Regression Analysis for Cure Rate under a Random Cure Time Model

Weijing Wang
NSC midterm report: 92-2118-M-009-007
Email: wjwang@stat.nctu.edu.tw

Institute of Statistics, National Chiao-Tung University
Hsin-Chu, Taiwan, R.O.C.

May, 2004

Summary

This article considers regression analysis for cure models in presence of competing risks. The model is formulated by a mixture representation. The main interest is in the incidence part, which measures the probability of a specific type of failure or cure rate. Assuming a binary regression model, several inference methods for estimating the regression parameters are proposed to handle the missing cure status due to censoring. The latency distributions, despite of less interest, play an essential role to utilize the partial but biased information provided by censored data. Alternatively a distribution-free procedure is also developed given that the quality of the data is good enough in terms of sufficient follow-up.

Key words: Cause-specific hazard; Competing Risks; Cure models; EM algorithm; Illness-Death Models; Imputation; Logistic regression; Missing Data; Mixture model; Sufficient follow-up; Susceptibility.

1 Introduction

Cure models allow for the possibility of not developing the event of interest despite of long-term follow-up. Several versions of cure models have appeared in the literature which differ in how “cure is defined. Under the classical setting, cure is not clearly specified in the sense that cured (or immune) individuals are not directly observed but always mixed with temporarily censored but susceptible ones. The book by Maller and Zhou (1996) is an excellent reference on the subject. Another type of the models assumes that a patient is cured if he/she does not develop the event of interest within a prespecified time period. One can refer to Laska and Meinser (1992) for further references. For the model considered in the article, cure is determined by the order of competing events. Throughout the paper, we will use the severe acute respiratory syndrome (SARS) as an illustrating example of such models. SARS is a life-threatening acute disease that resulted in a global outbreak in 2003. The phenomenon can be described by a two-path model, sometimes known as an illness-death model, depicted in Figure 1, where subjects may follow two different paths, $1 \to 2 \to 3$ or $1 \to 3$. For the SARS example, the intermediate state refers to hospital discharge with recovery and the absorbing state is death. Betensky and Schoenfled (2001) call the third type as cure models with random cure times.

Cure models are often expressed by a mixture formulation that contains two components: one component is related to long-term incidence and the other is related to the latency distribution given the cure status. Covariates may have different influences on the two parts and the mixture formulation provides a flexible way to study the effects separately. For the classical type of model, Farewell (1982) assumed a logistic/Weibull model. Several authors, including Kuk and Chen (1992), Sy and Taylor (2000) and Peng and Dear (2000), have considered logistic/Cox regression model. Their methods differ in handling the baseline hazard function in the estimation. Taylor (1995) proposed a logistic/Kaplan-Meier approach to analyze the incidence model by leaving the latency distribution un-specified. However his method imposes a rigid assumption that the latency distribution does not depend on the covariates. For the cure model with fixed cure time, the focus is on the incidence part since the latency time is a fixed constant. Jung (1996) and Subramanian (2001) proposed inference methods to estimate parameters in a binary regression model under censoring.

In this paper, we consider regression analysis for the third type of cure models in which cure is determined by the order of competing risks. The primary goal is to assess the effects of covariates on the cure rate. Using the SARS example, let T_1 be the time to hospital discharge and T_2 be the time to death. Define $\Delta = I(T_1 \leq T_2)$ as the indicator for failure type which is the indicator for the cure event. Given the value of Δ, two latency distributions can be Defined, namely $Q_1(t) = \Pr(T_1 > t | T_1 \leq T_2)$ and $Q_2(t) = \Pr(T_2 > t | T_1 > T_2)$. Notice that $Q_j(t)$ $(j = 1, 2)$ also represent the survival functions of the sojourn times in the two-path model and the cause-specific survival functions in the context of competing risks. We assume that $\Pr(T_1 \leq T_2 | Z) = \Pr(\Delta = 1 | Z) = \pi(\beta' Z)$, where $Z : p \times 1$ denotes a vector of covariates, $\pi^{-1}(\cdot)$ is the link function which is monotonic and differentiable. A common example is the logistic regression model

$$\pi(\beta' Z) = \frac{\exp(\beta' Z)}{1 + \exp(\beta' Z)}.$$

In general, covariates also affect the latency distributions and thus we should write $Q_{1,Z}(t) = \Pr(T_1 > t | T_1 \leq T_2, Z)$ and $Q_{2,Z}(t) = \Pr(T_2 > t | T_1 > T_2, Z)$. In the SARS example, it seems more important to investigate what factors affect whether a patient can be cured than to study how long it takes for them to recover. Therefore we prefer not making rigid assumptions on the forms of $Q_{j,Z}(t)$ $(j = 1, 2)$.

The major objective is to develop inference methods for estimating β when the cure status, Δ, may be unknown due to censoring. In Section 2, we first review the inference procedure which complete information of Δ is available and then discuss the likelihood inference given censored data. In Section 3 several inference methods for estimating β are proposed. Section 4 contains concluding remarks.

2 Preliminary Analysis

Let $\{(T_{1i}, T_{2i}, \Delta_i)$ $(i = 1, \ldots, n)\}$ be identically and independently replications of (T_1, T_2, Δ). When the observation period is long enough such that Δ_i $(i = 1, \ldots, n)$ are completely observed, standard techniques for generalized linear models can be applied to estimate β. Based on complete data, $\{(\Delta_i, Z_i)$ $(i = 1, \ldots, n)\}$, where $\Delta_i = I(T_{1i} \leq T_{2i})$, the likelihood function becomes

$$L(\beta) = \prod_{i=1}^{n} \pi(\beta' Z_i)^{\Delta_i} \{1 - \pi(\beta' Z_i)\}^{1-\Delta_i}, \tag{1}$$

which gives the score equation

$$\tilde{U}(\beta) = \sum_{i=1}^{n} \{\Delta_i - \pi(\beta' Z_i)\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i, \tag{2}$$

where $\pi_\phi(t) = \partial \pi(t)/\partial t$ and $\bar{\pi}(t) = 1 - \pi(t)$.

In practice it happens that subjects may drop out from the study or, at the end of the follow-up, some patients still have not developed the events interest. Let C be the censoring variable and assume that it is independent of both T_1 and T_2. Under competing risks, one observes $X_i = T_{1i} \wedge T_{2i} \wedge C_i$, $\delta_{1i} = I(T_{1i} \leq T_{2i} \wedge C_i)$, $\delta_{2i} = I(T_{2i} \leq T_{1i} \wedge C_i)$ for $i = 1, \ldots, n$. Letting $\delta_{3i} = I(C_i \leq T_{1i} \wedge T_{2i})$, $\delta_{1i} + \delta_{2i} + \delta_{3i} = 1$. Note that when $\delta_{1i} = 1$, $\Delta_i = 1$, while $\delta_{2i} = 1$, $\Delta_i = 0$. However the value of Δ_i is unknown if $\delta_{3i} = 1$.

Without loss of generality, assume that T_1, T_2 and C are all continuous variables. Based on censored data, the likelihood function becomes

$$L_C \propto \prod_{i=1}^{n} \left\{ [\pi(\beta' Z_i) f_{1,Z}(x_i)]^{\delta_{1i}} [\bar{\pi}(\beta' Z_i) f_{2,Z}(x_i)]^{\delta_{2i}} [S_Z(x_i;\beta)]^{\delta_{3i}} \right\}, \tag{3}$$

where $f_{j,Z}(x) = -\frac{\partial}{\partial x} Q_{j,Z}(x)$ for $j = 1, 2$ and

$$S_Z(x_i;\beta) = \pi(\beta' Z_i) Q_{1,Z}(x_i) + \bar{\pi}(\beta' Z_i) Q_{2,Z}(x_i).$$

Notice that when censoring exists, the likelihood function of β contains nuisance parameters related to $Q_{j,Z}(t)$ $(j = 1, 2)$ which implies that additional assumption on the latency distributions is required if likelihood-based inference is pursued. Parametric regression models may be assumed for $Q_{j,Z}(t)$ $(j = 1, 2)$. Larsen and Dinse (1985) considered the same framework as discussed here and assumed a logistic/piecewise-exponential model. In the next section, we review maximum likelihood estimation and then propose other inference methods under more flexible assumptions.

3 The Proposed Methods

3.1 EM algorithm for maximum likelihood estimation

Previous analysis implies that, when censoring is present, likelihood estimation of β requires additional knowledge on the latency distributions. Assume temporarily that a parametric form is imposed on $Q_{j,Z}(t)$ $(j = 1, 2)$. Since direct maximization of L_C in (3) is difficult, the EM algorithm can be used to obtain the maximum likelihood estimator. The likelihood based on pseudo-data, $\{(\Delta_i, \delta_{1i}, \delta_{2i}, X_i, Z_i)\ (i = 1, \ldots, n)\}$, is

$$L_f \propto \prod_{i=1}^{n} \left\{ \left(\pi(\beta' Z_i) \left[h_{1,Z_i}(x_i)\right]^{\delta_{3i}} \left[Q_{1,Z_i}(x_i)\right]\right)^{\Delta_i} \left(\overline{\pi}(\beta' Z_i) \left[h_{2,Z_i}(x_i)\right]^{\delta_{3i}} \left[Q_{2,Z_i}(x_i)\right]\right)^{1-\Delta_i} \right\},$$

where $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$ and $h_{j,Z}(x) = \left[-\frac{\partial}{\partial x} Q_{j,Z}(x)\right] \Big/ Q_{j,Z}(x)$ for $j = 1, 2$. The E-step takes the expectation of $\log L_f$ with respect to the distribution of the unobserved Δ_is, given the observed data and the current parameter values. The expected log-likelihood, denoted as $l_f(\beta, Q_1, Q_2, w^{(m)})$, becomes

$$\sum_{i=1}^{n} \left\{ \sum_{j=1}^{2} \left[\delta_{ji} \cdot ln(h_{j,Z_i}(x_i))\right] + w_i^{(m)} \cdot ln(\pi(\beta' Z_i) Q_{1,Z_i}(x_i)) + (1 - w_i^{(m)}) \cdot ln(\overline{\pi}(\beta' Z_i) Q_{2,Z_i}(x_i)) \right\},$$

where

$$w_i^{(m)} = \delta_{1i} + I(\delta_{1i} = \delta_{2i} = 0) \cdot \frac{\pi(\beta^{(m)\prime} Z_i) Q_{1,Z_i}^{(m)}(x_i)}{\pi(\beta^{(m)\prime} Z_i) Q_{1,Z_i}^{(m)}(x_i) + \overline{\pi}(\beta^{(m)\prime} Z_i) Q_{2,Z_i}^{(m)}(x_i)}$$

and $\beta^{(m)}$, $Q_{j,Z_i}^{(m)}$ $(j = 1, 2)$ denoted the current parameter values at the mth iteration. Note that the last term in $w_i^{(m)}$ is the conditional probability that the ith patient will be eventually cured given that the cure event has not occurred by time x_i.

Further, we can write

$$l_f(\beta, Q_1, Q_2, w^{(m)}) = L_{Q_1} + L_{Q_2} + L_\beta$$

where

$$L_{Q_1} = \sum_{i=1}^{n} \left\{ \delta_{1i} \cdot ln(h_{1,Z_i}(x_i)) + w_i^{(m)} \cdot ln(Q_{1,Z_i}(x_i)) \right\}$$

$$L_{Q_2} = \sum_{i=1}^{n} \left\{ \delta_{2i} \cdot ln(h_{2,Z_i}(x_i)) + (1 - w_i^{(m)}) \cdot ln(Q_{2,Z_i}(x_i)) \right\}$$

and

$$L_\beta = \sum_{i=1}^{n} \left\{ w_i^{(m)} \cdot ln(\pi(\beta' Z_i)) + (1 - w_i^{(m)}) \cdot ln(\bar{\pi}(\beta' Z_i)) \right\}.$$

It is important to note that in the M-step of the algorithm, L_{Q_1}, L_{Q_2} and L_β can be maximized separately by treating $w_i^{(m)}$ as a fixed constant. The EM procedure is iterative in a way that the estimates obtained previously are used to update the value of $w_i^{(m)}$ in the current maximization step. Maximization of L_β is straightforward while maximization of L_{Q_j} depends on the form of $Q_{j,Z}(x)$ $(j = 1, 2)$. For parametric analysis, convergence and properties of the resulting estimates follow the standard results.

3.2 Imputation by Conditional Mean

One alternative to adjust for censoring is to directly modify the score equation in (2). Based on the available data, we can impute the missing Δ_i, if $\delta_{3i} = 1$, by its conditional mean which equals

$$\tilde{\pi}_i = E[\Delta_i | T_{1i} \wedge T_{2i} > C_i, C_i = x_i, Z_i] = \frac{Q_{1,Z_i}(x_i)\pi(\beta' Z_i)}{Q_{1,Z_i}(x_i)\pi(\beta' Z_i) + Q_{2,Z_i}(x_i)\bar{\pi}(\beta' Z_i)}.$$

Here we propose two estimating functions of β which can avoid going through the maximization procedure of $L_{Q_j,Z}$ $(j = 1, 2)$. Both methods modify the results in Wang (2003) by assuming that the covariates take finite number of values. By partitioning the sample according to the value of Z, we can use Wangs method to obtain the non-parametric estimators, $\hat{Q}_{j,z_k}(t)$ $(j = 1, 2)$ and $\hat{p}_{z_k}(x)$ for $k = 1, \ldots, J$, where $\hat{p}_Z(x)$ is an estimator of

$$p_z(x) = E(\Delta | \delta_3 = 1, X = x, Z = z).$$

An estimating function by model-based imputation is given by

$$U_1(\beta) = \sum_{i=1}^{n} \left\{ \hat{\Delta}_i - \pi(\beta' Z_i) \right\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i, \tag{4}$$

where $\hat{\Delta}_i = 1$ for $\delta_{1i} = 1$; $\hat{\Delta}_i = 0$ for $\delta_{2i} = 1$; and if $\delta_{3i} = 1$, we set

$$\hat{\Delta}_i = \frac{\hat{Q}_{1,Z_i}(x_i)\pi(\beta' Z_i)}{\hat{Q}_{1,Z_i}(x_i)\pi(\beta' Z_i) + \hat{Q}_{2,Z_i}(x_i)\bar{\pi}(\beta' Z_i)}.$$

An estimating function by model-free imputation is given by

$$U_2(\beta) = \sum_{i=1}^{n} \left\{ \tilde{\Delta}_i - \pi(\beta' Z_i) \right\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i, \tag{5}$$

where $\tilde{\Delta}_i = 1$ for $\delta_{1i} = 1$; $\tilde{\Delta}_i = 0$ for $\delta_{2i} = 1$; and if $\delta_{3i} = 1$, we set $\tilde{\Delta}_i = \hat{p}_{z_i}(x_i)$. Denote $\hat{\beta}_j$ as the solution of $U_j(\beta) = 0$ for $j = 1, 2$.

3.3 Inverse Probability Weighting

When censoring is present, there is higher chance of first observing an event associated with smaller failure time. To see this, one can express $\delta_1 = \Delta I(T_1 \leq C)$ which implies that δ_1 is a biased proxy of Δ in the presence of censoring. The larger value of failure time T_1, the higher chance that Δ will be censored. When covariates exist, it can be shown that

$$E[I(\delta_1 = 1)|T_1, T_2, Z] = E[I(T_1 \leq T_2 \wedge C)|T_1, T_2, Z] = I(T_1 \leq T_2|Z) G_Z(T_1),$$

where $G_Z(t) = \Pr(C > t|Z)$. It follows that

$$E(\Delta|Z) = E\left(\frac{\delta_1}{G_Z(X)} \middle| Z \right) = E\left[E\left(\frac{\Delta I(T_1 \leq C)}{G_Z(T_1)} \middle| T_1, T_2, Z \right) \right].$$

To simplify the analysis, we let $G_Z(t) = \Pr(C > t)$ which is estimated by the Kaplan-Meier estimator

$$\hat{G}(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{k=1}^{n} I(X_k = u, \delta_{3k} = 1)}{\sum_{k=1}^{n} I(X_k \geq u)} \right\}.$$

Replacing Δ_i by $\delta_{1i}/\hat{G}(X_i)$, one obtain the following estimating function

$$U_3(\beta) = \sum_{i=1}^{n} \left\{ \frac{I(\delta_{1i} = 1)}{\hat{G}(X_i)} - \pi(\beta' Z_i) \right\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i.$$

The estimator of β can be obtained as a solution of $U_3(\beta) = 0$, denoted as $\hat{\beta}_3$. The technique of inverse probability weighting has been widely used in recent literature of survival analysis such as the papers by Jung (1996) and Lin, Sun and Ying (1999).

The proposed way of bias adjustment by inverse weighting implicitly assumes that there is no information about Δ beyond the observational period. Specifically, define

τ_t=sup$\{t : \Pr(T_1 \wedge T_2 > t) > 0\}$ and τ_c=sup$\{t : \Pr(C > t) > 0\}$. If $\tau_t > \tau_c$, which implies that $\Pr(T_1 \wedge T_2 > \tau_c) > 0$, we have

$$\begin{aligned} E\left(\frac{\delta_1}{G(X)}\middle| T_1, T_2\right) &= \begin{cases} \Delta & \text{if } \ T_1 \wedge T_2 \le \tau_c \\ 0 & \text{if } \ T_1 \wedge T_2 > \tau_c \end{cases} \\ &= I(T_1 \wedge T_2 \le \tau_c)\Delta \ne \Delta. \end{aligned}$$

The requirement of $\tau_t \le \tau_c$ is a condition of sufficient follow-up. We can show that, when $\tau_t \le \tau_c$ and under some regularity conditions, $\hat{\beta}_3$ is consistent and $\sqrt{n}(\hat{\beta}_3 - \beta_0)$ converges to a mean-zero normal random variable, where β_0 is the true value of β. The asymptotic variance of $\hat{\beta}_3$ is also derived.

Alternatively, we can use (δ_1, δ_2) jointly to construct an estimating equation. By the same idea, we have the estimating function

$$U_4(\beta) = \sum_{i=1}^{n} \left\{ \frac{\delta_{1i} - \delta_{2i}}{\hat{G}(X_i)} - (\pi(\beta' Z_i) - \bar{\pi}(\beta' Z_i)) \right\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i.$$

The finite-sample comparison of $U_3(\beta)$ and $U_4(\beta)$ could be explored in the simulation study.

3.4 Imputation by unconditional mean

In the context of competing risks, $\Pr(\Delta = 1)$ can be measured from the incidence function. Specifically let $T_1^* = T_1$ if $T_1 \le T_2$ and $T_1^* = \infty$ if $T_1 > T_2$. It follows that

$$S^*(t) = \Pr(T_1^* > t) = \Pr(T_1 > t, T_1 \le T_2) + \Pr(T_1 > T_2).$$

Hence

$$\lim_{t \to \infty} \Pr(T_1^* > t) = \Pr(T_1 > T_2) = \Pr(\Delta = 0).$$

When a nonparametric estimator of $S^*(t)$ is available, denoted as $S^*_{np}(t)$, we may consider the estimating equation

$$U_5(\beta) = \sum_{i=1}^{n} \left\{ (1 - \hat{S}^*_{np}(x_{max})) - \pi(\beta' Z_i) \right\} \frac{\pi_\phi(\beta' Z_i)}{\pi(\beta' Z_i)\bar{\pi}(\beta' Z_i)} Z_i, \tag{6}$$

where x_{max} is the maximum value of X_i $(i = 1, \ldots, n)$ with $\delta_{1i} = 1$. There exist several nonparametric estimators of $S^*(t)$.

3.5 Comparison of the methods

The formulation of mixture models suggests that the incidence probability and the latency distributions can be modeled separately. Due to the problem of identifiability as discussed in Li et al. (2001), classical cure models rely on joint estimation of the two components even if only one part is of major interest. For the cure model considered here, the event of cure is explicitly defined and hence identifiability is not an issue. Consequently the requirement on the lengthy of the follow-up period is less strict. In Section 2.1 we have seen that as long as $C_i > T_{1i} \wedge T_{2i}$ for all $i = 1, \ldots, n$, it is natural to estimate the two components separately. The focus here is on the incidence part. When there exist some observations with $C_i > T_{1i} \wedge T_{2i}$, additional information is needed beside the binary regression model itself.

In the likelihood-based analysis, model specification on $Q_{j,Z}(t)$ $(j = 1, 2)$ is required. Specifically for a censored observation with $\delta_{1i} = \delta_{2i} = 0$, the knowledge of the latency distributions is used in w_i or $\tilde{\pi}_i$ for weight assignment since the observed sojourn time as well as the covariate still reveal useful information. Ignoring such information would lead to bias results. However making additional assumptions increases the possibility of making mistakes. Parametric modeling on the latency distributions is usually the standard approach. Flexibility of the imposed models is an important concern. In this article, we discuss nonparametric analysis for the latency distributions. It is important to note that although no distributional assumption is made, nonparametric analysis requires that the subjects are independently and identically distributions or homogeneous in some sense. Thats why we need to assume that $Q_{j,Z}(t)$ $(j = 1, 2)$ do not depend on Z or Z is discrete so that the sample can be partitioned into homogeneous subgroups. The likelihood-based estimator of using logistic/Kaplan-Meier approach by Taylor (1995) is computationally intensive since it involves back-and-forward iterations between β and high dimensional parameters in $Q_{j,Z}(t)$ $(j = 1, 2)$. The proposed estimators β_1 and β_2, that use the plug-in approach to handle the nuisance parameters separately, are easier to implement. It is possible that finding the root of

$U_1(\beta) = 0$ is difficult since it is a complicated function of β. One alternative is to modify $\hat{\Delta}_i$ by

$$\check{\Delta}_i = \frac{\hat{Q}_{1,Z_i}(x_i)\pi(\hat{\beta}'_{(k-1)}Z_i)}{\hat{Q}_{1,Z_i}(x_i)\pi(\hat{\beta}'_{(k-1)}Z_i) + \hat{Q}_{2,Z_i}(x_i)\bar{\pi}(\hat{\beta}'_{(k-1)}Z_i)},$$

where $\hat{\beta}_{(k-1)}$ is the previously estimated value of β for $k \geq 1$. A reasonable choice of $\hat{\beta}_{(0)}$ is $\hat{\beta}_2$.

Now define $\tau_t = \sup\{t : \Pr(T_1 \wedge T_2 > t) > 0\}$ and $\tau_c = \sup\{t : Pr(C > t) > 0\}$. The estimators, $\hat{\beta}_j$ $(j = 1, 2, 3)$, all rely on the assumption of sufficient follow-up, that is $\tau_t \leq \tau_c$. The validity of $\hat{\beta}_3$ strongly depends on this condition since

$$E\left(\frac{\delta_1}{G_z(X)}\middle| T_1, T_2, Z\right) = I(T_1 \wedge T_2 \leq \tau_c) \cdot \Delta.$$

In general,

$$E\left(\frac{\delta_1}{G_Z(X)}\middle| Z\right) = E\left(I(T_1 \wedge T_2 \leq \tau_c) \cdot \Delta | Z\right) \neq E(\Delta|Z).$$

Since Wangs nonparametric estimators reply on the assumption of sufficient follow-up, $\hat{\beta}_1$ and $\hat{\beta}_2$ are also affected via the imputed values. However Wang proposed a modification when this condition is violated and can also be used here to correct the problem to some degree.

4 Concluding Remarks

Literature on illness-death models is abundant. There have been increasing research interests to analyze the problem in the framework of competing risks or mixture models. One advantage of the mixture formulation is that it allows Separate modeling for incidence and latency distributions. ¿From our analysis, the purpose of joint estimation is for correctly utilizing partial information provided by censored data. One can avoid making model assumption on the latency part if follow-up is sufficient. We see that there is a tradeoff between making strong assumptions on the quality of data or on the model. The techniques used to handle the effect of censoring can be viewed as applications of the principles for handling missing data that are reviewed in the book by Little and Rubin (2002).

References

Andersen, P. K., Borgan, ∅., Gill, R. D., and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Betensky, R. A. and Schoenfeld, D. A. (2001) Nonparametric estimation in a cure model with random cure times. *Biometrics*, **57**, 282–286.

Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant survival data. *J. Am. Statist. Ass.*, **72**, 27–36.

Kuk, A. Y. C. and Chen, C. (1992) A mixture model combining logistic regression with proportional hazards regressions. *Biometrika*, **79**, 531–541.

Farewell, V. T. (1982) The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041–1046.

Fine, J. P., Jiang, H. and Chappell, R. (2001) On semi-competing risks data. *Biometrika*, **88**, 907–919.

Hougaard, P. (2000) *Analysis of Multivariate Survival Analysis*. New York: Springer-Verlag.

Klein, J. P. and Moeschberger, M. L. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Laska, E. M. and Meisner, M. J. (1992) Nonparametric estimation and testing in a cure model. *Biometrics*, **48**, 1223–1234.

Larson, M. G. and Dinse, G. E. (1985) A mixture model for the regression analysis of competing risks data. *Applied Statistics*, **34**, 201–211.

Li, C.-S., Taylor, J. M. G. and Sy, J. P. (2001) Identifiability of cure models. *Statistics & Probability Letters*, **54**, 389–395.

Lin, D. Y., Sun, W. and Ying, Z. (1999) Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*, **86**, 59–70.

Maller, R. A. and Zhou, S. (1992) Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, 731–739.

Maller, R. A. and Zhou, S. (1994) Testing for sufficient follow-up and outliers in survival data. *J. Am. Statistic. Assoc.*, **89**, 1499–1506.

Maller, R. A. and Zhou, S. (1996) *Survival Analysis with Long-Term Survivors.* Wiley: New York.

Peng, Y. and Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**, 237-243.

Prentice, R.L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Taylor, J. M. G. (1995) Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899–907.

Wang, W. (2003) Inference on the association parameters for copula models under dependent censoring. *J. R. Statist. Soc.* B, **65**, 257–273.

Zhao, L. P. and LeMarchand, L. (1992) An analytical method for assessing patterns of familiar aggregation in case-control studies.

Genetic Epidemiology, **9**, 141–154.