Biostatistics (2002), **3**, 4, *pp.* 547–563
Printed in Great Britain

The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure

NGAYEE J. LAW*

Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, CA 94502, USA

ngayee.law@celeradiagnostics.com

JEREMY M. G. TAYLOR

Department of Biostatistics, University of Michigan, Ann Arbor, 1420 Washington Heights, Ann Arbor, MI 48109, USA

HOWARD SANDLER

Department of Radiation Oncology, University of Michigan, Ann Arbor, 1500 East Medical Center Dr., Ann Arbor, MI 48109, USA

SUMMARY

In this paper we present an extension of cure models: to incorporate a longitudinal disease progression marker. The model is motivated by studies of patients with prostate cancer undergoing radiation therapy. The patients are followed until recurrence of the prostate cancer or censoring, with the PSA marker measured intermittently. Some patients are cured by the treatment and are immune from recurrence. A joint-cure model is developed for this type of data, in which the longitudinal marker and the failure time process are modeled jointly, with a fraction of patients assumed to be immune from the endpoint. A hierarchical nonlinear mixed-effects model is assumed for the marker and a time-dependent Cox proportional hazards model is used to model the time to endpoint. The probability of cure is modeled by a logistic link. The parameters are estimated using a Monte Carlo EM algorithm. Importance sampling with an adaptively chosen t-distribution and variable Monte Carlo sample size is used. We apply the method to data from prostate cancer and perform a simulation study. We show that by incorporating the longitudinal disease progression marker into the cure model, we obtain parameter estimates with better statistical properties. The classification of the censored patients into the cure group and the susceptible group based on the estimated conditional recurrence probability from the joint-cure model has a higher sensitivity and specificity, and a lower misclassification probability compared with the standard cure model. The addition of the longitudinal data has the effect of reducing the impact of the identifiability problems in a standard cure model and can help overcome biases due to informative censoring.

Keywords: Cure models; Longitudinal models; Monte Carlo EM; Prostate cancer; Prostate specific antigen; Survival analysis.

*To whom correspondence should be addressed

1. INTRODUCTION

A cure model is a special case of the mixture model, in which there are two groups of subjects: the susceptible group and the cure group. In the cure group, the subjects are immune from developing the endpoint. In the susceptible group, the subjects are susceptible to developing the endpoint and the time to event is the primary interest. All subjects in the cure group are censored since cure can never be observed. Cure models are particularly appropriate in cancer where patients can be cured by treatment and there is scientific interest in factors associated with the probability of cure and factors associated with the time to recurrence for non-cured individuals.

The work to date on cure models involves fixed baseline covariates. Farewell (1977) considered a logistic model for the mixture probability and parametric survival models for the failure time process. Yamaguchi (1992) applied a cure model with a logistic mixture probability model and an accelerated failure time model with generalized gamma distribution. Maller and Zhou (1996) studied the cure model extensively, specifically nonparametric failure time models for one sample and parametric failure time models. More recent work has focused on nonparametric failure time models. Taylor (1995) assumed a model with a logistic mixture probability and a completely unspecified failure time process, estimated by a Kaplan–Meier-type estimator. Kuk and Chen (1992), Sy and Taylor (2000), Peng and Dear (2000), considered a semi-parametric Cox proportional hazards model for the failure time process. Kuk and Chen (1992) used a Monte Carlo approximation of the marginal likelihood to estimate the regression parameters, and the EM algorithm to estimate the baseline hazard. Sy and Taylor (2000) and Peng and Dear (2000) obtained the MLEs of the parameters using EM algorithm. Chen and Ibrahim (2001) studied a non-mixture semi-parametric cure model. They implemented a Monte Carlo EM (MCEM) algorithm for maximum likelihood estimation.

One problem associated with the cure model is identifiability. This arises due to the lack of information at the end of the follow-up period, since a significant proportion of subjects are censored before the end of the follow-up period. The consequence of this problem is that it is hard to distinguish the effect of the covariates on the mixture probability and the failure time process and parameter estimates can be unstable. A potential solution to this problem is to include time-dependent information which may be associated with the failure time process.

The model we develop in this paper is motivated by studies of prostate cancer patients undergoing radiation therapy. The use of radiation is believed to cure a fraction of these patients, so fitting a cure model is appropriate for these data. In this field an important disease progression marker is prostate-specific antigen (PSA). A significant rise of PSA following treatment is believed to be an indicator of treatment failure, and clinical recurrence of the disease is expected to follow. Thus the longitudinal disease progression marker and the failure time process are modeled jointly, in a cure model setting. The model developed in this study is called the joint-cure model.

Joint modeling of a disease progression marker and a failure time process has increased in popularity recently (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Xu and Zeger, 2001; Wang and Taylor, 2001; Tsiatis and Davidian, 2001), especially in the modeling of CD4 counts and the time to AIDS for HIV positive patients. For example, Faucett and Thomas (1996) modeled the CD4 counts and the time to AIDS jointly, using a linear random-effects model for CD4 counts and a time-dependent proportional hazards model for the time to AIDS. They did the estimation from a Bayesian approach by Gibbs sampling. Wulfsohn and Tsiatis (1997) studied the same model as Faucett and Thomas, but they estimated the parameters by maximum likelihood using the MCEM algorithm. From these studies, the joint models were found to be able to reduce the bias and increase the efficiency of estimates compared with separate models. The research in the current paper both extends such joint models to a mixture model setting and extends cure models to include longitudinal data.

2. MODEL

2.1 *Notation*

Each individual in the study is associated with a set of data including the incidence data, the failure time data for the clinical event of interest (clinical recurrence), the longitudinal disease progression marker data, and the baseline covariates data. Let Z denote the fixed baseline covariates. The m post-treatment PSA measurements of an individual are denoted by the vector Y^*, and u is the corresponding measurement time vector. The observed follow-up time, denoted by t, is the minimum of the event time and the censoring time. The censoring indicator δ is equal to 1 if the event is observed, and is equal to 0 otherwise. Let D denote the cure group indicator, it is a partially observed latent variable. For a subject in the susceptible group, $D = 1$; otherwise, $D = 2$. Note that D can only be observed if the subject develops the endpoint, then $D = 1$.

We assume that censoring is random and non-informative (Kalbfleisch and Prentice, 1980), that the censoring distribution does not depend on the missing data, and that missingness in the data are ignorable (Rubin, 1976).

2.2 *Incidence model*

Assume the probability of an individual to be in the susceptible group is

$$P(D = 1; Z) = \frac{\exp(b^T Z)}{1 + \exp(b^T Z)}.$$

2.3 *Longitudinal model*

For the longitudinal data, we use a specific model designed to mimic the patterns of the post-treatment PSA. The response is given by

$$\log(Y^*(t) + 1) = \log(\psi^*(t) + 1) + e(t)$$

where $Y^*(t)$ is the function of the observed PSA value at time t, $\psi^*(t)$ is the corresponding function of the 'true' PSA, and $e(t) \sim_{\text{iid}} \quad N(0, \sigma_e^2)$ is measurement error. Let $Y = \log(Y^* + 1)$ be the vector of the transformed observed post-treatment PSA and $\psi(u) = \log(\psi^*(u) + 1)$ be the transformed 'true' PSA values.

The true PSA marker process $\psi^*(t)$ is modeled by a nonlinear exponential decay- exponential growth model (Kaplan *et al.*, 1991; Zagars and Pollack, 1993):

$$\psi^*(t) = r_1 \exp(-r_2 \times t) + r_3 \exp(r_4 \times t) \tag{1}$$

where r_1, r_2, r_3, and r_4 are the unobserved random effects associated with ψ^* (with r_1, r_2, r_3, and $r_4 > 0$). In this model r_2 characterizes the initial decline in PSA after radiation and r_4 characterizes the later possible increase.

We use different random-effects models for the subjects in the cure group and the susceptible group. A log-transformation on the random effects is used to ensure they are positive. For the random effects of subjects in the susceptible group, we assume $(R|D = 1; Z) \sim N(Z_1^T \mu_1, \Sigma_1)$, where R denotes the vector of the transformed random effects $\begin{pmatrix}\log r_1 & \log r_2 & \log r_3 & \log r_4\end{pmatrix}^T$. For the random effects of subjects in the cure group, $(R_{(4)}|D = 2; Z) \sim N(Z_2^T \mu_2, \Sigma_2)$ where $R_{(4)}$ is a subset of the vector R, $\begin{pmatrix}\log r_1 & \log r_2 & \log r_3\end{pmatrix}^T$ and $r_4 \mid (D = 2) \equiv 0$. The assumption for the cure group reduces the PSA

process to an exponential decay model. The terms $Z_1^T\mu_1$ and $Z_2^T\mu_2$ are the mean vectors of the random effects in the susceptible group and the cure group respectively, where $Z_1^T = (I_4 \otimes Z)^T$ and $Z_2^T = (I_3 \otimes Z)^T$, and Σ_1 and Σ_2 are the corresponding covariance matrices. Thus the baseline covariates are allowed to influence the trajectories of PSA in different ways in the two groups.

The probability density functions of R in the susceptible group and the cure group are given by

$$h_1(R|D=1;Z) = |2\pi\Sigma_1|^{-1/2}\exp\left\{-\tfrac{1}{2}(R - Z_1^T\mu_1)^T\Sigma_1^{-1}(R - Z_1^T\mu_1)\right\}$$
$$h_2(R|D=2;Z) = |2\pi\Sigma_2|^{-1/2}\exp\left\{-\tfrac{1}{2}(R_{(4)} - Z_2^T\mu_2)^T\Sigma_2^{-1}(R_{(4)} - Z_2^T\mu_2)\right\} \times \mathrm{I}(r_4 \equiv 0)$$

respectively.

The probability density function of Y conditional on the random effects is

$$g(Y \mid R;u) = g(Y \mid \psi;u) = \frac{1}{(2\pi\sigma_e^2)^{m/2}}\exp\left\{-\frac{1}{2\sigma_e^2}(Y - \psi(u))^T(Y - \psi(u))\right\}.$$

2.4 *Conditional failure time model*

For the failure time model in the susceptible group, conditional on the unobserved random effects, we assume a time-dependent proportional hazards model:

$$\lambda(t|D=1,R;Z) = \lambda_0(t)\exp(\gamma\psi(t) + \beta^T Z).$$

The hazard depends on the current 'true' PSA value and the baseline covariates. The baseline hazard function $\lambda_0(t)$ takes an unspecified form.

The conditional survival function and the conditional probability density function are

$$S(t \mid D=1,R;Z) = \exp\left\{-\int_0^t [\lambda_0(v)\exp(\gamma\psi(v) + \beta^T Z)]\,\mathrm{d}v\right\}$$

and $f(t \mid D=1,R;Z) = \lambda(t \mid D=1,R;Z)S(t \mid D=1,R;Z)$.

Note that the parameter vector β represents the direct effect of the baseline covariates on the relative hazard. The baseline covariates also have an indirect effect on the relative risk through the PSA marker process represented by γ.

3. ESTIMATION

3.1 *Parameter estimation*

Let Ω denote the set of parameters $(b, \mu_1, \mu_2, \Sigma_1, \Sigma_2, \sigma_e^2, \gamma, \beta, \lambda_0)$. Let $X_{\mathrm{obs}} = \{Z_i, Y_i, u_i, t_i, \delta_i, i = 1,\ldots,n\}$ denote the set of observed data, and $X_{\mathrm{miss}} = \{R_i, D_i, i = 1,\ldots,n\}$ denote the set of missing data.

The observed data likelihood can be written as

$$L(\Omega \mid X_{\text{obs}}) = \prod_{i=1}^{n}\Bigg[\Bigg\{P(D_i = 1; Z_i)\int (f(t_i \mid D_i = 1, R_i; Z_i) \times g(Y_i \mid R_i; u_i)h_1(R_i \mid D_i = 1; Z_i))\,\mathrm{d}R_i\Bigg\}^{\mathrm{I}(\delta_i=1)} \times\Bigg\{P(D_i = 1; Z_i)\int (S(t_i \mid D_i = 1, R_i; Z_i)g(Y_i \mid R_i; u_i)h_1(R_i \mid D_i = 1; Z_i))\,\mathrm{d}R_i + P(D_i = 2; Z_i)\int (g(Y_i \mid R_i; u_i)h_2(R_i \mid D_i = 2; Z_i))\,\mathrm{d}R_i\Bigg\}^{\mathrm{I}(\delta_i=0)}\Bigg].$$

Because direct maximization of the observed data likelihood is complicated, we use the EM algorithm to obtain parameter estimates. The complete-data likelihood is the joint distribution of (Y, t, D, R), given by

$$L(\Omega \mid X_{\text{obs}}, X_{\text{miss}}) = \prod_{i=1}^{n} L_i$$

where

$$\begin{aligned} L_i = {} & \{P(D_i = 1; Z_i)\, f(t_i \mid D_i = 1, R_i; Z_i)g(Y_i \mid R_i; u_i)h_1(R_i \mid D_i = 1; Z_i)\}^{\mathrm{I}(\delta_i=1)\mathrm{I}(D_i=1)} \\ & \times\{P(D_i = 1; Z_i)S(t_i \mid D_i = 1, R_i; Z_i)g(Y_i \mid R_i; u_i)h_1(R_i \mid D_i = 1; Z_i)\}^{\mathrm{I}(\delta_i=0)\mathrm{I}(D_i=1)} \\ & \times\{P(D_i = 2; Z_i)g(Y_i \mid R_i; u_i)h_2(R_i \mid D_i = 2; Z_i)\}^{\mathrm{I}(\delta_i=0)\mathrm{I}(D_i=2)}. \end{aligned}$$

It can be rewritten as the product of five components, each of which involves a distinct set of parameters in Ω.

The expected complete-data log-likelihood is given by

$$E[\log L(\Omega \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] = \sum_{i=1}^{n}\int [\log L_i \times p(X_{\text{miss},i} \mid X_{\text{obs}}, \Omega^{(m)})]\,\mathrm{d}X_{\text{miss},i} \qquad (2)$$

For simplicity of notation, let P_{1i} denote $P(D_i = 1 \mid Z_i)$, g_i denote $g(Y_i \mid R_i; u_i)$, h_{1i} denote $h_1(R_i \mid D_i = 1; Z_i)$, h_{2i} denote $h_2(R_i \mid D_i = 2; Z_i)$, S_i denote $S(t_i \mid D_i = 1, R_i, Z_i)$, and f_i denote $f(t_i \mid D_i = 1, R_i, Z_i)$, and $P_{1i}^{(m)}$, $g_i^{(m)}$, $h_{1i}^{(m)}$, $h_{2i}^{(m)}$, $S_i^{(m)}$, and $f_i^{(m)}$ denote the corresponding functions substituted with the current parameter values $\Omega^{(m)}$ at the mth EM step.

For subjects who are observed to develop the endpoint, we know that $D_i = 1$, and the probability density function of R_i is given by

$$p(R_i \mid X_{\text{obs}}, \Omega^{(m)}) = \frac{f_i^{(m)} g_i^{(m)} h_{1i}^{(m)}}{\int \{f_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\}\,\mathrm{d}R_i} \qquad (3)$$

For a censored subject, the joint conditional distributions of D_i and R_i are given by

$$p(D_i = 1, R_i \mid X_{\text{obs}}, \Omega^{(m)}) = \frac{P_{1i}^{(m)} S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}}{\int \{P_{1i}^{(m)} S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\}\,\mathrm{d}R_i + \int \{(1 - P_{1i}^{(m)}) g_i^{(m)} h_{2i}^{(m)}\}\,\mathrm{d}R_i}$$

and

$$p(D_i = 2, R_i \mid X_{\text{obs}}, \Omega^{(m)}) = \frac{(1 - P_{1i}^{(m)}) g_i^{(m)} h_{2i}^{(m)}}{\int \{P_{1i}^{(m)} S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\} \, \mathrm{d}R_i + \int \{(1 - P_{1i}^{(m)}) g_i^{(m)} h_{2i}^{(m)}\} \, \mathrm{d}R_i}$$

We can rewrite the above joint distributions of D_i and R_i in the following form:

$$p(D_i = 1, R_i \mid X_{\text{obs}}, \Omega^{(m)}) = p(R_i \mid D_i = 1, X_{\text{obs}}, \Omega^{(m)}) \times p(D_i = 1 \mid X_{\text{obs}}, \Omega^{(m)})$$

and

$$p(D_i = 2, R_i \mid X_{\text{obs}}, \Omega^{(m)}) = p(R_i \mid D_i = 2, X_{\text{obs}}, \Omega^{(m)}) \times p(D_i = 2 \mid X_{\text{obs}}, \Omega^{(m)})$$

where

$$p(R_i \mid D_i = 1, X_{\text{obs}}, \Omega^{(m)}) = \frac{S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}}{\int \{S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\} \mathrm{d}R_i} \tag{4}$$

$$p(R_i \mid D_i = 2, X_{\text{obs}}, \Omega^{(m)}) = \frac{g_i^{(m)} h_{2i}^{(m)}}{\int \{g_i^{(m)} h_{2i}^{(m)}\} \mathrm{d}R_i} \tag{5}$$

$$p(D_i = 1 \mid X_{\text{obs}}, \Omega^{(m)}) = \frac{\int \{P_{1i}^{(m)} S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\} \mathrm{d}R_i}{\int \{P_{1i}^{(m)} S_i^{(m)} g_i^{(m)} h_{1i}^{(m)}\} \mathrm{d}R_i + \int \{(1 - P_{1i}^{(m)}) g_i^{(m)} h_{2i}^{(m)}\} \mathrm{d}R_i} \tag{6}$$

$$p(D_i = 2 \mid X_{\text{obs}}, \Omega^{(m)}) = 1 - p(D_i = 1 \mid X_{\text{obs}}, \Omega^{(m)}). \tag{7}$$

The conditional expectation of $\log L_i$ in equation (2) can be computed by

$$E[\log L_i \mid X_{\text{obs}}, \Omega^{(m)}] = \int [\log L_i \times p(R_i \mid X_{\text{obs}}, \Omega^{(m)})] \, \mathrm{d}R_i \tag{8}$$

for an event, or

$$E[\log L_i \mid X_{\text{obs}}, \Omega^{(m)}] = \sum_{d=1}^{2} \left\{ \int [\log L_i \times p(R_i \mid D_i = d, X_{\text{obs}}, \Omega^{(m)})] \, \mathrm{d}R_i \times p(D_i = d \mid X_{\text{obs}}, \Omega^{(m)}) \right\} \tag{9}$$

otherwise. The expectations in the above expressions are approximated by Monte Carlo integration using importance sampling.

The conditional probability of D_i given by (6), when evaluated at the MLE, can be used to estimate the probability that a censored subject is in the susceptible group. This probability will be referred as the conditional probability of recurrence hereafter.

In the M-step, the expected complete-data log-likelihood (denoted by $Elc^{(m)}$) can be expressed as

$$\begin{aligned} Elc^{(m)} &= E[\log L(\Omega \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] \\ &= Elc_1^{(m)}(b) + Elc_2^{(m)}(\sigma_e^2) + Elc_3^{(m)}(\mu_1, \Sigma_1) + Elc_4^{(m)}(\mu_2, \Sigma_2) + Elc_5^{(m)}(\lambda_0, \gamma, \beta) \end{aligned}$$

where

$$Elc_1^{(m)}(b) = E[\log L(b \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] \tag{10}$$

$$Elc_2^{(m)}(\sigma_e^2) = E[\log L(\sigma_e^2 \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] \tag{11}$$

$$Elc_3^{(m)}(\mu_1, \Sigma_1) = E[\log L(\mu_1, \Sigma_1 \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] \tag{12}$$

$$Elc_4^{(m)}(\mu_2, \Sigma_2) = E[\log L(\mu_2, \Sigma_2 \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}] \tag{13}$$

$$Elc_5^{(m)}(\lambda_0, \gamma, \beta) = E[\log L(\lambda_0, \gamma, \beta \mid X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \Omega^{(m)}]. \tag{14}$$

The functions $Elc_1^{(m)}$, $Elc_2^{(m)}$, $Elc_3^{(m)}$, $Elc_4^{(m)}$, and $Elc_5^{(m)}$ involve distinct parameters in Ω. Thus we can maximize the expected complete-data log-likelihood by maximizing the five terms separately. For b in $Elc_1^{(m)}$, (μ_1, Σ_1) in $Elc_3^{(m)}$, and (μ_2, Σ_2) in $Elc_4^{(m)}$, we use Newton–Raphson to obtain the estimates at each M-step. For σ_e^2 in $Elc_2^{(m)}$, a closed form algebraic solution exists. For $(\gamma, \beta, \lambda_0)$ in $Elc_5^{(m)}$, we use the profile likelihood approach proposed by Johansen (1983). Details of the parameter estimation procedure can be found in Law (2000). The methods of parameter estimation are extensions of the work of Wulfsohn and Tsiatis (1997), Sy and Taylor (2000) and Henderson *et al.* (2000).

3.2 *Implementation details*

In the E-step, we use the Monte Carlo integration to evaluate the required integrals. Wei and Tanner (1990) first introduced the MCEM method by executing the E-step with a Monte Carlo process. Importance sampling is used here to implement the Monte Carlo E-step.

For an event, the conditional expectation of the complete-data log-likelihood (given in (8)) can be approximated by

$$\int [\log L_i \times p(R_i \mid X_{\text{obs}}, \Omega^{(m)})]\mathrm{d}R_i \approx \frac{1}{K}\sum_{k=1}^{K} \log L_{ik} \times \frac{p(R_{ik}^* \mid X_{\text{obs}}, \Omega^{(m)})}{p_i^*(R_{ik}^* \mid X_{\text{obs}}, \Omega^{(m)})}$$

where $p_i^*(R \mid X_{\text{obs}}, \Omega^{(m)})$ is the importance sampling distribution for the random effects of subject i conditional on the observed data and current parameter values. The Monte Carlo sample $R_{ik}^*, k = 1, \ldots, K$, are generated from $p_i^*(R \mid X_{\text{obs}}, \Omega^{(m)})$. The term $\log L_{ik}$ is the function of $\log L_i$ substituted with $R_{ik}^*, k = 1, \ldots, K$.

Similarly, the conditional expectation of the complete-data log-likelihood for a censored subject (given in equation (9)) can be approximated by

$$\sum_{d=1}^{2}\left\{\int [\log L_i \times p(R_i \mid D_i = d, X_{\text{obs}}, \Omega^{(m)})]dR_i \times p(D_i = d \mid X_{\text{obs}}, \Omega^{(m)})\right\}$$
$$\approx \sum_{d=1}^{2}\left\{\frac{1}{K}\sum_{k=1}^{K}\left[\log L_{ik} \times \frac{p(R_{ik}^* \mid D_i = d, X_{\text{obs}}, \Omega^{(m)})}{p_i^*(R_{ik}^* \mid D_i = d, X_{\text{obs}}, \Omega^{(m)})}\right] \times p(D_i = d \mid X_{\text{obs}}, \Omega^{(m)})\right\}$$

where $p_i^*(R \mid D_i = 1, X_{\text{obs}}, \Omega^{(m)})$ and $p_i^*(R \mid D_i = 2, X_{\text{obs}}, \Omega^{(m)})$ are the importance sampling distributions for the random effects of the censored subject i, conditional on that the subject being in the cure group or in the susceptible group respectively. The Monte Carlo sample $R_{ik}^*, k = 1, \ldots, K$, are generated from $p_i^*(R \mid D_i = 1, X_{\text{obs}}, \Omega^{(m)})$ or $p_i^*(R \mid D_i = 2, X_{\text{obs}}, \Omega^{(m)})$.

The precision of the Monte Carlo approximations depends on the choice of the importance sampling distribution and the Monte Carlo sample size K. In this study, a multivariate Student's t distribution is

used for the importance sampling. The mean and covariance of the importance sampling distribution can be found by a Laplace-type approximation. Let $t_i(R)$ denote the conditional distribution of the log of the random effects for subject i (given by (3), (4), or (5)). The mean vector μ_{ti} is the maximizer of $t_i(R)$, obtained by solving the equation $\frac{\partial}{\partial R} t_i(R) = 0$. The Laplace approximation of the covariance matrix is given by $-[\frac{\partial^2}{\partial R^2} t_i(R)]^{-1} \mid_{R=\mu_{ti}}$. Note that the denominators of the conditional distributions also involve intractable integrals of functions over the random effects. These integrals can be approximated similarly using Monte Carlo integration.

For the Monte Carlo sample size K we follow the suggestion of Wei and Tanner (1990): use a small K at early EM steps, and increase K along with the number of iterations. In a generalized linear model setting, Booth and Hobert (1999) developed an automation scheme for varying the Monte Carlo sample size and monitoring the convergence of the MCEM algorithm. For the complicated joint-cure model in this study, we adopt a more subjective way of increasing the Monte Carlo sample size K with the EM steps. Graphical methods are used to monitor the convergence of the MCEM algorithm.

3.3 *Standard error estimation*

Sy and Taylor (2001) suggested three methods to compute the standard errors for the parameter estimates in a cure model involving a logistic incidence link and a proportional hazards model with time-fixed covariates. All the three methods involve the inverse of an information matrix derived from the observed likelihood. The first method treats λ_0 as a vector of parameters. Sy and Taylor (2001) observed that this method gave satisfactory standard errors. We use the same approach with a slight modification to include the extra parameters in the longitudinal models.

The observed data information is not a by-product of the EM step. Louis (1982) showed that the observed data information can be expressed as the difference between the 'complete-data information' and 'the missing data information', i.e.

$$-\frac{\partial^2}{\partial \Omega^2} \log L(\Omega; X_{\text{obs}}) = -\frac{\partial^2}{\partial \Omega^2} E[\log L(\Omega; X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \hat{\Omega}] - \text{Var}\left(\frac{\partial}{\partial \Omega} \log L(\Omega; X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \hat{\Omega}\right). \tag{15}$$

The first term of the right-hand side of (15) is the complete-data information I_c, and the second term represents the missing data information.

The complete-data information has a block diagonal form with five blocks, with expressions involving the expectation conditional on the final MLE $\hat{\Omega}$. Monte Carlo integration discussed in the previous section is performed to approximate the expectation.

It can be shown that the second term in (15) can be written as

$$\text{Var}\left(\frac{\partial}{\partial \Omega} \log L(\Omega; X_{\text{obs}}, X_{\text{miss}}) \mid X_{\text{obs}}, \hat{\Omega}\right) = \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial \Omega} \log L_i\right)^2 \mid X_{\text{obs}}, \hat{\Omega}\right] - \sum_{i=1}^{n} \left(E\left[\frac{\partial}{\partial \Omega} \log L_i \mid X_{\text{obs}}, \hat{\Omega}\right]\right)^2$$

where L_i is the complete-data likelihood of subject i. The second term of the above equation equals zero at MLE. The expectation of the first term can be approximated by Monte Carlo integration.

4. PROSTATE CANCER STUDY

4.1 *Data description*

The model is fitted to a data set of 458 prostate cancer patients who had received radiation therapy. The clinical recurrence endpoint (local recurrence or distant metastasis) was observed in 92 patients. The baseline covariates considered in this study are the baseline PSA (bPSA), T stage (categorical: T1, T2, T3/4), Gleason Score (GS). Post-treatment PSA were measured about every 6 months. A total of 4226 post-treatment PSA values were measured, the median number of PSA measurement for each patient was 9 (ranged from 1 to 26). Most patients in this study had T stage II (67%), GS of 6 (25%) or 7 (31%). The mean baseline PSA value is 22.

4.2 *Model specifications*

We fit two models to the data. The first model is the joint-cure model with a logistic link for the probability of cure, a Cox regression model with both time-fixed and time-dependent (the true current PSA value) covariates for the relative risk of recurrence of the susceptible, and a hierarchical nonlinear mixed-effects model for the post-treatment PSA measurements. The second model is a logistic-Cox cure model (Sy and Taylor, 2000) with a logistic link for the probability of cure, a Cox proportional hazards model with time-fixed covariates for the relative risk of recurrence of the susceptible, and no longitudinal PSA component. Details of the methods of parameter estimation for the second model can be found in Sy and Taylor (2000).

In the Monte Carlo E step, a multivariate Student's t distribution with 10 degree of freedom is used for the importance sampling. We start with a Monte Carlo sample size K of 200, and increase it gradually to the final sample size of 3000. The final estimate for each of the parameters is the average of the estimates from the last 10 EM iterations. Computations are performed using Matlab.

4.3 *Results of fits*

Table 1 gives estimates of the regression parameters for the two models parameters. Estimates of the remaining parameters are given in Law (2000).

The joint-cure model and the logistic-Cox model give similar inference for most of the parameters except T stage in the incidence component and baseline PSA in the failure time model. In the incidence component of the joint-cure model, all the baseline covariates in the study have significant effects on the probability of recurrence. The estimates are consistent with intuition. All of the baseline covariates are also significant in the conditional failure time model, except that baseline PSA has a negative sign rather than the expected positive sign. This implies that patients with higher baseline PSA have a higher recurrence probability, but the recurrence will occur later. This initially surprising result may be due to the high correlation between the baseline PSA and the current PSA in the conditional failure time model. The current PSA value is highly significant in the model, therefore the higher the current PSA value, the higher the risk of recurrence is. Furthermore, for a given current PSA value at a particular time, someone with a low baseline PSA, would have a faster average rise in PSA, and thus could have a higher risk than someone who started with a high baseline value. For all patients (whether they are in the cure group or the susceptible group), baseline PSA has the most significant effects on r_1 and r_3. The rate of decline (r_2) of the post-treatment PSA is not affected by any of the baseline covariates studied for patients in both groups. In the susceptible group, the rate of rise (r_4) after the initial decline is significantly determined by both T stage and GS, but not by the baseline PSA.

Figure 1 shows the estimated profiles of PSA for selected censored subjects. The PSA profiles for subject i are estimated by substituting the means of $P(R_i \mid D_i = 1, X_{\text{obs}}, \hat{\Omega})$ and $P(R_i \mid D_i = 2, X_{\text{obs}}, \hat{\Omega})$ into (1). Thus there are two estimated PSA profiles conditional on the subject being in the cure group

Table 1. *Results of Fit. (a) Incidence model and conditional failure time model (b) Longitudinal model (joint-cure model)—susceptible group (c) Longitudinal model (joint-cure model)—cure group*

(a)

		Joint-cure model			**Logistic-Cox model**		
Parameter		Estimate	S.E.	*p*-value	Estimate	S.E.	*p*-value
$\boldsymbol{b}$							
	intercept	1.06	0.48	0.03	−0.29	0.45	0.52
	T stage			<0.01			0.46
	T1	−1.79	0.62	<0.01	−0.82	1.17	0.48
	T2	−1.27	0.50	0.01	−0.63	0.52	0.23
	bPSA	1.19	0.20	<0.01	0.82	0.23	<0.01
	GS	0.25	0.11	0.03	0.34	0.18	0.06
β							
	T stage			<0.01			<0.01
	T1	−1.82	0.49	<0.01	−2.17	0.71	<0.01
	T2	−0.84	0.25	<0.01	−1.08	0.26	<0.01
	bPSA	−0.55	0.14	<0.01	0.13	0.17	0.42
	GS	0.27	0.09	<0.01	0.36	0.14	<0.01
γ							
	$\psi(t)$	1.07	0.09	<0.01	–	–	–

(b)

Parameter	Estimate	S.E.	*p*-value
μ_1 r_1			
intercept	2.554	0.075	<0.01
T stage			0.87
T1	−0.014	0.112	0.90
T2	0.029	0.075	0.70
bPSA	0.892	0.038	<0.01
GS	−0.015	0.023	0.51
μ_1 r_2			
intercept	−1.188	0.112	<0.01
T Stage			0.89
T1	0.078	0.160	0.63
T2	0.039	0.110	0.72
bPSA	0.027	0.056	0.63
GS	−0.033	0.033	0.32
μ_1 r_3			
intercept	−1.005	0.291	<0.01
T Stage			0.22
T1	0.412	0.421	0.33
T2	−0.224	0.299	0.46
bPSA	0.633	0.148	<0.01
GS	−0.168	0.089	0.06
μ_1 r_4			
intercept	−2.416	0.168	<0.01
T Stage			<0.01
T1	−1.002	0.264	<0.01
T2	−0.488	0.171	<0.01
bPSA	0.037	0.089	0.68
GS	0.212	0.053	<0.01

(c)

Parameter	Estimate	S.E.	*p*-value
μ_2 r_1			
intercept	2.830	0.300	<0.01
T stage			<0.01
T1	−0.681	0.316	0.03
T2	−0.374	0.306	0.22
bPSA	0.863	0.063	<0.01
GS	0.059	0.035	0.10
μ_2 r_2			
intercept	−0.834	0.272	<0.01
T stage			0.72
T1	−0.244	0.302	0.42
T2	−0.192	0.278	0.49
bPSA	0.073	0.077	0.34
GS	0.011	0.040	0.78
μ_2 r_3			
intercept	−0.605	0.224	<0.01
T stage			0.13
T1	0.293	0.248	0.24
T2	0.419	0.234	0.07
bPSA	0.399	0.062	<0.01
GS	−0.012	0.035	0.72

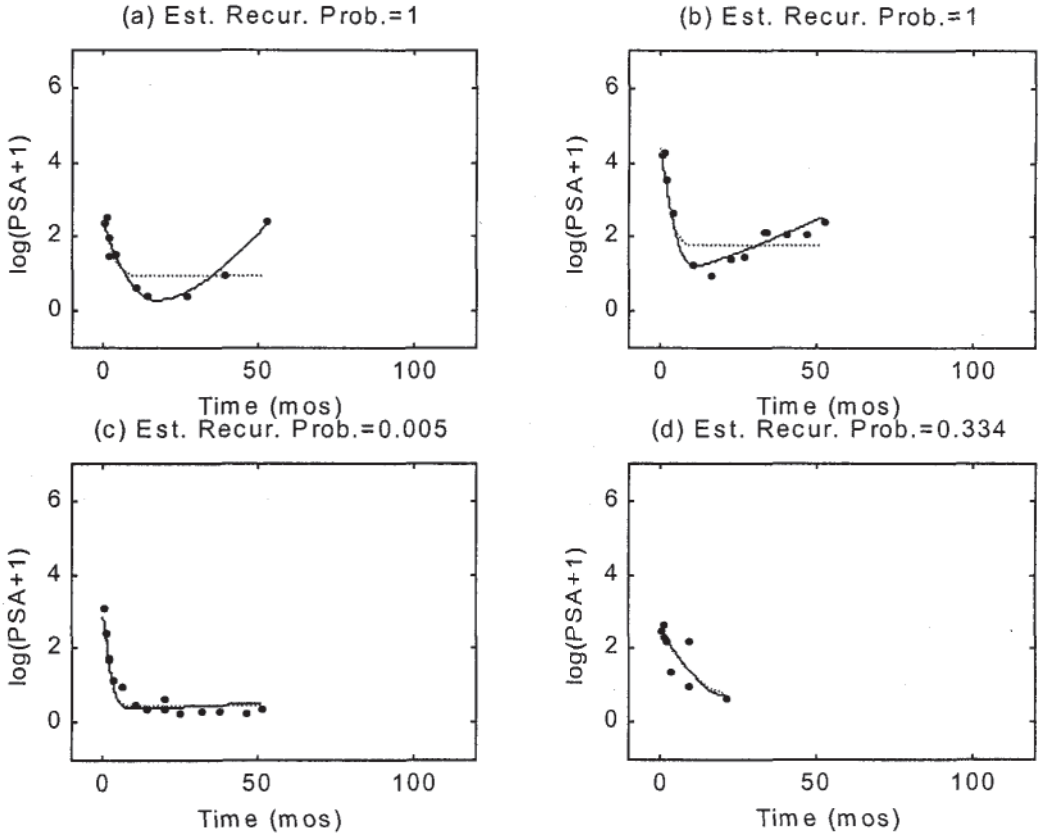

Fig. 1. Plots of individual estimated PSA profiles for selected censored subjects. The solid curve is the PSA profile conditional on the subject being in the susceptible group, and the dotted curve is the profile conditional on the cure group. The observed PSA values are denoted by ●.

or the susceptible group. The final estimates of the conditional recurrence probability are given for the subjects. The plots (a) and (b) show that the susceptible group PSA profile fits better than the cure group PSA profile. The estimated recurrence probabilities using the joint-cure model for these individuals are 1. Based on the logistic-Cox model, the estimated recurrence probabilities for (a) and (b) are 0.232 and 0.681 respectively. These later estimates are based primarily on the baseline covariates and did not take into account the rising patterns in the post-treatment PSA profiles. The fact that the $D = 2$ fits so poorly in Figures 1(a) and (b) is not surprising, since PSA can only decrease in that group. Thus the assumptions in the longitudinal model are having a substantial influence on the assignment of these two individuals to the non-cured groups.

The estimated probabilities are lower for the two subjects of plots (c) and (d), and the corresponding plots show that both the susceptible group profile and the cure group profile fit well. Based on the leveling off pattern of the post-treatment PSA profile, the subject in (c) is very likely to be from the cure group. The susceptible group PSA profile fits well due to the flexibility of the exponential decay-exponential growth model. However, such a small rate of rise of the PSA profile means that the subject is very unlikely to be from the susceptible group. However, the logistic-Cox model estimated an 0.511 recurrence

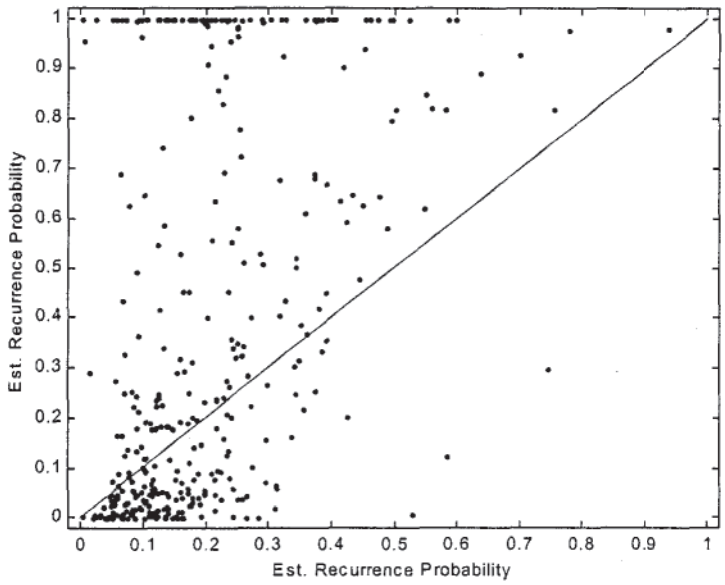

Fig. 2. Plots of estimated conditional recurrence probability from the joint-cure model (y-axis) versus the logistic-Cox model (x-axis).

probability for this individual. The subject in plot (d) has only a few PSA measurements. It is difficult to classify the subject into the cure group or the susceptible group based on the observed PSA profile. Thus the estimated recurrence probability of 0.334 is based mostly on the baseline covariates. The estimated recurrence probability by the logistic-Cox model is 0.513.

Figure 2 shows the plot of the estimated conditional recurrence probability from the joint-cure model versus the logistic-Cox model. There is little correlation between the estimates from the two models. The logistic-Cox model tends to give lower estimates than the joint-cure model, and tends not to assign probabilities near zero or one.

Figure 3 shows the residual plots for the longitudinal model. The jth residual of the ith subject is defined as

$$\text{residual}_{ij} = \{Y_{ij} - \hat{\psi}_i(u_{ij} \mid D_i = 1)\} \times \hat{P}(D_i = 1) + \{Y_{ij} - \hat{\psi}_i(u_{ij} \mid D_i = 2)\} \times \hat{P}(D_i = 2)$$

where $\hat{P}(D_i = 1)$ and $\hat{P}(D_i = 2)$ are estimated by (6) and (7), and $\hat{\psi}_i(u_{ij} \mid D_i = 1)$ and $\hat{\psi}_i(u_{ij} \mid D_i = 2)$ are the estimated PSA profiles for subject i. The residual plots show that the nonlinear hierarchical mixed effects model fit the longitudinal PSA data well.

5. A SIMULATION STUDY

We performed a small simulation study to compare the performance of the joint-cure model and the logistic-Cox model in the presence of the longitudinal disease progression marker, in particular the fits of the models, and the classification of the censored subjects into the cure/susceptible groups. We consider a design and parameter values which mimic those obtained from the prostate cancer data set; 50 replications are done.

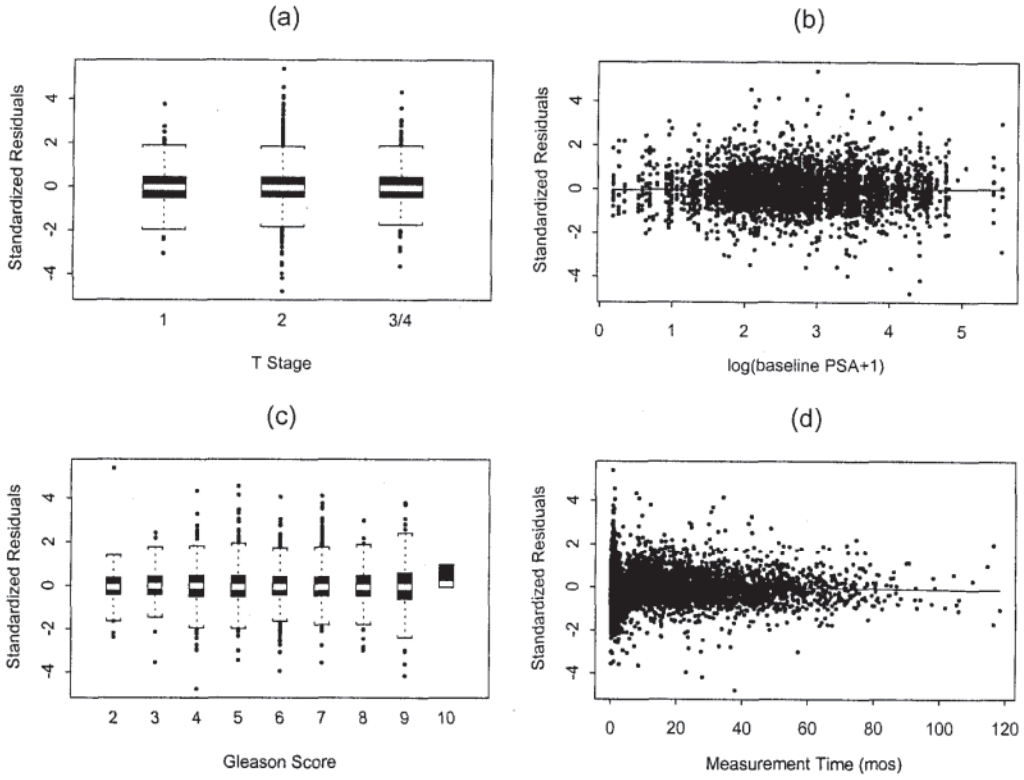

Fig. 3. Residual plots of the longitudinal model in the joint-cure model.

5.1 *Simulation designs*

The data are simulated from the joint-cure model as described in previous sections. A single continuous baseline covariate is considered in the model, which is assumed to influence the incidence probability and the conditional failure time model, but not the longitudinal model. True values for the regression parameters are given in Table 2. The baseline hazard $\lambda_0(t)$ is constant with $\lambda_0(t) = 1/100$. The censoring time follows an exponential distribution. The maximum follow-up time is 120 months. The post-treatment PSA value is measured every 6 months for each individual. The baseline covariate is generated from a uniform distribution between –0.5 and 0.5 with 25 grid points. 500 subjects are simulated in each replication.

5.2 *Simulation results*

The median proportion of cured subjects is 56.4%. Among the subjects in the susceptible group, the median censoring rate is 43.9%. There are 13.2% of subjects censored after the maximum observed event time. The median number of post-treatment PSA measurements is 7.

Table 2 gives the bias, the variance, and the MSE of $\hat{b}_0$, $\hat{b}_1$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\mu}_1$, and $\hat{\mu}_2$. When compared with the logistic-Cox model, the joint-cure model gives less bias (except for $\hat{b}_1$), variance, and smaller MSE for $\hat{b}_0$, $\hat{b}_1$, $\hat{\beta}$.

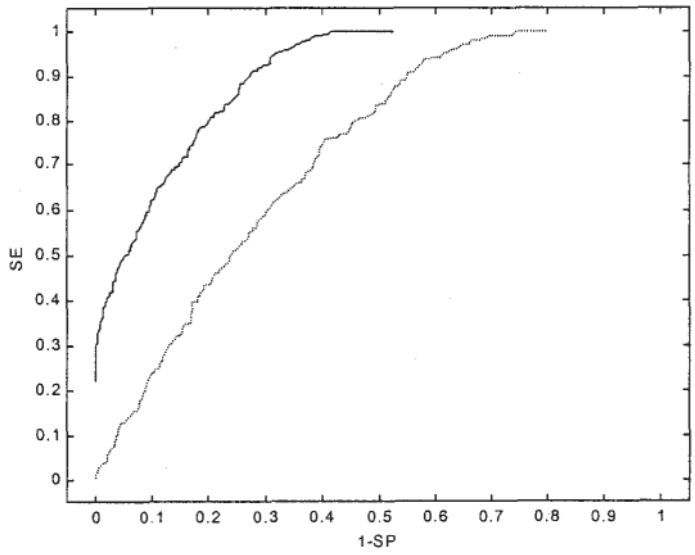

Fig. 4. Simulation study: median receiver operating characteristics curves (sensitivity versus 1-specificity) of the joint-cure model (solid curve) and the logistic-Cox model (dotted curve).

Table 2. *Simulation results: bias, variance, MSE*

	True value	Bias	Variance	MSE
Joint-cure model				
$\hat{b}_0$	−0.1999	0.0050	0.0097	0.0097
$\hat{b}_1$	1.3261	0.1173	0.1410	0.1548
$\hat{\beta}$	1.3802	0.0286	0.1432	0.1440
$\hat{\gamma}$	1.0657	0.0570	0.0807	0.0839
$\hat{\mu}_1$				
$\log(r_1)$	2.5730	0.0009	0.0001	0.0001
$\log(r_2)$	1.1489	0.0015	0.0002	0.0002
$\log(r_3)$	−1.2283	0.0028	0.0005	0.0005
$\log(r_4)$	−2.9038	−0.0026	0.0004	0.0004
$\hat{\mu}_2$				
$\log(r_1)$	2.4561	−0.0022	0.0002	0.0002
$\log(r_2)$	−1.0258	−0.0008	0.0002	0.0002
$\log(r_3)$	−0.1854	0.0014	0.0002	0.0002
Logistic-cox model				
$\hat{b}_0$	−0.1999	−0.0151	0.0123	0.0125
$\hat{b}_1$	1.3261	0.1151	0.2903	0.3035
$\hat{\beta}$	1.3802	−0.0968	0.2397	0.2490

Table 3. *Simulation results: bias, asymptotic standard error, empirical standard deviation*

	Asymptotic standard error	Empirical standard deviation
$\hat{b}_0$	0.116	0.100
$\hat{b}_1$	0.419	0.379
$\hat{\beta}$	0.375	0.382
$\hat{\gamma}$	0.271	0.287
$\hat{\mu}_1$		
$\log(r_1)$	0.018	0.012
$\log(r_2)$	0.020	0.016
$\log(r_3)$	0.041	0.022
$\log(r_4)$	0.027	0.021
$\hat{\mu}_2$		
$\log(r_1)$	0.017	0.015
$\log(r_2)$	0.018	0.015
$\log(r_3)$	0.013	0.013

The empirical standard deviation and the mean of the asymptotic standard errors from the 50 replications under the joint-cure model are compared. The empirical standard deviation is computed as the sample standard deviation of the parameter estimates over the 50 replications. The results for the regression parameters of the joint-cure model are given in Table 3. Overall the values are similar, except for μ_1 and μ_2, where the asymptotic standard errors appear to be too large in some cases.

The group classification performance of the conditional recurrence probability $P(D_i = 1 \mid X_{\text{obs}}, \hat{\Omega})$ by the joint-cure model and the logistic-Cox model is compared. At each pre-specified cut-off value w^*, varying from 0 to 1 at intervals of 0.001, the censored subjects are classified into the cure group if the conditional recurrence probability is smaller than w^* or the susceptible group otherwise. We found that the joint-cure model always has higher sensitivity and lower misclassification probability than the logistic-Cox model. The specificity of the joint-cure model is generally higher at low cut-off values, but is lower around the medium cut-off. Figure 4 gives the median receiver operating characteristics (ROC) curve. On average, the joint-cure model is better than the logistic-Cox model in terms of the classification performance as evaluated by the ROC curve.

6. Discussion

We see three major advantages of adding a longitudinal component to the cure model: one is to reduce the bias due to informative censoring, a second is to improve individual predictions and a third is to reduce the well known identifiability problems in a cure model.

Informative censoring is undoubtedly present in many prostate cancer studies. Clinicians may decide to start salvage therapy in some patients when they observe a sharp rise of post-treatment PSA in those patients before any clinical event is observed. The usual salvage therapy is hormonal treatment which rapidly reduces the PSA to very low levels. We censored the patients at this time point and thus they are informatively censored. By including the longitudinal PSA measurements in the joint-cure model, we hope to approximately achieve a situation of missing at random missingness mechanism (Rubin, 1976) and thus eliminate the informative censoring problem. If the longitudinal PSA data were not included in the analysis we would be in a situation of non-ignorable missingness.

Under the model setting in this study, the only aspects of the PSA process we consider in the conditional failure time model is the current PSA value. Other assumptions are possible. For example, the rate of decline of the post-treatment PSA following radiation therapy, or the rate of rise of the PSA after the initial decline, can be included into the failure time model. Further research can also be done in the use of parametric failure time models, different longitudinal models, refinement of the MCEM algorithm, or the use of Gibbs sampling in the estimation procedure.

An important aspect of using complex models is to assess whether they fit the data. We have used some graphical schemes, but clearly some more formal methods would be useful. Despite the biological rationale for including a cured fraction, it would be interesting to see if a simpler joint longitudinal-Cox model, without a cured fraction, gave a good fit to the data. Although the longitudinal-Cox model is nested in the joint-cure model, the nesting is non-standard and it is unclear whether standard likelihood ratio tests would work.

The focus of the model developed in this paper is on describing the process which gives rise to the observations and describing patterns in the data. The model can have a variety of uses depending on the scientific goals. For example, one might be interested in which covariates are associated with the probability of cure. The model could also be used to address issues of surrogacy of a baseline treatment variable, specifically by examining whether this baseline variable affects the hazard of the clinical event conditional on the longitudinal measurements. We have presented individual estimates of the probability of cure for a censored subject, an estimate of the uncertainty of that probability would also be a useful measure. From a clinical perspective one might want to make individual predictions about the distribution of future event times or future values of the longitudinal variable conditional on the observed longitudinal data.

REFERENCES

BOOTH, J. G. AND HOBERT, J. H. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society* B **62**, 265–285.

CHEN, M. H. AND IBRAHIM, J. G. (2001). Maximum likelihood methods for cure rate models with missing covariates. *Biometrics* **57**, 43–52.

FAREWELL, V. T. (1977). A model for a binary variable with time censored observations. *Biometrika* **64**, 43–46.

FAUCETT, C. L. AND THOMAS, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

HENDERSON, R., DIGGLE, P. AND DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

JOHANSEN, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51**, 165–174.

KALBFLEISCH, J. D. AND PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

KAPLAN, I. D., COX, R. S. AND BAGSHAW, M. A. (1991). A model of prostatic carcinoma tumor kinetics based on prostate specific antigen levels after radiation therapy. *Cancer* **68**, 400–404.

KUK, A. Y. C. AND CHEN, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.

LAW, N. J. (2000). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure, Unpublished Ph.D. dissertation, University of California, Los Angeles.

LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society* B **44**, 226–233.

MALLER, R. AND ZHOU, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley.

PENG, Y. AND DEAR, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56**, 237–243.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

SY, J. P. AND TAYLOR, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227–236.

SY, J. P. AND TAYLOR, J. M. G. (2001). Standard errors for the Cox proportional hazards cure model. Modelling and Data Analysis in Cancer Studies. *Mathematical and Computer Modelling* **33**, 1237–1251.

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

TSIATIS, A. A. AND DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrica* **88**, 447–458.

WANG, Y. AND TAYLOR, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 895–905.

WEI, G. C. G. AND TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.

WULFSOHN, M. S. AND TSIATIS, A. A. (1997). Joint models for survival and longitudinal data. *Biometrics* **53**, 330–339.

XU, J. AND ZEGER, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

YAMAGUCHI, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in Japan. *Journal of the American Statistical Association* **87**, 284–292.

ZAGARS, G. K. AND POLLACK, A. (1993). The fall and rise of prostate-specific antigen. *Cancer* **72**, 832–842.

[*Received December 12, 2000; revised October 30, 2001; accepted for publication December 12, 2001*]