more accurate, while an increase on the clinics only made the estimates more accurate.

7. FINAL DISCUSSION

This paper extends the idea considered in Lai and Yau [18], by considering the mixture model with bivariate random effects and different link functions for the cure probability (including the logit link).

We adopted a parametric approach, in order to select a link between the proposed functions. Parameter estimation was performed based on the EM algorithm, but in the maximization step we used the REML approach proposed in McGilchrist and Yau (1995). Simulation studies were considered, suggesting an acceptable performance for the proposed estimation method. The new model was also applied to a real data set, and the complementary loglog link function outperformed the two symmetrical links (logit and probit) we considered. This was more evident in the estimation of the cure probability, especially for extreme values of the covariates.

REFERENCES

1. Balakrishnan, N. and Pal, S. (2016). Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Statistical methods in medical research*, 25(4), 1535-1563.
2. Barreto-Souza, W. (2015). Long-term survival models with over dispersed number of competing causes. *Computational Statistics & Data Analysis*, 91, 51-63.
3. Bazán, J.L., Bolfarine, H. and Branco, M.D. (2010). A Framework for Skew-Probit Links in Binary Regression. *Communications in Statistics - Theory and Methods*, 39, 678-697.
4. Bazán, J.L., Branco, M.D. and Bolfarine, H. (2006). A skew item response model. *Bayesian analysis*, 1, 861-892.
5. Bazán, J.L., Romeo, J.S. and Rodrigues, J. (2014). Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics*, 28, 467-482.
6. Berkson, J. and Gage, R. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47, 501-515.
7. Cancho, V.G., Louzada, F. and Ortega, E.M. (2013). The power series cure rate model: an application to a cutaneous melanoma data. *Communications in Statistics - Simulation and Computation*, 42, 586-602.
8. Chen, M.H., Dey, D.K. and Shao, Q.M. (1999a). A New Skewed Link Model for Dichotomous Quantal Response Data. *Journal of the American Statistical Association*, 94, 1172-1186.
9. Chen, M.H., Ibrahim, J.G. and Sinha, D. (1999b). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909-919.
10. Donohue, M.C., Overholser, R., Xu, R. and Vaida, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*, 98(3), 685-700.

11. Fang, H.B., Li, G. and Sun, J. (2005). Maximum likelihood estimation in a semi-parametric logistic/proportional hazards mixture model. *Scandinavian Journal Statistical*, 32, 59-75.
12. Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38, 1041-1046.
13. Friedman, M. (1987). Piecewise Exponential Models for Survival Data with Covariates. *The Annals of Statistics*, 10, 101-113.
14. Gallardo, D.I., Bolfarine, H. and Pedroso-de-Lima, A.C. (2013). Promotion time cure rate model with random effects: an application to a multi-centre clinical trial of carcinoma. *Statistics Research Letters*, 2, 44-52.
15. Gallardo, D.I., Gómez, H.W. and Bolfarine, H. (2016). A new cure rate model based on the Yule-Simon distribution with application to a melanoma data set. *Journal of Applied Statistics*, 44, 1153-1164.
16. Gallardo, D.I., Gómez, Y.M. and De Castro, M. (2017a). A note on the EM algorithm for the destructive negative cure rate model. *Journal of Statistical Computation and Simulation*, 87, 2291-2297.
17. Gallardo, D.I., Gómez, Y.M. and De Castro, M. (2018). A flexible cure rate model based on the polylogarithm distribution. *Journal of Statistical Computation and Simulation*, 88(11), 2137-2149.
18. Gallardo, D.I., Romeo, J.S. and Meyer, R. (2017b). A simplified estimation procedure based on the EM algorithm for the power series cure rate model. *Communications in Statistics - Simulation and Computation*, 46, 6342-6359.
19. Grande, R.H., Pedroso-de-Lima, A.C., Rodrigues, L.E. and Witzel, M.F. (2000). Clinical evaluation of an adhesive used as a fissure sealant. *American Journal of Dentistry*, 13(4), 167-170.
20. Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31, 423-447.
21. Lai, X. and Yau, K. (2008). Long-term survivor model with bivariate random effects: Applications to bone marrow transplant and carcinoma study data. *Statistics in Medicine*, 27, 5692-5708.
22. Lai, X. and Yau, K. (2010). Extending the long-term survivor mixture model with random effects for clustered survival data. *Computational Statistics & Data Analysis*, 54, 2103-2112.
23. Lopes, C.M. and Bolfarine, H. (2012). Random effects in promotion time cure rate models. *Computational Statistics and Data Analysis*, 56, 75-87.
24. Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics*, 60, 545-574.
25. Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure estimation. *Statistica Sinica*, 20, 661-674.
26. McGilchrist, C.A. and Yau, K. (1995). The derivation of BLUP, ML, REML estimation methods for generalized linear mixed models. *Communications in Statistics - Theory and Methods*, 24, 2963-2980.
27. Miller, R.G. (1974). The jackknife: A review. *Biometrika*, 61, 1-15.
28. Pal, S. and Balakrishnan, N. (2016). Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime. *Statistics & Probability Letters*, 116, 9-20.

29. Pal, S. and Balakrishnan, N. (2017). An EM type estimation procedure for the destructive exponentially weighted Poisson regression cure model under generalized gamma lifetime. *Journal of Statistical Computation and Simulation*, 87(6), 1107-1129.
30. Rodrigues, J., Cancho, V.G., Castro, M.A. and Louzada-Neto, F. (2009a). On the unification of the long-term survival models. *Statistics & Probability Letters*, 79, 753-759.
31. Rodrigues, J., Castro, M.A., Balakrishnan, N. and Cancho, V.G. (2012). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*, 17, 333-346.
32. Rodrigues, J., De Castro, M., Cancho, V.G. and Balakrishnan, N. (2009b). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, 139, 3605-3611.
33. Venables, W.N. and Smith, D.M. (2015). R Development Core Team. An introduction to R. Available at http://cran.r-project.org/doc/manuals/R-intro.pdf.
34. Yau K. and Ng, A. (2001). Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in Medicine*, 20, 1591-1607.

APPENDIX A: THE BLUP ESTIMATION PROCEDURE

The Table 4 presents the cumulative distribution function of Logistic, normal and Gumbel distributions in their standard versions, with the respective first and second derivatives.

Table 4: Cumulative Distribution Functions for the Logistic, Normal and Gumbel Distributions and its First and Second Order Derivatives

Distribution	$\psi(u)$	$\psi'(u)$	$\psi''(u)$
Logistic	$\frac{e^u}{1+e^u}$	$\frac{e^u}{(1+e^u)^2}$	$\frac{e^u(1-e^u)}{(1+e^u)^3}$
Normal	$\Phi(u)$	$\phi(u)$	$-u\phi(u)$
Gumbel	$1-e^{-e^u}$	e^{u-e^u}	$-e^{u-e^u}(e^u-1)$

The score vector and **B**, minus the second derivative of ℓ in relation to $\boldsymbol{\varsigma}$, are given by

$$\frac{\partial \ell}{\partial \varsigma} = \begin{bmatrix} \boldsymbol{X}^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}^T \\ \boldsymbol{G}^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^T \end{bmatrix} \begin{bmatrix} \partial \ell_1/\partial \boldsymbol{\xi} \\ \partial \ell_1/\partial \boldsymbol{\vartheta} \end{bmatrix} - \frac{1}{\sigma_u^2\sigma_v^2(1-\rho^2)} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \sigma_v^2\boldsymbol{U} - \rho\sigma_u\sigma_v\boldsymbol{V} \\ \sigma_u^2\boldsymbol{V} - \rho\sigma_u\sigma_v\boldsymbol{U} \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{X}^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}^T \\ \boldsymbol{G}^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{G}^T \end{bmatrix} \begin{bmatrix} \partial^2 \ell_1/\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}^T & \partial^2 \ell_1/\partial \boldsymbol{\xi}\partial \boldsymbol{\vartheta}^T \\ \partial^2 \ell_1/\partial \boldsymbol{\vartheta}\partial \boldsymbol{\xi}^T & \partial^2 \ell_1/\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \mathbf{0} & \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z} & \mathbf{0} & \boldsymbol{G} \end{bmatrix}$$

$$+ \frac{1}{\sigma_u^2\sigma_v^2(1-\rho^2)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_v^2\boldsymbol{I}_J & -\rho\sigma_u\sigma_v\boldsymbol{I}_J \\ 0 & 0 & -\rho\sigma_u\sigma_v\boldsymbol{I}_J & \sigma_u^2\boldsymbol{I}_J \end{bmatrix},$$

with $\boldsymbol{\xi} = \boldsymbol{X\gamma} + \boldsymbol{GU}$ and $\boldsymbol{\vartheta} = \boldsymbol{Z\beta} + \boldsymbol{GV}$. X, Z and G are the design matrices of γ, β and the random effects U and V, respectively. To simplify the notation, let

$$\pi_{ij} = \psi(\vartheta_{ij}), \pi'_{ij} = \psi'(\vartheta_{ij}) \text{ and } \pi''_{ij} = \psi''(\vartheta_{ij}), 1 \le i \le n_j, 1 \le j \le J,$$

and lets $\mathrm{H}_0(\cdot\,|\boldsymbol{\lambda})$ the cumulative hazard function. The first and second partial derivatives of ℓ_1 with respect to the components of $\boldsymbol{\xi}$ and $\boldsymbol{\vartheta}$ as follows.

$$\frac{\partial \ell_1}{\partial \xi_{ij}} = \tilde{y}_{ij}\{\delta_{ij} - e^{\xi_{ij}}\mathrm{H}_0(\mathrm{t}_{ij}|\boldsymbol{\lambda})\}, -\frac{\partial^2 \ell_1}{\partial \xi_{ij}^2} = \tilde{y}_{ij}e^{\xi_{ij}}\mathrm{H}_0(\mathrm{t}_{ij}|\boldsymbol{\lambda}),$$

$$\frac{\partial \ell_1}{\partial \vartheta_{ij}} = \tilde{y}_{ij}\frac{\pi'_{ij}}{\pi_{ij}} + (1-\tilde{y}_{ij})\frac{\pi'_{ij}}{(1-\pi_{ij})},$$

$$-\frac{\partial^2 \ell_1}{\partial \vartheta_{ij}^2} = \tilde{y}_{ij}\left\{\frac{\pi_{ij}\pi''_{ij} - \pi'^2_{ij}}{\pi_{ij}^2}\right\} + (1-\tilde{y}_{ij})\left\{\frac{(1-\pi_{ij})\pi''_{ij} - \pi'^2_{ij}}{(1-\pi_{ij})^2}\right\},$$

$$-\frac{\partial^2 \ell_1}{\partial \xi_{ij}\partial \vartheta_{ij}} = 0, \text{ if } 1 \le i \le n_j, 1 \le j \le J,$$

$$-\frac{\partial^2 \ell_1}{\partial \xi_{ij}\partial \xi_{i'j'}} = -\frac{\partial^2 \ell_1}{\partial \vartheta_{ij}\partial \vartheta_{i'j'}} = -\frac{\partial^2 \ell_1}{\partial \xi_{ij}\partial \vartheta_{i'j'}} = 0, \text{ if } 1 \le i \ne i' \le n_j \text{ and } 1 \le j \ne j' \le J.$$

APPENDIX B: ASYMPTOTIC VARIANCES FOR THE VARIANCE COMPONENT PARAMETER ESTIMATES

It can be verified that

$$\frac{\partial \boldsymbol{\Omega}}{\partial \sigma_u^2} = \boldsymbol{Q}_1 + \frac{\rho\sigma_v}{2\sigma_u}\boldsymbol{Q}_2, \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2} = \frac{-1}{\sigma_u^4\,(1-\rho^2)}\boldsymbol{Q}_1 + \frac{\rho}{2\sigma_u^3\sigma_v\,(1-\rho^2)}\boldsymbol{Q}_2,$$
$$\frac{\partial \boldsymbol{\Omega}}{\partial \sigma_v^2} = \boldsymbol{Q}_3 + \frac{\rho\sigma_u}{2\sigma_v}\boldsymbol{Q}_2, \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2} = \frac{-1}{\sigma_v^4\,(1-\rho^2)}\boldsymbol{Q}_3 + \frac{\rho}{2\sigma_u\sigma_v^3\,(1-\rho^2)}\boldsymbol{Q}_2,$$
$$\frac{\partial \boldsymbol{\Omega}}{\partial \rho} = \sigma_u\sigma_v\boldsymbol{Q}_2 \text{and} \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho} = \frac{2\rho}{(1-\rho^2)^2}\left[\frac{1}{\sigma_u^2}\boldsymbol{Q}_1 + \frac{1}{\sigma_v^2}\boldsymbol{Q}_3\right] - \frac{(1+\rho^2)}{\sigma_u\sigma_v(1-\rho^2)^2}\boldsymbol{Q}_2,$$

with $\boldsymbol{Q}_1, \boldsymbol{Q}_2$ and $\boldsymbol{Q}_3$ defined in (12). Moreover, let

$$\boldsymbol{D}_1 = \widehat{\boldsymbol{A}}_{33}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2}, \boldsymbol{D}_2 = \boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_u^2}, \boldsymbol{D}_3 = \widehat{\boldsymbol{A}}_{33}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2}, \boldsymbol{D}_4 = \boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma_v^2}, \boldsymbol{D}_5 = \widehat{\boldsymbol{A}}_{33}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho}$$
$$\text{and } \boldsymbol{D}_6 = \boldsymbol{\Omega}\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho}.$$

Following Lai and Yau (2008), the variance of $\boldsymbol{\tau}$ is given by the inverse of matrix $2\boldsymbol{K}$, where the elements of matrix $\boldsymbol{K}$ are given by

$$K_{11} = tr\,(\boldsymbol{D}_1 - \boldsymbol{D}_2)^2, K_{12} = tr\,(\boldsymbol{D}_1\boldsymbol{D}_3 + \boldsymbol{D}_2\boldsymbol{D}_4 - \mathbf{2}\boldsymbol{D}_1\boldsymbol{D}_4),$$
$$K_{13} = tr\,(\boldsymbol{D}_1\boldsymbol{D}_5 + \boldsymbol{D}_2\boldsymbol{D}_6 - \mathbf{2}\boldsymbol{D}_1\boldsymbol{D}_6),$$
$$K_{22} = tr\,(\boldsymbol{D}_3 - \boldsymbol{D}_4)^2, K_{23} = tr\,(\boldsymbol{D}_3\boldsymbol{D}_5 + \boldsymbol{D}_4\boldsymbol{D}_6 - \mathbf{2}\boldsymbol{D}_3\boldsymbol{D}_6)$$
$$\text{and } K_{33} = tr\,(\boldsymbol{D}_5 - \boldsymbol{D}_6)^2.$$

APPENDIX C: ESTIMATION OF PARAMETERS FOR THE BASELINE HAZARD

For the Weibull distribution to non-cured individuals, the cumulative hazard is $H_0(t|\boldsymbol{\lambda}) = \lambda t^\alpha$. The first and second partial derivatives of ℓ_p with respect to the parameters of the baseline hazard and matrix are given as follows.

$$\frac{\partial \ell_p}{\partial \lambda} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\tilde{y}_{ij}\left(\frac{\delta_{ij}}{\lambda} + e^{\xi_{ij}}\mathrm{t}_{ij}^{\alpha}\right),$$
$$\frac{\partial \ell_p}{\partial \alpha} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\tilde{y}_{ij}\left(\delta_{ij}\left(\frac{1}{\alpha} + \log t_{ij}\right) + \lambda \mathrm{t}_{ij}^{\alpha}\log t_{ij}\right),$$
$$-\frac{\partial^2 \ell_p}{\partial \lambda^2} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\frac{\tilde{y}_{ij}\delta_{ij}}{\lambda^2}, -\frac{\partial^2 \ell_p}{\partial \alpha^2} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\tilde{y}_{ij}\mathrm{t}_{ij}^{\alpha}\log t_{ij}$$
$$\text{and } -\frac{\partial^2 \ell_p}{\partial \lambda \partial \alpha} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\frac{\tilde{y}_{ij}\delta_{ij}}{\alpha^2} - \lambda \mathrm{t}_{ij}^{\alpha}\log^2 t_{ij}\right).$$

For the PE model, $H_0(t|\boldsymbol{\lambda}) = \sum_{l=1}^{L}\lambda_l \nabla_l\,(t)$, with $\nabla_l\,(t)$ defined in (9). The elements of $\partial\ell_p/\partial\boldsymbol{\lambda}$ and $\partial^2\ell_p/\,(\partial^T\lambda\partial\lambda)$ are given by

$$\frac{\partial \ell_p}{\partial \lambda_l} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\tilde{y}_{ij}\left(\frac{\delta_{ij}}{\lambda_l}I\big(a_{l-1} < t_{ij} \le a_l\big) - e^{\xi_{ij}}\nabla_l\,(t_{ij})\right), l = 1, \dots, L,$$
$$-\frac{\partial^2 \ell_p}{\partial \lambda_l^2} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\frac{\tilde{y}_{ij}\delta_{ij}}{\lambda_l^2}I\big(a_{l-1} < t_{ij} \le a_l\big), l = 1, \dots, L,$$
$$-\frac{\partial^2 \ell_p}{\partial \lambda_l \partial \lambda_{l'}} = 0,\ 1 \le l \ne l' \le L.$$

APPENDIX D: TABLES FOR SIMULATION STUDY 1

Table 5: Estimated Biases for 500 Simulations of REML Estimators (Survival Function of Non-Cured Considered from Weibull Model)

Parameter	True Value	Average Bias	SE_1	SE_2	CP
Simulation set ($\lambda = 0.01, \alpha = 1, C = 500$)					
Simulation 1 ($\rho = 0.2$)					
γ	-0.5	0.052	0.276	0.219	0.941
β_0	0.5	-0.044	0.269	0.225	0.934
β_1	-0.5	0.001	0.298	0.253	0.943
σ_u^2	0.5	-0.066	0.307	0.245	0.905
σ_v^2	0.5	-0.058	0.295	0.273	0.897
ρ	0.2	0.055	0.459	0.390	0.837
Simulation 2 ($\rho = 0.4$)					
γ	-0.5	0.033	0.287	0.217	0.937
β_0	0.5	-0.035	0.267	0.223	0.929
β_1	-0.5	0.028	0.294	0.252	0.954
σ_u^2	0.5	-0.044	0.314	0.254	0.907
σ_v^2	0.5	-0.027	0.309	0.288	0.902
ρ	0.4	0.051	0.426	0.356	0.871
Simulation 3 ($\rho = 0.6$)					
γ	-0.5	0.013	0.280	0.217	0.941
β_0	0.5	-0.039	0.272	0.213	0.932
β_1	-0.5	0.027	0.296	0.249	0.955
σ_u^2	0.5	-0.055	0.281	0.249	0.909
σ_v^2	0.5	-0.031	0.336	0.286	0.903
ρ	0.6	0.045	0.357	0.312	0.885
Simulation 4 ($\rho = 0.8$)					
γ	-0.5	0.013	0.278	0.213	0.947
β_0	0.5	-0.045	0.256	0.209	0.929
β_1	-0.5	0.009	0.288	0.248	0.950
σ_u^2	0.5	-0.033	0.283	0.260	0.920
σ_v^2	0.5	0.031	0.364	0.317	0.903
ρ	0.8	0.014	0.246	0.247	0.957

Table 6
Estimated Biases for 500 Simulations of REML Estimators
(Survival Function of Non-Cured Considered from PE Model)

Parameter	True Value	Average Bias	SE_1	SE_2	CP
Simulation set ($\lambda = 0.01, \lambda_2 = 0.02, C = 500$)					
Simulation 1 ($\rho = 0.2$)					
γ	-0.5	0.034	0.257	0.257	0.938
β_0	0.5	-0.011	0.285	0.242	0.935
β_1	-0.5	0.013	0.260	0.250	0.946
σ_u^2	0.5	-0.045	0.270	0.253	0.902
σ_v^2	0.5	-0.009	0.321	0.299	0.899
ρ	0.2	0.022	0.454	0.385	0.846
Simulation 2 ($\rho = 0.4$)					
γ	-0.5	0.012	0.239	0.215	0.936
β_0	0.5	-0.011	0.263	0.235	0.925
β_1	-0.5	-0.002	0.271	0.249	0.952
σ_u^2	0.5	-0.030	0.289	0.259	0.905
σ_v^2	0.5	-0.018	0.332	0.294	0.901
ρ	0.4	-0.005	0.396	0.363	0.850
Simulation 3 ($\rho = 0.6$)					
γ	-0.5	0.033	0.238	0.213	0.937
β_0	0.5	-0.010	0.247	0.230	0.928
β_1	-0.5	0.012	0.251	0.251	0.958
σ_u^2	0.5	-0.020	0.325	0.263	0.905
σ_v^2	0.5	0.024	0.328	0.315	0.901
ρ	0.6	-0.010	0.326	0.318	0.868
Simulation 4 ($\rho = 0.8$)					
γ	-0.5	0.036	0.239	0.215	0.941
β_0	0.5	-0.019	0.232	0.212	0.926
β_1	-0.5	-0.004	0.263	0.251	0.952
σ_u^2	0.5	-0.006	0.283	0.270	0.912
σ_v^2	0.5	0.028	0.293	0.314	0.898
ρ	0.8	-0.014	0.209	0.242	0.968

APPENDIX E: TABLES FOR SIMULATION STUDY 2

Table 7
Estimated Biases and Maximum Biases Estimates for Cure Rate Estimates using the 3 Links Proposed to the Cure

Set parameter: $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300$
Link used to simulate: Logit

	$J = 5, n_j = 6$		$J = 5, n_j = 12$		$J = 10, n_j = 6$	
Link to cure	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**
Case I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$						
Logistic	0.004 (0.099)	0.244 (0.161)	-0.005 (0.071)	0.200 (0.120)	-0.001 (0.070)	0.246 (0.123)
Probit	-0.005 (0.095)	0.250 (0.159)	-0.012 (0.070)	0.209 (0.121)	-0.011 (0.070)	0.269 (0.124)
C-loglog	0.012 (0.097)	0.275 (0.163)	-0.001 (0.070)	0.229 (0.125)	0.008 (0.070)	0.285 (0.131)
Case II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$						
Logistic	-0.010 (0.094)	0.268 (0.166)	-0.008 (0.070)	0.218 (0.134)	-0.010 (0.066)	0.274 (0.131)
Probit	-0.023 (0.091)	0.274 (0.164)	-0.018 (0.065)	0.222 (0.131)	-0.024 (0.066)	0.296 (0.134)
C-loglog	-0.003 (0.092)	0.294 (0.165)	-0.005 (0.066)	0.237 (0.138)	-0.001 (0.066)	0.313 (0.143)
Case III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$						
Logistic	0.006 (0.093)	0.237 (0.146)	0.005 (0.064)	0.200 (0.124)	0.001 (0.063)	0.230 (0.115)
Probit	-0.004 (0.093)	0.243 (0.146)	-0.004 (0.063)	0.206 (0.124)	-0.009 (0.062)	0.244 (0.118)
C-loglog	0.013 (0.094)	0.266 (0.149)	0.007 (0.064)	0.224 (0.129)	0.009 (0.063)	0.257 (0.125)

Link used to simulate: Probit

	$J = 5, n_j = 6$		$J = 5, n_j = 12$		$J = 10, n_j = 6$	
Link to cure	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**
Case I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$						
Logistic	0.007 (0.088)	0.286 (0.175)	0.002 (0.058)	0.238 (0.134)	0.003 (0.060)	0.298 (0.132)
Probit	-0.005 (0.086)	0.257 (0.161)	-0.008 (0.056)	0.213 (0.123)	-0.010 (0.059)	0.275 (0.126)
C-loglog	0.015 (0.087)	0.283 (0.164)	0.008 (0.058)	0.243 (0.132)	0.014 (0.061)	0.303 (0.137)

Link to cure	Av. Bias (sd)	Max. Bias (sd)	Av. Bias (sd)	Max. Bias (sd)	Av. Bias (sd)	Max. Bias (sd)
Case II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$						
Logistic	0.008 (0.088)	0.311 (0.187)	0.002 (0.067)	0.264 (0.161)	0.001 (0.063)	0.326 (0.163)
Probit	-0.010 (0.082)	0.277 (0.172)	-0.014 (0.060)	0.230 (0.144)	-0.016 (0.061)	0.296 (0.152)
C-loglog	0.015 (0.085)	0.310 (0.181)	0.005 (0.061)	0.259 (0.154)	0.011 (0.062)	0.322 (0.160)
Case III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$						
Logistic	0.014 (0.086)	0.283 (0.166)	0.009 (0.066)	0.240 (0.138)	0.008 (0.061)	0.302 (0.136)
Probit	0.001 (0.081)	0.263 (0.154)	-0.003 (0.063)	0.206 (0.124)	-0.005 (0.060)	0.272 (0.133)
C-loglog	0.021 (0.083)	0.288 (0.159)	0.013 (0.064)	0.236 (0.134)	0.018 (0.063)	0.302 (0.142)

Table 8
Estimated Biases and Maximum Biases Estimates for Cure Rate Estimates using the 3 Links Proposed to the Cure (Continuation)

Set parameter: $\gamma = -0.5, \beta_0 = -0.5, \beta_1 = 0.5, \lambda = 0.01, \alpha = 1, C = 300$
Link used to simulate: C-loglog

	$J = 5, n_j = 6$		$J = 5, n_j = 12$		$J = 10, n_j = 6$	
Link to cure	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**	**Av. Bias (sd)**	**Max. Bias (sd)**
Case I: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.5$						
Logistic	-0.005 (0.100)	0.263 (0.163)	-0.004 (0.069)	0.229 (0.137)	-0.008 (0.068)	0.290 (0.125)
Probit	-0.015 (0.096)	0.277 (0.156)	-0.014 (0.066)	0.214 (0.129)	-0.021 (0.065)	0.284 (0.118)
C-loglog	0.002 (0.096)	0.260 (0.161)	-0.001 (0.066)	0.225 (0.132)	0.000 (0.064)	0.299 (0.123)
Case II: $\sigma_u^2 = 1.0, \sigma_v^2 = 1.0, \rho = 0.5$						
Logistic	0.005 (0.103)	0.309 (0.191)	-0.009 (0.072)	0.259 (0.148)	-0.009 (0.072)	0.319 (0.143)
Probit	-0.017 (0.099)	0.314 (0.183)	-0.024 (0.069)	0.241 (0.137)	-0.031 (0.064)	0.303 (0.131)
C-loglog	0.001 (0.099)	0.301 (0.185)	-0.007 (0.069)	0.249 (0.141)	-0.003 (0.064)	0.317 (0.137)
Case III: $\sigma_u^2 = 0.5, \sigma_v^2 = 0.5, \rho = 0.8$						
Logistic	0.013 (0.099)	0.256 (0.160)	0.002 (0.064)	0.283 (0.126)	0.000 (0.072)	0.273 (0.124)
Probit	0.019 (0.099)	0.257 (0.161)	-0.012 (0.062)	0.271 (0.117)	-0.012 (0.071)	0.272 (0.121)
C-loglog	0.002 (0.098)	0.241 (0.164)	0.008 (0.062)	0.257 (0.123)	0.007 (0.070)	0.265 (0.126)