A score test for assessing the cured proportion in the long-term survivor mixture model

Yun Zhao[1], Andy H. Lee[1,*,†], Kelvin K. W. Yau[2], Valerie Burke[3] and Geoffrey J. McLachlan[4]

[1]*School of Public Health, Curtin Health Innovation Research Institute, Curtin University of Technology, GPO Box U 1987, Perth, WA 6845, Australia*
[2]*Department of Management Sciences, City University of Hong Kong, Hong Kong, China*
[3]*School of Medicine and Pharmacology, University of Western Australia, Perth, Australia*
[4]*Centre for Statistics, University of Queensland, Brisbane, Australia*

SUMMARY

The long-term survivor mixture model is commonly applied to analyse survival data when some individuals may never experience the failure event of interest. A score test is presented to assess whether the cured proportion is significant to justify the long-term survivor mixture model. Sampling distribution and power of the test statistic are evaluated by simulation studies. The results confirm that the proposed test statistic performs well in finite sample situations. The test procedure is illustrated using a breast cancer survival data set and the clustered multivariate failure times from a multi-centre clinical trial of carcinoma. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: cured proportion; long-term survivors; mixture model; random effects; score test

1. INTRODUCTION

Long-term survivor or cure models are applicable when a certain fraction of the population never experiences the failure event of interest. Typically, a proportion of individuals completely recover after treatment and do not suffer relapse or death from the disease. A popular approach for analysing such survival data is to formulate the model in terms of a mixture of two-component distributions; one component for the cured group and another component for members of the uncured subpopulation. A review of survival analysis with long-term survivors can be found in

---

*Correspondence to: Andy H. Lee, School of Public Health, Curtin Health Innovation Research Institute, Curtin University of Technology, GPO Box U 1987, Perth, WA 6845, Australia.
†E-mail: andy.lee@curtin.edu.au

Reference [1]. The cure model has been extended to proportional hazards and semiparametric settings [2–4]. For correlated outcomes, a long-term survivor mixture model incorporating random effects was developed to handle multivariate failure times when the survival data are obtained from a multi-centre clinical trial [5].

In practice, it is important to assess whether the cured proportion is significant to justify the fitting of the long-term survivor mixture model. In this paper, a score statistic is presented to test the single-component Weibull survival model against the alternative long-term survivor mixture model. Both independent and correlated situations are considered, the latter renders the introduction of random effects in the model to explain the variability shared by members within a subgroup, or shared by multiple failure observations within an individual. The advantage of the score test statistic lies in its computational simplicity of fitting just the Weibull model under the null hypothesis. In the related context of mixture models, score tests for zero-inflation and over-dispersion have been proposed for correlated count data [6–8].

After briefly reviewing the long-term survivor mixture model with random effects in Section 2, the hypotheses concerning the cured proportion and corresponding score test are specified in Section 3. A simulation study is conducted in Section 4 to investigate the sampling distribution of the score test statistic and its power properties. Two practical examples are then used in Section 5 to illustrate the test procedure in detail. Finally, some remarks regarding applicability of the methodology are given in Section 6.

2. LONG-TERM SURVIVOR MIXTURE MODEL WITH RANDOM EFFECTS

Suppose the data are

$$(t_{ij}, x_{ij}, \delta_{ij}), \quad j = 1, \ldots, n_i, \ i = 1, \ldots, M \quad \text{and} \quad \sum_{i=1}^{M} n_i = N$$

where t_{ij} is the jth observed survival time within cluster i, δ_{ij} is the censoring indicator for the failure event and x_{ij} is the associated vector of covariates. There are n_i observations in each cluster. Let Y_{ij} be an unobserved binary variable with value 1 denoting a failure event (uncured) but 0 if the individual will never experience the event (cured). The uncured probability p can be specified by a logistic form:

$$p(x_{ij}) = P(Y_{ij} = 1 | x_{ij}) = \frac{1}{1 + e^{\xi_{ij}}}$$

where $\xi_{ij} = w'_{ij}\gamma$, $w_{ij} = (1, x'_{ij})'$ and the parameter vector γ represents the effects of covariates on p. The unconditional survival function can then be expressed as

$$S(t_{ij}; x_{ij}) = 1 - p(x_{ij}) + p(x_{ij})S_u(t_{ij}; x_{ij})$$

where $S_u(t_{ij}; x_{ij}) = S_{u0}^{e^{\eta_{ij}}}$ is the conditional survival function for the uncured group, $\eta_{ij} = x'_{ij}\beta$ is the linear predictor and β is a $q \times 1$ vector of regression coefficients. The conditional hazard function for the uncured group may be assumed to follow Cox's proportional hazards form

$$h_u(t_{ij}; x_{ij}) = h_{u0}(t_{ij})e^{\eta_{ij}}$$

where $h_{u0}(t_{ij})$ is the baseline hazard function for the uncured group.

In order to account for the clustering of the multivariate failure times, unobserved random effect terms are introduced into the long-term survivor mixture model via the cured probability and the hazard function, viz,

$$\eta_{ij} = x_{ij}'\boldsymbol{\beta} + U_i, \quad \xi_{ij} = w_{ij}'\boldsymbol{\gamma} + V_i$$

where the random effects U_i and V_i are assumed to be independent and follow the normal $N(0, \theta_1)$ and $N(0, \theta_2)$ distribution, respectively. Here, U_i denotes the unobserved random effect of the ith cluster for uncured individuals, with larger values implying a higher risk and shorter survival, whereas V_i denotes the unobserved random effect of the ith cluster determining the cured probability, with larger values implying a higher probability of belonging to the cure group.

Because Cox's partial likelihood approach cannot eliminate the baseline hazard function, a parametric form may be adopted. Suppose the Weibull distribution $h_{u0}(t) = \lambda \alpha t_{ij}^{\alpha-1}$, λ, $\alpha > 0$, is used as the baseline hazard [5], the log-likelihood function of the survival time becomes:

$$l_1 = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \{(1 - \delta_{ij}) \log S(t_{ij}; x_{ij}) + \delta_{ij} \log f(t_{ij}; x_{ij})\} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \{\log S(t_{ij}; x_{ij}) + \delta_{ij} \log h(t_{ij}; x_{ij})\}$$

Let $\mathbf{u} = (U_1, \ldots, U_M)'$ and $\mathbf{v} = (V_1, \ldots, V_M)'$. For parameter estimation, a best linear unbiased predictor (BLUP) type log-likelihood can be constructed as $l = l_1 + l_2$, with l_1 as defined above and

$$l_2 = -\frac{1}{2}[M \log 2\pi\theta_1 + (1/\theta_1)\mathbf{u}'\mathbf{u} + M \log 2\pi\theta_2 + (1/\theta_2)\mathbf{v}'\mathbf{v}]$$

Here, l may be considered as a penalized log-likelihood with l_2 being the penalty function for the conditional log-likelihood l_1 when the random effects are conditionally fixed. Let $\Omega = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{u}, \mathbf{v})$ be the vector of unknown parameters. The estimating procedure starts from a BLUP estimate of Ω as the initial step and extends to obtain residual maximum quasi-likelihood (REMQL) estimators of Ω along with random component variance estimates for θ_1 and θ_2. Details of fitting the long-term survivor mixture model with random effects and associated numerical algorithms can be found in Reference [5].

3. SCORE TEST FOR THE CURED PROPORTION

To assess whether the cured proportion is significant to justify the long-term survivor mixture model, a score test procedure is developed as follows. We are interested in testing the null hypothesis $H_0: p = 1$ against the alternative hypothesis $H_1: p < 1$, which is equivalent to testing $H_0^*: \varphi = 0$ versus $H_1^*: \varphi > 0$, by letting $\varphi = (1 - p)/p$, and $0 \leqslant \varphi < \infty$ for $0 \leqslant 1 - p < 1$.

Under the null hypothesis, the model reduces to the standard Weibull random effects model with conditional log-likelihood

$$l_1 = \sum_{i=1}^{M} \sum_{j=1}^{n_i} l_{1ij} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} (1 - \delta_{ij}) \log[\varphi + S_{u0}^{e^{\eta_{ij}}}] + \delta_{ij} \log[h_{u0} e^{\eta_{ij}} S_{u0}^{e^{\eta_{ij}}}] - \log[1 + \varphi]$$

and $l_2 = -(1/2)[M \log(2\pi\theta_1) + (1/\theta_1)\mathbf{u}'\mathbf{u}]$. Let $\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1$ be the corresponding REMQL estimates for the single-component Weibull model. Given the first derivatives of l with respect to $\boldsymbol{\beta}$, $\mathbf{u}$, θ_1, φ and the Fisher information matrix $\boldsymbol{\Gamma}(\boldsymbol{\beta}, \mathbf{u}, \theta_1, \varphi)$, the score test statistic for $H_0^*: \varphi = 0$ is constructed as

$$S(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1, 0) = U'(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1, 0)\hat{\boldsymbol{\Gamma}}^{-1} U(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1, 0)$$

where the score function

$$U(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1, 0) = \left(0, \ldots, 0, \sum_{i=1}^{M} \sum_{j=1}^{n_i} \left[(1-\delta_{ij}) \frac{1}{S_{u0}^{e^{\hat{\eta}_{ij}}}} - 1 \right] \right)$$

$\hat{\eta}_{ij}$ is evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1)$ and the baseline parameters λ and α may be estimated by a profile likelihood approach through a two-dimensional grid search method [5]. The entries of $\hat{\boldsymbol{\Gamma}}$ are obtained from the second derivatives of l evaluated at $\varphi = 0$ and $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$ and $\hat{\theta}_1$; see the Appendix for details. It should be noted that when the cured proportion is close to 100 per cent, the score test statistic will increase indefinitely. On the other hand, when the cured proportion is close to 0 per cent, the score test statistic will approach 0.

Now, denote the inverse of $\boldsymbol{\Gamma}(\boldsymbol{\beta}, \mathbf{u}, \theta_1, \varphi)$ by

$$\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}$$

with the corresponding partition same as $\boldsymbol{\Gamma}$, the score statistic is then expressed as

$$S = \mathbf{H}_{22} \left(\sum_{i=1}^{M} \sum_{j=1}^{n_i} \left[(1-\delta_{ij}) \frac{1}{S_{u0}^{e^{\hat{\eta}_{ij}}}} - 1 \right] \right)^2$$

where $\mathbf{H}_{22} = (\hat{\boldsymbol{\Gamma}}_{22} - \hat{\boldsymbol{\Gamma}}_{21} \hat{\boldsymbol{\Gamma}}_{11}^{-1} \hat{\boldsymbol{\Gamma}}_{12})^{-1}$ and $\hat{\boldsymbol{\Gamma}}_{11}$, $\hat{\boldsymbol{\Gamma}}_{12}$, $\hat{\boldsymbol{\Gamma}}_{22}$ are defined in the Appendix. For the case of independent observations, $\mathbf{u} = \mathbf{0}$, so that

$$S(\hat{\boldsymbol{\beta}}, 0) = U'(\hat{\boldsymbol{\beta}}, 0) \hat{\boldsymbol{\Gamma}}^{-1} U(\hat{\boldsymbol{\beta}}, 0)$$

Further simplifications of the mathematical formulae are possible but details are omitted for brevity. In view of the one-sided alternative hypothesis $H_1^*: \varphi > 0$, according to statistical theory on tests involving the boundary of the parameter space [1, Chapter 7], the reference distribution of the score test statistic S is asymptotically $0.5(\chi_0^2 + \chi_1^2)$ with P-value given by $0.5 P\{\chi_1^2 > S\}$, i.e. the limiting distribution follows a mixture of a degenerate point mass at 0 and a χ_1^2 distribution in equal mixing proportions.

4. SAMPLING DISTRIBUTION AND POWER STUDY

4.1. Sampling distribution

A simulation study is conducted to investigate the distribution of the score statistic under finite sample situations. The simulation design essentially follows that of Yau and Ng [5], except that $\mathbf{u}$ is set to 0 for simplicity. The working model under the null hypothesis $H_0: p = 1$ is taken to be the standard Weibull model with linear predictor

$$\eta_i = -0.5 x_i, \quad i = 1, 2, \ldots, N$$

The single covariate x_i is generated as a Bernoulli random variable with probability 0.2. Realizations of the unobservable variable Y are generated in which an individual has a probability of $1 - p(x_i)$

being cured ($Y=0$) and a probability of $p(x_i)$ being uncured ($Y=1$), with $p(x_i)$ given by the logistic model defined in Section 2 and $\gamma_0 = \gamma_1 = -0.5$. The failure times of the cured individuals are assigned to be infinite. For each uncured individual, a failure time is generated from the conditional probability density function

$$f_u(t; x_{ij}) = h_{u0}(t) \exp \eta_{ij} S_{u0}(t)^{\exp \eta_{ij}}$$

We assume the baseline hazard $h_{u0}(t_{ij})$ for the uncured group to follow a Weibull distribution with known parameters $\lambda = 0.005$ and $\alpha = 1$. If the generated failure time exceeds a constant censoring time C, it is taken to be censored at time C. The sampling distribution of S is considered for $C = 600$, 700 and 800 and sample sizes $N = 100, 200, 400, 600, 800$ and 1000.

The empirical ordered S statistics based on 500 replications are compared with the corresponding quantiles of a 50:50 mixture of the χ_1^2 distribution and a point mass at 0. The Q–Q plots for $C = 800$ and various N's are presented in Figure 1; results for other constant C values are similar. It is evident that the sampling distribution of S follows closely the asymptotic reference distribution even for moderate sample sizes.

4.2. Empirical power

Performance of the score test procedure is next evaluated under the long-term survivor mixture model. Parameter values are specified as those in Section 4.1, except that $\lambda = 0.01$. For a given uncured proportion p and significance level a, the empirical power of the score test is calculated using the estimated upper tail probabilities of S at $\chi_1^2(1-a)$ under the alternative hypothesis $H_1 : p \neq 1$, i.e.

$$P\{S > \chi_1^2(1-a)\} \approx \sum_{k=1}^{500} I[S_k > \chi_1^2(1-a)|H_1]/500$$

where S_k is the observed score statistic at the kth replication trial, $k = 1, \ldots, 500$. A range of uncured proportions ($p = 0.65, 0.75, 0.85, 0.95$) are considered, together with the commonly adopted significance levels $a = 0.1, 0.05$ and 0.01 and sample sizes $N = 400, 600, 800$ and 1000 are used. The constant censoring time C is set at 600, 700 and 800, producing moderate censorships of about 20–30 per cent in the generated data sets.

The results in Table I show that the score test is generally powerful in rejecting the null hypothesis for moderate cured fractions, and remain satisfactory for large p except when N and the preset censored time C are both small. As expected, a more powerful test can be produced by increasing the sample size and the prescribed level of significance. The empirical power also improves with C as the percentage of censoring decreases.

To further investigate the performance of the score test when the cure assumption is invalid, 500 replicated data sets are generated under the null hypothesis $p = 1$. The working model is thus a Weibull survival model. With $C = 800$ and other parameter values same as those specified in the empirical power study, the observed score test statistics and corresponding P-values are computed for the generated data sets. The results show that the proportion of committing the Type I error is small and decreases with the sample size N. Specifically, the proportions of Type I error are found to be: 4.2 per cent ($N = 200$), 3.8 per cent ($N = 400$), 3.6 per cent ($N = 600$), 3.6 per cent ($N = 800$), 2.8 per cent ($N = 1000$). Therefore, the score test also performs satisfactorily when cure is not a reasonable assumption.
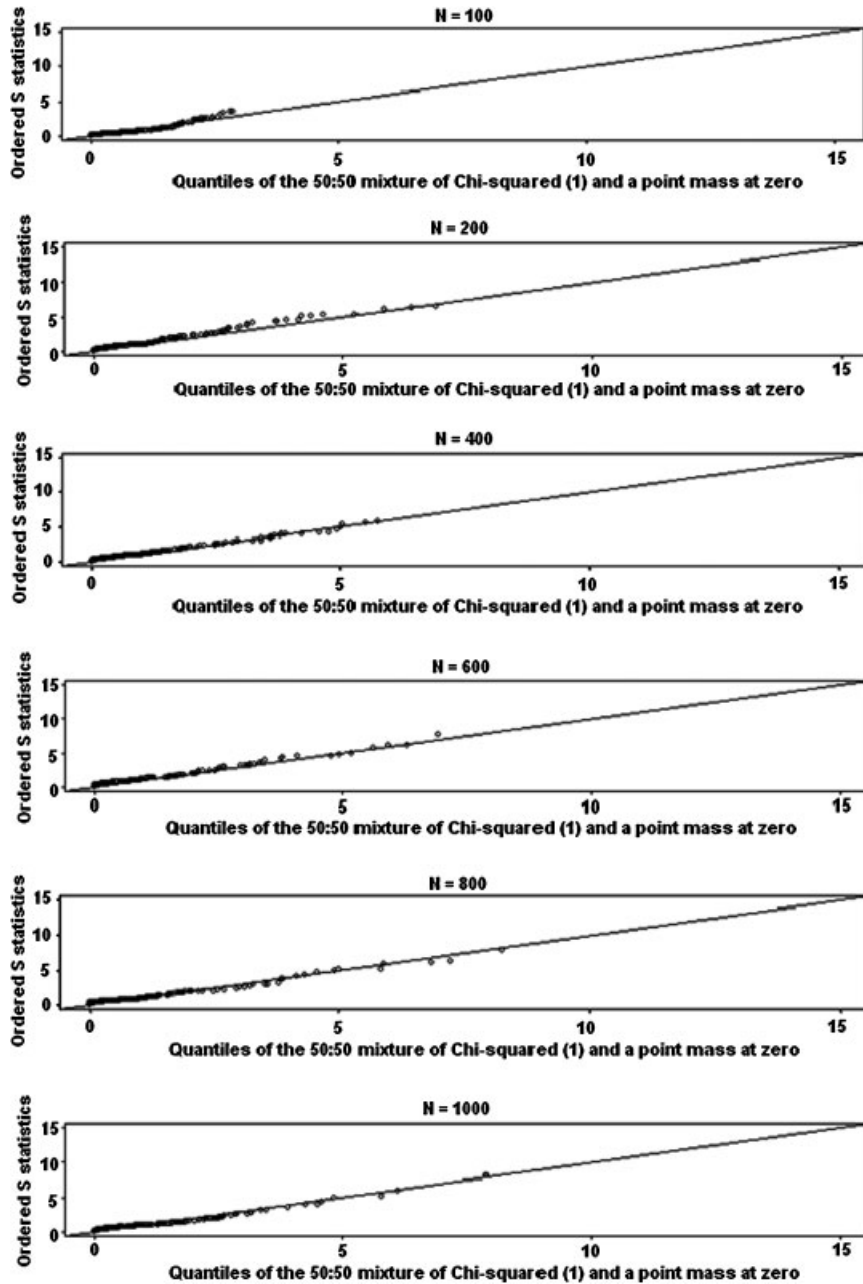
Figure 1. Q–Q plots of ordered score statistics against the quantiles of a 50:50 mixture of the χ_1^2 distribution and a point mass at 0.

Table I. Empirical power of the score statistic S based on 500 replications generated from a long-term survivor mixture model.

	$p=0.65$ a			$p=0.75$ a			$p=0.85$ a			$p=0.95$ a		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$C=600$												
N												
400	0.98	0.85	0.23	0.99	0.96	0.65	0.97	0.91	0.51	0.39	0.21	0.04
600	1.00	0.99	0.66	1.00	1.00	0.92	1.00	1.00	0.92	0.80	0.63	0.25
800	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	0.99	0.74	0.56	0.19
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.82	0.42
$C=700$												
N												
400	1.00	0.99	0.70	1.00	1.00	0.97	1.00	1.00	0.94	0.78	0.57	0.18
600	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.95	0.68
800	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.92	0.68
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94
$C=800$												
N												
400	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.89	0.50
600	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
800	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

5. APPLICATIONS

5.1. Breast cancer data

Consider the survival times (in years) of 45 breast cancer patients after mastectomy [1, Chapter 6]. This data subset, with 19 (42.2 per cent) censored observations, came from a clinical study to assess the long-term prognosis by lectin binding [9]. In the original study, 179 patients were recruited and the paraffin sections of their primary breast cancers were dewaxed and stained by a lectin derived from the albumin gland of the Roman snail, *Helix pomatia*, with a hope that the metastatic potential of an individual breast cancer might be assessed at the time of operation [9]. A single covariate was available, indicating the status of positive staining or negative staining of the cancer cells. In this example, the outcome was cause-specific mortality so that deaths due to other causes were treated as censored observations, whereas for those patients whose cancer cells being removed/cleared completely after treatment, they would be regarded as 'cured'. In the presence of such cured observations, a plateau in the survival curve is expected for sufficiently long follow-up time.

Results from fitting the Weibull and the long-term survivor mixture models suggest that patients with positive staining are associated with a reduced survival; with $\hat{\beta}=0.865$ (SE 0.107) and 1.308 (SE 0.136) being significant under both the single-component model and two-component mixture model, respectively. For the latter, the estimated γ regression coefficients in the logistic part are -0.357 (SE 1.947) and -0.241 (SE 1.033). Based on the long-term survivor model, the cured proportion estimate for the positive staining group (0.355) is less than that for the negative staining
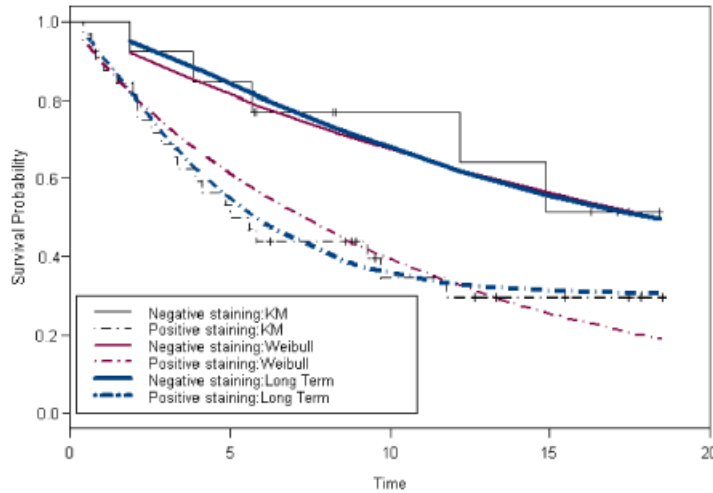
Figure 2. Kaplan–Meier curves and estimated survival functions from fitting the Weibull and long-term survivor mixture models for the breast cancer data.

group (0.412). The overlaid Kaplan–Meier curves and the estimated survival functions from fitting the Weibull and the long-term survivor models are plotted in Figure 2. The three estimation approaches are generally consistent and the negative staining group exhibited a longer survival time than the positive staining group. The appropriateness of the long-term survivor model is further justified by the observed score test statistic S of 14.298 (P-value <0.001) with respect to its asymptotic reference distribution.

5.2. Multi-centre carcinoma data

The score test is next illustrated using clustered multivariate failure times data [5] obtained from a multi-centre clinical trial [10]. The data set contains information on 66 patients with squamous carcinoma of the pharyngeal tongue. The six participating centres are regarded as a random sample from all institutions. Death from causes unrelated to the carcinoma is treated as censored, whereas patients responded favourably to radiation therapy and subsequently free of disease symptoms may be considered cured. The study objective was to determine the effect of tumour stage classification on both the survival and the cured proportion, with $x=1$ indicating a massive tumour with extension to adjoining tissue and $x=0$ refers to a small primary tumour. A total of 47 deaths and 19 (29 per cent) censored observations were recorded.

According to the log-rank test, patients with a small primary tumour survived significantly longer than those with a massive tumour, the Kaplan–Meier estimates being 798.56 (SE 89.16) days and 391.12 (SE 94.29) days, respectively, with P-value $=0.009$. Fitting a single-component Weibull model leads to $\hat{\beta}=0.656$ (SE 0.305) and random component variance estimate $\hat{\theta}=0.065$ (SE 0.123). On the other hand, REMQL estimates for $(\gamma_0, \gamma_1, \beta, \theta_1, \theta_2)$ are -0.945 (SE 0.406), -0.79 (SE 0.868), 0.618 (SE 0.294), 0.20 (SE 0.549), 0.011 (SE 0.105), respectively, based on the long-term survivor mixture model with random effects. The latter results suggest that the presence of a massive tumour will significantly reduce the patient's survival, with estimated hazard ratio 1.86, but have little impact on the cured proportion. Under the assumption of long-term survivors, the
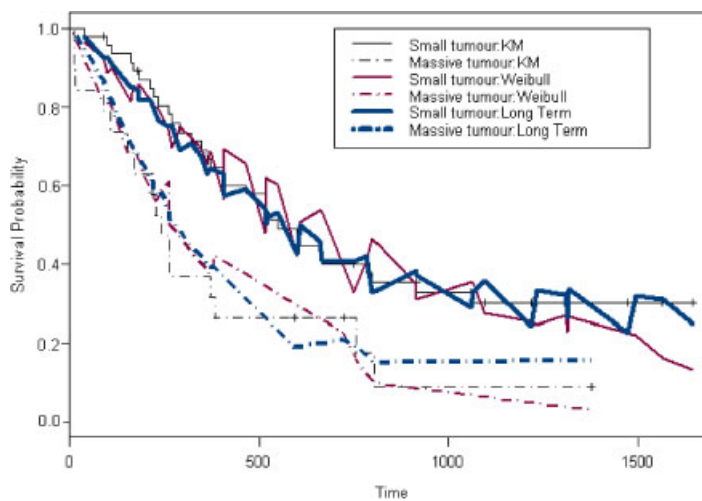
Figure 3. Kaplan–Meier curves and estimated survival functions from fitting the Weibull and long-term survivor mixture models for the multi-centre carcinoma data.

cured probability of patients with a small primary tumour and a massive tumour are estimated to be 0.28 and 0.15, respectively. The Kaplan–Meier curves and the estimated survival functions from fitting the Weibull and the long-term survivor models are plotted together in Figure 3, all of which indicate that patients with a small primary tumour experienced better survival than their counterparts with a massive tumour. The score test statistic S is 16.286 with P-value <0.001, confirming the existence of long-term survivors among these carcinoma patients.

6. DISCUSSION

A score test is proposed for assessing the proportion of long-term survivors. Unlike the likelihood ratio test, the advantage of the score statistic lies in its computational convenience. The score test does not require, under the null hypothesis, the long-term survivor mixture model to be fitted. The test procedure has been implemented as an Splus computer program available from the corresponding author. Although asymptotic inferences for survival mixture models are generally lacking, our simulation results show that the score test has high power and the test statistic follows a 50:50 mixture distribution of χ_1^2 and a point mass at 0 for the settings considered. From a practical viewpoint, the nominal significance level obtained enables the assessment of cured proportion for the subpopulation of long-term survivors.

The Weibull distribution is widely adopted as the parametric baseline in survival analysis. For the simulation study, alternative parametric baseline distributions are likely to give different results. However, we anticipate the difference will not be large as long as the corresponding survival patterns are similar.

Applications to the breast cancer data and the multi-centre clinical trial of carcinoma demonstrate the usefulness of the test procedure in both independence and clustered multivariate failure time situations. In the presence of a significant cured proportion in the sample, fitting of the alternative long-term survivor mixture model [5] is recommended, and comparisons should be made with its

corresponding single-component Weibull model. On the other hand, if the score test suggests little departure from the Weibull model, the resulting inferences can be made with increased confidence. In practice, sufficiently long follow-up time is needed in order to observe the cured proportion, which is revealed as a plateau in the survival curve. To determine whether the follow-up time is 'sufficient', knowledge about the disease is required.

In this paper, the random effects U_i and V_i in the long-term survivor model are assumed to be independent. With correlated random effects U_i and V_i, analogous score test statistics can be developed along similar lines but with corresponding adjustments in the formulation of the BLUP type log-likelihood and the derivative expressions. Moreover, the proposed methodology can be extended for testing the single-component model against an alternative two-component survival mixture model [11] in addition to model selection methods such as Akaike's information criterion and Bayesian information criterion. Generalizations to time-dependent covariates and nonparametric mixture model settings are also feasible and will be pursued in the future research.

APPENDIX

The derivation of the Fisher information matrix in Section 3 is presented here.

$$\frac{\partial l_{1ij}}{\partial \eta_{ij}} = (1-\delta_{ij})\frac{S_{u0}^{e^{\eta_{ij}}} e^{\eta_{ij}} \ln S_{u0}}{\varphi + S_{u0}^{e^{\eta_{ij}}}} + \delta_{ij}(1+e^{\eta_{ij}} \ln S_{u0})$$

$$\frac{\partial l_{1ij}}{\partial \varphi} = (1-\delta_{ij})\frac{1}{\varphi + S_{u0}^{e^{\eta_{ij}}}} - \frac{1}{1+\varphi}$$

$$\frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2} = (1-\delta_{ij})S_{u0}^{e^{\eta_{ij}}} e^{\eta_{ij}} \ln S_{u0}\frac{(1+e^{\eta_{ij}}\ln S_{u0})\varphi + S_{u0}^{e^{\eta_{ij}}}}{(\varphi + S_{u0}^{e^{\eta_{ij}}})^2} + \delta_{ij}e^{\eta_{ij}} \ln S_{u0}$$

$$\frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi} = (1-\delta_{ij})\frac{-S_{u0}^{e^{\eta_{ij}}} e^{\eta_{ij}} \ln S_{u0}}{(\varphi + S_{u0}^{e^{\eta_{ij}}})^2}$$

When $\varphi = 0$,

$$E\left(-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi}\right) = E\left[(1-\delta_{ij})\frac{e^{\eta_{ij}} \ln S_{u0}}{S_{u0}^{e^{\eta_{ij}}}}\right], \quad E\left(-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\right) = E[-e^{\eta_{ij}} \ln S_{u0}]$$

Let $\mathbf{S} = S_{u0}^{e^{\eta}} = \exp(-\lambda t_{ij}^{\alpha} e^{\eta_{ij}})$, $\mathbf{S1} = -\log \mathbf{S} = -\partial^2 l_{1ij}/\partial \eta_{ij}^2$, $\mathbf{S2} = (1-\delta_{ij}) - \mathbf{S1}/S_{u0}^{e^{\eta}} = -\partial^2 l_{1ij}/\partial \eta_{ij} \partial \varphi$,

$$\mathbf{D}' = \left(\frac{\partial \eta_{11}}{\partial \boldsymbol{\beta}}, \dots, \frac{\partial \eta_{1n_1}}{\partial \boldsymbol{\beta}}, \dots, \frac{\partial \eta_{M1}}{\partial \boldsymbol{\beta}}, \dots, \frac{\partial \eta_{Mn_M}}{\partial \boldsymbol{\beta}}\right)_{q \times N}$$

$$\mathbf{C}' = \left(\frac{\partial \eta_{11}}{\partial \mathbf{u}}, \dots, \frac{\partial \eta_{1n_1}}{\partial \mathbf{u}}, \dots, \frac{\partial \eta_{M1}}{\partial \mathbf{u}}, \dots, \frac{\partial \eta_{Mn_M}}{\partial \mathbf{u}}\right)_{M \times N} \quad \text{and} \quad \mathbf{I}'_N = (1, 1, \dots, 1)_{1 \times N}$$

The first-order derivatives of l with respect to $\boldsymbol{\beta}$, $\mathbf{u}$, θ_1, φ are

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial l_{1ij}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} (1-\delta_{ij}) \frac{S_{u0}^{e^{\eta_{ij}}} e^{\eta_{ij}} x'_{ij} \ln S_{u0}}{\varphi + S_{u0}^{e^{\eta_{ij}}}} + \delta_{ij} x'_{ij}(1+e^{\eta_{ij}} \ln S_{u0})$$

$$\frac{\partial l}{\partial \mathbf{u}} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial l_{1ij}}{\partial \mathbf{u}} + \frac{\partial l_2}{\partial \mathbf{u}} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} (1-\delta_{ij}) \frac{S_{u0}^{e^{\eta_{ij}}} e^{\eta_{ij}} \ln S_{u0}}{\varphi + S_{u0}^{e^{\eta_{ij}}}} + \delta_{ij}(1+e^{\eta_{ij}} \ln S_{u0}) - \frac{\mathbf{u}}{\theta_1}$$

$$\frac{\partial l}{\partial \varphi} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial l_{1ij}}{\partial \varphi} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} (1-\delta_{ij}) \frac{1}{\varphi + S_{u0}^{e^{\eta_{ij}}}} - \frac{1}{1+\varphi}$$

$$\frac{\partial l}{\partial \theta_1} = \frac{\partial l_2}{\partial \theta_1} = -\frac{1}{2}\left[\frac{M}{\theta_1} - \frac{\mathbf{u}'\mathbf{u}}{\theta_1^2}\right]$$

Entries of the Fisher information matrix $\hat{\boldsymbol{\Gamma}}$ can be computed from the second-order derivatives of l evaluated at $\varphi = 0$, as follows:

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\beta\beta}} = E\left[-\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)^2\right]$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}'} E\left[-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\right] \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}} = \mathbf{D}'\mathbf{S1D}$$

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\beta u}} = E\left[-\frac{\partial^2 l}{\partial \mathbf{u} \partial \boldsymbol{\beta}'}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \boldsymbol{\beta} \partial \mathbf{u}} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2} \frac{\partial \eta_{ij}}{\partial \mathbf{u}} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right]$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}'} E\left[-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\right] \frac{\partial \eta_{ij}}{\partial \mathbf{u}} = \mathbf{D}'\mathbf{S1C}$$

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}\theta_1} = E\left[-\frac{\partial^2 l}{\partial \boldsymbol{\beta}' \partial \theta_1}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \boldsymbol{\beta} \partial \theta_1} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2} \frac{\partial \eta_{ij}}{\partial \theta_1} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right] = 0$$

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}\varphi} = E\left[-\frac{\partial^2 l}{\partial \boldsymbol{\beta}' \partial \varphi}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \boldsymbol{\beta} \partial \varphi} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}} E\left[-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi}\right]$$

$$= \mathbf{D}'\mathbf{S2I}_N$$

$$\hat{\boldsymbol{\Gamma}}_{\mathbf{uu}} = E\left[-\frac{\partial^2 l}{\partial \mathbf{u} \partial \mathbf{u}'}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \mathbf{u} \partial \mathbf{u}'} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\left(\frac{\partial \eta_{ij}}{\partial \mathbf{u}}\right)^2\right] + \frac{1}{\theta_1}$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial \eta_{ij}}{\partial \mathbf{u}'} E\left[-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij}^2}\right] \frac{\partial \eta_{ij}}{\partial \mathbf{u}} + \frac{1}{\theta_1} = \mathbf{C}'\mathbf{S1C} + \frac{1}{\theta_1}$$

$$\hat{\mathbf{\Gamma}}_{\mathbf{u}\theta_1} = E\left[-\frac{\partial^2 l}{\partial \mathbf{u}' \partial \theta_1}\right] = -\frac{\mathbf{u}}{\theta_1^2}$$

$$\hat{\mathbf{\Gamma}}_{\mathbf{u}\varphi} = E\left[-\frac{\partial^2 l}{\partial \mathbf{u}' \partial \varphi}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} -E\left[\frac{\partial l_{1ij}}{\partial \eta_{ij}}\frac{\partial^2 \eta_{ij}}{\partial \mathbf{u} \partial \varphi} + \frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi}\frac{\partial \eta_{ij}}{\partial \mathbf{u}}\right]$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{n_i} \frac{\partial \eta_{ij}}{\partial \mathbf{u}} E\left[-\frac{\partial^2 l_{1ij}}{\partial \eta_{ij} \partial \varphi}\right] = \mathbf{C}'\mathbf{S}2\mathbf{I}_N$$

$$\hat{\mathbf{\Gamma}}_{\theta_1\theta_1} = E\left[-\frac{\partial^2 l}{\partial \theta_1^2}\right] = \frac{\mathbf{u}'\mathbf{u}}{\theta_1^3} - \frac{M}{2\theta_1^2}$$

$$\hat{\mathbf{\Gamma}}_{\theta_1\varphi} = E\left[-\frac{\partial^2 l}{\partial \theta_1 \partial \varphi}\right] = 0$$

$$\hat{\mathbf{\Gamma}}_{\varphi\varphi} = E\left[-\frac{\partial^2 l}{\partial \varphi^2}\right] = \sum_{i=1}^{M} \sum_{j=1}^{n_i} E\left[(1-\delta_{ij})\frac{1}{(S_{u0}^{e^{\eta_{ij}}})^2} - 1\right]$$

The matrix $\mathbf{\Gamma}(\mathbf{\beta}, \mathbf{u}, \tau, \varphi)$ can be partitioned as $\begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{12}' & \mathbf{\Gamma}_{22} \end{pmatrix}$, where

$$\mathbf{\Gamma}_{22} = \Gamma_{\varphi\varphi}, \quad \mathbf{\Gamma}_{12}' = (\Gamma_{\mathbf{\beta}\varphi}, \Gamma_{\mathbf{u}\varphi}, \Gamma_{\theta_1\varphi}) \quad \text{and} \quad \mathbf{\Gamma}_{11} = \begin{bmatrix} \Gamma_{\mathbf{\beta}\mathbf{\beta}} & \Gamma_{\mathbf{\beta}\mathbf{u}} & \Gamma_{\mathbf{\beta}\theta_1} \\ & \Gamma_{\mathbf{u}\mathbf{u}} & \Gamma_{\mathbf{u}\theta_1} \\ & & \Gamma_{\theta_1\theta_1} \end{bmatrix}$$

Hence $\hat{\mathbf{\Gamma}}(\hat{\mathbf{\beta}}, \hat{\mathbf{u}}, \hat{\theta}_1, 0) = \begin{pmatrix} \hat{\mathbf{\Gamma}}_{11} & \hat{\mathbf{\Gamma}}_{12} \\ \hat{\mathbf{\Gamma}}_{12}' & \hat{\mathbf{\Gamma}}_{22} \end{pmatrix}$, where

$$\hat{\mathbf{\Gamma}}_{22} = \sum_{i=1}^{M} \sum_{j=1}^{n_i} E\left[(1-\delta_{ij})\frac{1}{S^2} - 1\right], \quad \hat{\mathbf{\Gamma}}_{12}' = (-\mathbf{I}_N'\mathbf{S}2\mathbf{D}, -\mathbf{I}_N'\mathbf{S}2\mathbf{C}, 0)_{1\times(q+M+1)} \quad \text{and}$$

$$\hat{\mathbf{\Gamma}}_{11} = \begin{pmatrix} \mathbf{D}'\mathbf{B}\mathbf{D} & \mathbf{D}'\mathbf{B}\mathbf{C} & 0 \\ \mathbf{C}'\mathbf{B}\mathbf{D} & \mathbf{C}'\mathbf{B}\mathbf{C} & -\dfrac{\mathbf{u}}{\theta_1} \\ 0 & -\dfrac{\mathbf{u}'}{\theta_1} & \dfrac{\mathbf{u}'\mathbf{u}}{\theta_1^3} - \dfrac{M}{2\theta_1^2} \end{pmatrix}_{(q+M+1)\times(q+M+1)}$$

REFERENCES

1. Maller RA, Zhou X. *Survival Analysis with Long-term Survivors*. Wiley: Chichester, 1996.
2. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.
3. Tsodikov AD. Estimation of survival based on proportional hazards when cure is a possibility. *Mathematical and Computer Modeling* 2001; **33**:1227–1236.
4. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**:237–243.
5. Yau KKW, Ng ASK. Long-term survivor mixture model with random effects: application to a multi-centre clinical trial of carcinoma. *Statistics in Medicine* 2001; **20**:1591–1607.
6. Jung BC, Jhun M, Song SH. Testing for overdispersion in a censored Poisson regression model. *Statistics* 2006; **40**:533–543.
7. Xiang L, Lee AH, Yau KKW, McLachlan GJ. A score test for zero-inflation in correlated count data. *Statistics in Medicine* 2006; **25**:1660–1671.
8. Xiang L, Lee AH, Yau KKW, McLachlan GJ. A score test for overdispersion in zero-inflated Poisson mixed regression model. *Statistics in Medicine* 2007; **26**:1608–1622.
9. Leathem AJ, Brooks SA. Predictive value of lectin binding on breast-cancer recurrence and survival. *Lancet* 1987; **1**(8541):1054–1056.
10. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: New York, 1980.
11. Ng ASK, McLachlan GJ, Yau KKW, Lee AH. Modelling the distribution of ischaemic stroke-specific survival time using an EM-based mixture approach with random effects adjustment. *Statistics in Medicine* 2004; **23**:2729–2744.