Statistica Sinica **14**(2004), 835-862

# JOINT LONGITUDINAL-SURVIVAL-CURE MODELS AND THEIR APPLICATION TO PROSTATE CANCER

Menggang Yu[1], Ngayee J. Law[2], Jeremy M. G. Taylor[1,3] and Howard M. Sandler[3]

[1]*Indiana University,* [2]*Roche Molecular Systems, Inc. and* [3]*University of Michigan*

*Abstract:* Many scientific investigations generate both longitudinal data and survival data. Methods for the combined analysis of both kinds of data have been developed in recent years, with the main emphasis being on modeling and estimation. In cancer research it is common for there to be long term survivors or cured patients and methods have been developed to analyze such data. In this article, we review both joint models for the analysis of longitudinal and survival data and cure models. We then present a joint longitudinal-survival-cure model to analyze data from a study of prostate cancer patients treated with radiation therapy. In this model each patient is assumed to be either cured or susceptible to clinical recurrence. The cured fraction is modeled as a logistic function of baseline covariates. The longitudinal PSA data is modeled as a non-linear hierarchical mixed model, with different models for the cured and susceptible groups. The clinical recurrences are modeled as a time-dependent proportional hazards model for those in the susceptible group. The baseline variables are covariates in both the failure time and longitudinal models. We use both a Monte Carlo EM algorithm and Markov chain Monte Carlo techniques to fit the model. The results from the two estimation methods are compared. We focus on both selected parameters of the model and derived interpretable quantities.

*Key words and phrases:* Cure models, joint longitudinal-survival models, prostate cancer.

## 1. Introduction

There are many circumstances in which both a repeatedly-measured biomarker outcome and the elapsed time to an event are collected on each individual in a medical study. These observed biomarker series are frequently important health indicators that represent the progression of a disease. Such data will typically have additional features and complications associated with them, including the presence of treatment group indicators and baseline covariates, measurement error in the biomarkers, and right censoring of the event time with the possibility of dependent censoring. The goals for studies with data of these types can be

quite variable. The goal might be assessing how the biomarker changes with time and how this is influenced by the baseline covariates; it might be determining how the risk of the event is influenced by the biomarker and the covariates; it could be determining whether the biomarker can be used as a surrogate endpoint or as an auxiliary variable in a clinical trial, or whether it could be used to make individual predictions of future event times for patients who are censored.

One such example is prostate cancer studies. Prostate cancer is a disease which occurs primarily in older men. Prostate-Specific Antigen (PSA), a protein produced only by the cells of the prostate gland, is a well known biomarker for prostate cancer. Common treatments for the patients with local prostate cancer include radiation therapy and surgery. After treatment, clinical recurrence of disease may occur after a period of time. The pattern of PSA after radiation therapy is well recognized as an important aspect of disease progression. Clinicians monitor the outcome of the treatments by measuring PSA regularly. In patients who undergo radiation therapy, a sharp rise in PSA after the initial decline is an indicator of treatment failure, and clinical recurrence (reappearance of tumor, either local recurrence or distant metastasis) is expected to follow. Thus the longitudinal PSA could be useful for predicting cancer recurrence for patients after radiation therapy. The latest value of PSA and the slope of its increase can be very informative about the progression of disease and the hazard of a clinical recurrence. If the pattern of PSA is suggestive of an increased risk of clinical recurrence, the patients may be put on salvage hormonal therapy to slow down progression of the disease.

The specific data we consider in this article are from the University of Michigan. The patients had carcinoma of the prostate and were treated with radiation therapy between 1985 and 1997. The endpoint of interest is clinical recurrence (local recurrence or distant metastasis). The history of the PSA marker process can be predictive of the residual time to cancer recurrence for a censored subject. Regarding the censored observations as missing data, it maybe possible to recover information from the censoring by modeling both the longitudinal biomarker information and the survival time. Figure 1 are plots of post-treatment PSA measurements for 30 randomly selected patients with clinical recurrence, 30 randomly selected censored patients, and all 23 patients who did not have a clinical event but received salvage hormonal therapy. We can see a clear pattern of decline in PSA after therapy, followed potentially at a later time by an increase. The PSA profiles are different among the three groups. We see a clear trend of PSA increase at the later follow-up time after the initial decline for failed patients, while we see much flatter curves for censored patients. For patients who received hormonal therapy as a salvage therapy, we also see a rising trend of the PSA measurements. The reason is that many patients received the hormonal therapy due to elevated PSA values.

For data such as these, modeling the event time process using biomarker observations as time-dependent covariates in a proportional hazards model has its difficulties. In order to use partial likelihood to estimate the effect of covariates on the hazard rate, we must know the values of the time varying biomarker for all subjects in the risk set at any failure time point, with the risk set being defined as the set of all patients who are still under study at a time just prior to that failure time point (Fleming and Harrington (1991)). However, in many clinical studies, subjects fail on a continuous basis while the marker is usually measured only at discrete time points, and thus no measurements for the marker exist for members in the risk set when a failure occurs between scheduled follow-up visits. A popular approach to addressing this problem is to pull forward the nearest preceding value of the marker and treat it as if it were the current value of the marker at the failure time, and then use partial likelihood to obtain estimates of the hazard and marker relationship. An adaption of this approach discussed by Gail (1981) is to use the preceding marker value as an estimate of the current value if it was observed within d time units prior to the failure, or else to exclude the subject from the risk set for that event. One problem with both of these approaches is that neither accounts for measurement error in the marker value, which can cause the estimated relative risk parameter in the time-dependent Cox model to be biased toward the null, and the extent of this bias is directly proportional to the amount of measurement error in the observed marker (Prentice (1982)).

An alternative approach to estimation of parameters in a hazard model is to jointly model both the marker process and the survival data. To do this, random effect models (Laird and Ware (1982)) are often used to model the marker process and the individual random effects are included in the survival model. Two major approaches to estimation are used in the joint modeling literature: a two-stage approach (Tsiatis, DeGruttola and Wulfsohn (1995), Bycott and Taylor (1998), Dafni and Tsiastis (1998)) and a likelihood based approach (Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Henderson, Diggle and Dobson (2000), Wang and Taylor (2001) and Xu and Zeger (2001)). While many joint modeling methods focus on estimation of the covariate effects in the survival model, other aspects of the data can also be of interest, for example the relationship between the biomarker and time or other covariates. The latter can be explored when estimation is based on the likelihood from the joint model, with the parameters in the marker process model being the ones of main interest.

For the prostate cancer application, we are interested in a number of different aspects, including both how PSA changes over time and how this is influenced by other covariates and how PSA influences the hazard of clinical recurrence. A complication in prostate cancer applications is the fact that some of the patients

may have their tumor completely killed by the treatment, and so will never experience clinical recurrence. These patients are considered to be "cured". We incorporate this aspect of the study into our joint modeling by using mixture cure models.

The objectives of the article is to review joint models and cure models, and then present Bayesian analysis and compare inferences to those obtained from the likelihood analysis of the same data (Law, Taylor and Sandler (2002)). The rest of this article is organized as follows. We give a literature review of joint models in Section 2 and of cure models in Section 3. In Section 4, we describe a joint cure model. In Section 5, we describe MLE and Bayesian estimation schemes. In Section 6, we demonstrate the methodology and compare the two methods using data from a prostate cancer study. Finally, we conclude the article with a discussion section.

## 2. Joint Models for Longitudinal and Survival Data

Joint models are a class of models to describe the joint behavior of a biomarker process and an associated survival process, where the biomarker process is observed at a series of times and the survival process gives rise to censored event times. The model building usually starts from separate models for each component and then links the models together. One way to do this is by building some characteristics of the longitudinal biomarker model into the survival model. Two estimation methods are popular in the literature: a two stage approach and a likelihood based approach.

### 2.1. Two-stage approach to estimation in a time-dependent proportional hazard model

The aim of the two-stage approach is estimation of the regression coefficient in a time-dependent Cox model while addressing the limitations in our knowledge of the true trajectory of a marker. In the first stage, a model, usually data driven, is assumed for the progression of the time-varying marker. Using this model, estimates of the missing marker values at each event time point are imputed for all subjects in the risk set, and this process is repeated across all risk sets. In the second stage, these imputes are treated as the true values of the biomarker at the time of each failure for purposes of fitting a time-dependent Cox proportional hazards model.

Tsiatis et al. (1995) used this method for the analysis of data from a placebo-controlled trial of ZDV in an AIDS study. They assumed a random intercept and slope model for the true log CD4 value in the placebo group with simple error structure. Conditional on the marker history, they used the conditional mean of the multivariate normal to impute the missing marker values, and showed why it

is appropriate to use this conditional mean in a Cox model. Dafni and Tsiastis (1998) considered the situation of $k$ different treatments, assuming a different random intercept and random slope model for each treatment group. Bycott and Taylor (1998) used a Brownian motion error term to capture the biological variation and heterogeneity in the pattern of CD4 trajectories seen in individuals over time.

Although the two-stage approach reduces the bias of the parameter estimate in the Cox model, there are several drawbacks (Wulfsohn and Tsiatis (1997)). The two-stage approach does not use any survival information in modeling the marker process, this could result in bias and loss of efficiency (Faucett and Thomas (1996)). The estimated marker values from stage one are regarded as fixed in stage two, thus the approach does not propagate uncertainty from stage one to stage two.

### 2.2. Likelihood based approach

An alternative, more unified approach is to base estimation and inference on the likelihood from a joint model of both the longitudinal biomarker data and survival data. We can expect more precise and accurate estimates of the strength of the relationship between the marker and the risk of failure from such an approach. Besides making more efficient use of the data for estimation of the parameters in the Cox model, the joint model allows estimation of many different aspects. The method simultaneously estimates the parameters that describe the marker process, as well as those that describe the risk of failure as a function of the marker process.

Let $Y_i^*(t)$ denote the hypothetical *true value* of the biomarker process at time $t$ for subject $i$ and $Y_i(t)$ the corresponding observed biomarker process at time $t$. Let $\bar{Y}_i^*(t)$ denote the history up to time $t$, $\{Y_i^*(s),\, s \le t\}$ and $\bar{Y}_i(t) = \{Y_i(s); s \le t\}$ be the corresponding observed marker values up to time $t$. Let $m_i$ be the number of longitudinal observations for subject $i$ and $t_{ij}$, $j = 1, \cdots, m_i$, be the corresponding time points when these observations $Y_{ij}$ are made. We denote $Y_i = (Y_{i1}, \cdots, Y_{im_i})$ as the vector of observed longitudinal observations. Let $T_i$ be the observed time, which is the minimum of the event time $T_i^0$ and the censoring time $C_i$. The censoring indicator $\delta_i$ equals 1 if the event is observed and 0 otherwise. Let $Z_i$ denote other time-independent covariates.

Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) considered the following model

$$Y_i(t) = Y_i^*(t) + e(t), \quad Y_i^*(t) = \theta_{0i} + \theta_{1i} t \tag{1}$$

with $e(t) \sim N(0, \sigma_e^2)$ and $cov(e(s), e(t)) = 0$, $s \neq t$, and

$$\boldsymbol{\theta}_i \,|\, Z_i \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \tag{2}$$

where $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i})'$, $\boldsymbol{\theta} = (\theta_0, \theta_1)'$, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{\theta_0\theta_0} & \sigma_{\theta_0\theta_1} \\ \sigma_{\theta_1\theta_0} & \sigma_{\theta_1\theta_1} \end{pmatrix}.$$

Here $\boldsymbol{\theta} = (\theta_0, \theta_1)'$ can depend on covariates $Z_i$. A proportional hazards model was assumed for the survival data,

$$\lambda(t \,|\, \boldsymbol{\theta}_i, \bar{Y}_i^*(t), Z_i) = \lambda_0(t) \exp[\, \gamma Y_i^*(t) + \boldsymbol{\beta} Z_i \,] = \lambda_0(t) \exp[\, \gamma(\theta_{0i} + \theta_{1i} t) + \boldsymbol{\beta} Z_i \,]. \tag{3}$$

Hence the hazard depends on the marker through its current true value. Of course it is possible to make the hazard depend on the history of the true marker trajectory of that individual using other functional forms such as the rate of change or total area under the curve (Henderson, Diggle and Dobson (2000)). Also we note that for the PSA profiles, a nonlinear random effect model will be needed instead of (1) since the typical trajectory as seen in Figure 1 is clearly nonlinear.

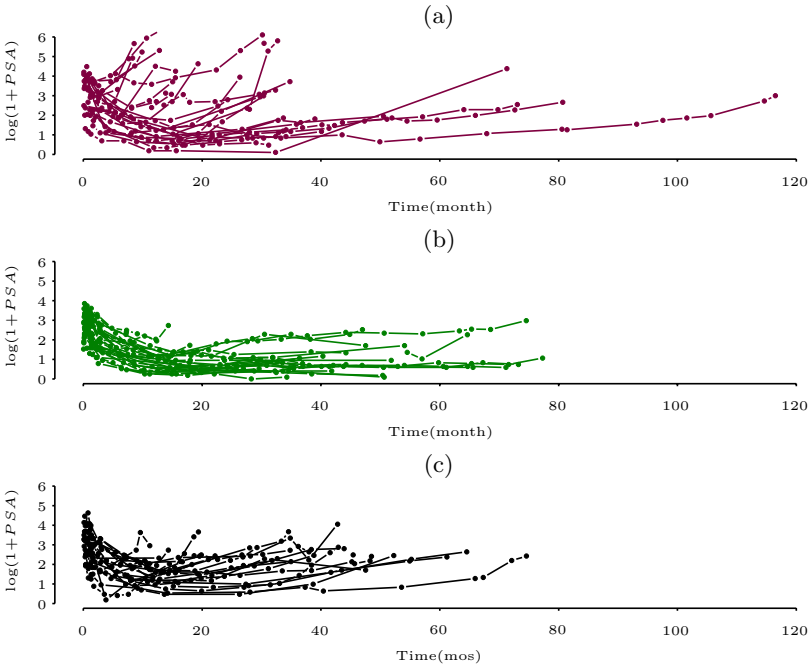


Figure 1. Observed post-treatment PSA measurements.

(a) 30 events. (b) 30 censored patients with no hormonal therapy.
(c) 23 hormonal therapy patients.

Under the models (1), (2) and (3), the complete data likelihood is

$$\prod_{i=1}^{n}\Big\{\Big[\prod_{j=1}^{m_i} g(Y_{ij}\,|\,\boldsymbol{\theta}_i,\sigma_e^2,Z_i)\Big]\,h(\boldsymbol{\theta}_i\,|\,\boldsymbol{\theta},\boldsymbol{\Sigma})\,f(t_i,\delta_i\,|\,\boldsymbol{\theta},\lambda_0,\gamma,\boldsymbol{\beta},Z_i)\Big\}, \tag{4}$$

where $g(Y_{ij}\,|\,\boldsymbol{\theta}_i,\sigma_e^2,Z_i)$ is the density of $Y_{ij}$ at time $t_{ij}$, $h(\boldsymbol{\theta}_i\,|\,\boldsymbol{\theta},\boldsymbol{\Sigma})$ is the density of the random effect $\boldsymbol{\theta}_i$ which is also normal with mean $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, and finally $f(t_i,\delta_i\,|\,\boldsymbol{\theta},\lambda_0,\gamma,\boldsymbol{\beta},Z_i)$ is the density from the survival part of the model and can be written as

$$\Big\{\lambda_0(t_i)\exp[\gamma(\theta_{0i}+\theta_{1i}t_i)+\boldsymbol{\beta}Z_i]\Big\}^{\delta_i}\,\exp\Big\{-\int_0^{t_i}\lambda_0(u)\exp[\gamma(\theta_{0i}+\theta_{1i}u)+\boldsymbol{\beta}Z_i]\mathrm{d}u\Big\}.$$

Since we do not observe the random effect $\boldsymbol{\theta}_i$, inference is based on the observed data likelihood

$$\prod_{i=1}^{n}\Big\{\int\Big[\prod_{j=1}^{m_i} g(Y_{ij}\,|\,\boldsymbol{\theta}_i,\sigma_e^2,Z_i)\Big]h(\boldsymbol{\theta}_i\,|\,\boldsymbol{\theta},\boldsymbol{\Sigma})f(t_i,\delta_i\,|\,\boldsymbol{\theta},\lambda_0,\gamma,\boldsymbol{\beta},Z_i)\mathrm{d}\boldsymbol{\theta}_i\Big\}. \tag{5}$$

The following assumptions are typically used (Faucett, Schenker and Taylor (2002) and Faucett, Schenker and Elashoff (1998)) to justify the validity of the likelihood approach. Let $\bar{Y}^*_{i,F}\equiv\bar{Y}^*_i(T_i)=\{Y^*_i(s),s\le T_i\}$ be the full history of true value of the biomarker process up to observed time $T_i$ and $\bar{Y}_{i,F}\equiv\bar{Y}_i(T_i)=\{Y_i(s),s\le T_i\}$ be the corresponding full history of the observed marker information, and let $\bar{Y}_{i,mis}$ denote the values in $\bar{Y}_{i,F}$ that are not included in the observed $Y_i$ due to only periodic follow-up.

1. The missing $\bar{Y}_{i,mis}$ of the longitudinal marker observations are ignorable (Rubin (1976)).
2. All censoring of survival time is noninformative given true marker history up to observed time $T_i$ and baseline covariates. Specifically, we assume that the censoring time $C_i$ and the event time $T_i^0$ are independent given $\bar{Y}_{i,F}$, $Z_i$ and $\bar{Y}^*_{i,F}$.
3. $C_i$ is independent of $\bar{Y}_{i,mis}$ and $\bar{Y}^*_{i,F}$, given $Y_i$ and $Z_i$.
4. The parameters of the censoring distribution $[\,C_i\,|\,\bar{Y}_{i,F},\bar{Y}^*_{i,F},Z_i\,]$ are distinct from the parameters in the joint distribution $[\,T_i^0,\bar{Y}_{i,F},\bar{Y}^*_{i,F}\,|Z_i\,]$.

For this joint model, Faucett and Thomas (1996) used Gibbs sampling for estimation. They further assumed that the baseline hazard $\lambda_0(t)$ was a step function. Although Gibbs sampling is computationally intensive, it provides a feasible approach to fit this large model without making simplifying assumptions and it is flexible enough to accommodate a variety of other expanded models. For example, assuming a more sophisticated form of the dependence between the hazard and marker process will not substantially increase the computation load.

Through a simulation study comparing joint analysis with an analysis that fits each sub-model separately using standard methods, Faucett and Thomas (1996) demonstrated that the joint analysis produced estimates that were less biased and more efficient.

Wulfsohn and Tsiatis (1997) used the EM algorithm for estimation. Closed-form maximum likelihood estimates exist for all parameters except $\gamma$, for which they used a one-step Newton-Raphson. In the E step, the main computation was in finding the conditional expectation of some functional of $\boldsymbol{\theta}_i$, namely, $E[h(\boldsymbol{\theta}_i) \mid t_i, \delta_i, Y_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}, \hat{\sigma_e^2}, \hat{\lambda}_0(t), \hat{\gamma}]$. This expectation was evaluated using the numerical integration technique of Gauss-Hermite quadrature. Law et al. (2002) who combined a joint model with a cure model for analysis of prostate cancer data, used a Monte Carlo method to evaluate expectations in their EM algorithm. This estimation method is an extension of the method developed by Sy and Taylor (2000) in the estimation of a cure model. It can also be viewed as an extension of the method by Wulfsohn and Tsiatis (1997) for a joint model of the failure time process and the longitudinal data.

Xu and Zeger (2001) extended the model given by (1), (2) and (3) by assuming a more general latent stochastic Gaussian process instead of the random effects model in (1) and (2). This latent process is also a covariate in the failure time model. To handle non-Gaussian longitudinal data such as binary outcomes, a link function is used to relate the mean of observed marker process to the latent process. Wang and Taylor (2001) used an Integrated Ornstein-Uhlenbeck process for the true underlying biomarker process $Y_i^*(t)$. Henderson et al. (2000) postulated a latent bivariate Gaussian process $\boldsymbol{\eta}(t) = \{\eta_1(t), \eta_2(t)\}$, with $\eta_1(t)$ and $\eta_2(t)$ linked to the marker and event processes respectively. Association between the marker process and the event process was then described through the cross-correlation between $\eta_1(t)$ and $\eta_2(t)$. Faucett et al. (1998) developed a joint model for a binary longitudinal covariate and censored survival data, where a Markov model was assumed for the covariate process.

Tsiatis and Davidian (2000) focussed on estimation of the parameters in the hazard model. They relaxed the normality assumption (2) for the random effects. Under the random effects model (1) and proportional hazards model (3), they used the conditional score approach to get estimating equations free of nuisance parameters. Solutions to these conditional score equations are consistent and asymptotically normal under some regularity conditions. Song, Davidian and Tsiatis (2001) extended this approach to multiple, possibly correlated markers.

Hogan and Laird (1997a/b) gave an excellent review of the methods for joint analysis of longitudinal data and survival data. They considered many of the above models from the perspective of repeated measures data with missing, possibly non-ignorable, observations. Models are classified as *selection* or *pattern mixture* models (Little (1993)). The formulation of the joint model in (1), (2) and

(3) can be viewed as a selection model where we first model the (hypothetical) complete data and then model the missing-data (survival) process conditional on the complete data. In *pattern-mixture* models, the samples are stratified by the pattern of missing data (i.e., by the time of drop-out/censoring) and then different models are assumed for the longitudinal data according to these patterns.

### 2.3. Degradation models

Whitmore, Crowder and Lawless (1998) proposed an degradation process-based approach which views the time-dependent covariates as sample paths of a diffusion process such as a Weiner process with constant drift, and then failure is determined by a threshold level (e.g., the first passage-time of the observed biomarker process) of the diffusion process. Properties of diffusion processes were utilized for inference and prediction. However, they only considered the situation where there is one measurement of the marker process and they did not discuss the incorporation of covariates. Lee, DeGruttola and Schoenfeld (2000) extended their methodology to biomedical applications and considered a generalized linear regression model to include baseline conditions and covariates. This model does not require the proportional hazards assumption, which could be viewed as an advantage. Other references in this area include Doksum and Normand (1995) and Berman (1990). An implicit assumption in degradation models is that there is a monotone progression of the biomarker. Due to the non-monotonic nature of the PSA trajectory after radiation treatment, it would be difficult to apply degradation models to our data.

## 3. Cure Model

A cure model is applicable when there are 'immunes' or 'long-term survivors' present in survival data. As a result, cured subjects are censored since cure can never be observed. On the other hand, susceptible subjects would eventually develop the endpoint if followed for long enough. Examples exist in clinical studies for many types of cancer for which a significant proportion of patients are cured. Interest in such studies can be on the effect of time independent covariates on the cure rate as well as on the time to event. In this section, we review the approaches of modeling cure in survival analysis without longitudinal data. How to incorporate longitudinal data when there are cured patients is discussed in Sections 4, 5 and 6. Two different approaches to cure rate models in survival analysis are prevalent in the literature.

### 3.1. Mixture cure models

A mixture model formulation is an attractive approach to analyzing such data, in that it contains two parts which can be interpreted separately. Berkson and Gage (1952) introduced this model by adding structure to the standard

survival model. The model can be formulated as follows. Assume that a certain fraction $p$ of the population are susceptibles and the remaining are not, then the survival function $S(t)$ for the population is given by $S(t) = pS_1(t)+(1-p)$, where $S_1(t)$ is the latent survival function for the non-cured group. Common parametric choices for $S_1(t)$ are exponential and Weibull distributions. Nonparametric choices for $S_1(t)$ have also been considered. The effects of time independent covariates on both the incidence probability $p$ and the survival distribution $S_1(t)$ for the susceptible group can be modeled. Specifically, let $(t_i, \delta_i, Z_i)$ be the observations, where $Z_i$ is a vector of time independent covariates, $t_i$ the observed or censored time, and $\delta_i$ the censoring indicator. Let $D_i$ indicate cure status for each subject with $D_i = 1$ for a non-cured subject and $D_i = 2$ for cured. Note that for a failed patient ($\delta_i = 1$), we know $D_i = 1$. Yet for a censored patient ($\delta_i = 0$), we do not observed $D_i$. The incidence model is typically given by

$$p(Z_i) = P(D_i = 1 \,|\, \mathbf{b}, Z_i) = \frac{e^{\mathbf{b}'Z_i}}{1 + e^{\mathbf{b}'Z_i}}, \tag{6}$$

where $\mathbf{b}$ is a vector of parameters. Among individuals for whom $D_i = 1$, the time to event is assumed to follow a parametric distribution. Farewell (1977, 1982) assumed a Weibull distribution

$$S_1(t_i|D_i = 1, Z_i) = \exp[-\lambda t_i^{\alpha}], \tag{7}$$

where $\alpha$ is a shape parameter and $\lambda$ is related to $Z_i$ by $\lambda = \exp(\boldsymbol{\gamma}' Z_i)$.

Different formulations can also be used in the above setting, especially in the survival model (7) for the susceptible group. 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 (1997) studied the cure model extensively, specifically nonparametric failure time models for one sample and parametric failure time regression models. 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) and 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 the EM algorithm. Li and Taylor (2002) considered a semi-parametric accelerated failure time model for the failure time process. A recent publication (Law et al. (2002)) has extended the cure model to incorporate a

longitudinal covariate. This model will be described more fully in Section 4, where it is applied to data from prostate cancer.

One problem associated with the cure model is identifiability (Farewell (1986) and Li, Taylor and Sy (2001)). 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. As a result, we can have difficulties in distinguishing models with high incidence of susceptibles and long tails of the failure time process from low incidence of susceptibles and short tails of the failure time process. Li et al. (2001) showed that the mixture cure model with a general model for the failure time process is identifiable if a parametric model such as (6) for incidence is assumed. They also considered other important special cases of the mixture cure models and non-mixture cure models, establishing conditions for identifiability. The incorporation of longitudinal data into the cure model is one way to help reduce the uncertainty about the tail of the failure time distribution for susceptibles (Law et al. (2002)). In our prostate cancer data, the use of longitudinal PSA values could be quite informative about the tail of the failure time distribution; because of its strong association with clinical recurrence, thus helping to reduce concerns about the lack of identifiability of the model. While the parameters in $p$ and $S_1(t)$ have nice interpretations, in some applications it may be the marginal survival distribution $S(t)$ and its dependence on $Z$ which is of most interest. This marginal survival distribution is also interpretable and easily obtained from the estimates of $p$ and $S_1(t)$.

### 3.2. Non-mixture cure models

It is easy to see that in the presence of covariates, the mixture cure model $S(t)$ does not have a proportional hazards structure if $S_1(t)$ is taken to be proportional hazards. In order to keep the proportional hazards structure, non-mixture cure models have been proposed (Yakovlev and Tsodikov (1996), Tsodikov (1998) and Chen, Ibrahim and Sinha (1999)). In these models, the probability of cure is incorporated into the proportional hazards model by assuming a bounded cumulative hazard $H(t)$ as $t \to \infty$ with $H(t) \leq \theta$, $\lim_{t\to\infty} H(t) = \theta$. One way to enforce this is to write $H(t) = \theta F(t)$, where $F(t)$ is the distribution function of a nonnegative random variable. Then the survival distribution $S(t)$ for the population can be written as

$$S(t) = e^{-\theta F(t)}. \tag{8}$$

We can see from (8) that the cure rate is $\lim_{t\to\infty} S(t) = e^{-\theta}$. Chen et al. (1999) showed that if $S(t)$ is taken to have a proportional hazards structure, then the conditional survival function $S_1(t)$ for the susceptible group no longer has a

proportional hazards structure. Hence in the non-mixture model, the survival distribution $S(t)$ for the entire population is modeled as a proportional hazard model, whereas in the mixture cure models, the non-cured group is often modeled as a proportional hazard model.

Covariates can be incorporated into the non-mixture cure model through $\theta$. One example is to use $\theta(Z_i) = \exp(\boldsymbol{\beta}' Z_i)$. Tsodikov (1998) treated $F(t)$ as nuisance and used marginal likelihood to estimate the cure rate $\theta(Z_i)$. Chen et al. (1999) specified a parametric form for $F(t)$ and used a Bayesian approach. In a recent article, Brown and Ibrahim (2004) have extended this non-mixture cure model to include a longitudinal covariate.

## 4. A Joint Longitudinal-Survival-Cure Model

In the joint-cure model, joint modeling of the disease progression marker and the failure time process is done in a cure model setting. The model we describe in this paper is motivated by a study of prostate cancer patients undergoing radiation therapy at the University of Michigan. We develop a statistical approach to analyzing the data from this study. The endpoint of interest is clinical recurrence. We assume that a fraction of the patients are cured by the treatment and are immune from recurrence. The model in this study involves the joint modeling of the failure time process and the longitudinal marker data, when there is a fraction of patients who are immune from the endpoint. The model has the ability to address the following types of questions. Which of the patients are being cured by the treatment? How does the probability of cure depend on the different baseline characteristics of the patients? What is the pattern of the post-treatment PSA profile for the cured and the susceptible patients? How does the pattern depend on the baseline characteristics of the patients? How does the relative risk of recurrence for the susceptible patients depend on the baseline characteristics of the patients, as well as the post-treatment PSA profiles?

### 4.1. Description of data

The data consist of 458 patients who had carcinoma of the prostate and were treated with radiation therapy between 1985 and 1997. Patients who had radiation therapy immediately following prostatectomy are excluded in the study. All patients were non-metastatic and node negative at the time of irradiation. Patients who received planned hormone therapy prior to or along with irradiation were excluded.

The endpoint of interest is clinical recurrence (local recurrence or distant metastasis). Patients who were free of clinical recurrence are considered to be censored at the last date of contact or at the time of death. Patients who started

salvage hormonal therapy are censored at the time of their salvage therapy. Of the 458 patients, 92 had either local recurrence (50) or distant metastasis (42). Among the 366 censored patients, 50 were dead, 23 had started hormonal therapy and 293 were censored at their latest follow-up time. The median follow-up time was 45 months (the range was from 1.2 month to 118 months). The median time to event was 25 months (the range was from 2.4 months to 72 months). There are 66 patients censored at follow-up time larger than 72 months, suggesting the presence of a cured group.

The baseline covariates included in the model are the baseline PSA value, T stage and Gleason score. T stage is a categorical variable with 3 levels: T1, T2 and T3-T4. Gleason score is treated as a continuous variable. The baseline PSA (bPSA) values are transformed by log(1+bPSA). In the fitting of the models, the continuous covariates are centered to improve convergence.

Post-treatment longitudinal PSA were measured about every 6 months. For patients who had started the hormonal therapy, only PSA measured prior to the hormonal therapy are used. For patients who developed clinical recurrence, the PSA measurements after the endpoint are included, since we assume that the event of recurrence does not change the trajectory of the PSA profile. A total of 4,226 post-treatment PSA values are included in the study, the median number of PSA measurement for each patient is 9 (the range is from 1 to 26). The pattern of the post-treatment PSA measurements is illustrated in Figure 1, which shows data for patients with events, censored patients who did not receive hormonal therapy, and censored patients who did receive hormonal therapy. The pattern of PSA after radiation therapy is an important aspect of disease progression and in many studies the endpoint is in fact based on a pattern of increasing PSA, instead of being based on clinical disease recurrence. There has been considerable debate over what constitutes a clinically meaningful rise in PSA, but unfortunately, the relationship between the pattern of PSA and disease recurrence has not been investigated in a precise quantitative way. Thus there is a need for developing models in which the pattern of PSA and the development of recurrence are considered as separate, but linked, entities.

The determination of "cure" is difficult among prostate cancer patients due to the slow progression of the disease. Since cure can never be observed, and recurrence is thought to be possible as late as 10 years after treatment, it is hard to determine if a patient is cured. The slow progression of the disease also means that a lot of the non-cured patients are censored before they experience clinical recurrence. Since the patterns of the post-treatment PSA are different in cured and non-cured patients, we can make use of the PSA data to help distinguish the cured and the non-cured among the censored patients. The post-treatment PSA profiles also have important information on the relative risk of recurrence

for those who are not cured, therefore modeling the failure time data and the disease progression marker data jointly is appropriate and necessary. Another complication in prostate cancer studies is informative censoring. As mentioned above, a sharp rise in PSA usually precedes clinical recurrence. However, having observed such a rise in the PSA levels, the clinician may initiate salvage hormonal therapy before the patient actually experiences a clinical recurrence. If the censoring time was taken as the date of salvage therapy, then this could lead to biased estimation of the survival distribution. The joint modeling approach can potentially eliminate the bias from this type of censoring.

The model used in the current paper is the same as that described in Law et al (2002), who developed a ML estimation procedure. Here we focus on different aspects of the model and data analysis results, and compare the ML approach and results with a Bayesian estimation technique.

### 4.2. Notation and model specification

Let $Z_i$ denote the $q+1$ fixed baseline covariates for subject $i$ including the intercept. The $m_i$ post-treatment PSA measurements of an individual are denoted by the vector $Y_i = (Y_{i1}, \cdots, Y_{im_i})$, with the corresponding measurement time vector $t_i = (t_{i1}, \cdots, t_{im_i})$. Let $T_i$ be the observed follow-up time, and $\delta_i$ be the corresponding censoring indicator. The cure group indicator is denoted by $D_i$. For a subject $i$ in the susceptible group, $D_i$ is equal to 1; otherwise, it is equal to 2.

**Incidence model**

The probability of an individual $i$ to be in the susceptible group is given by the logistic function:

$$P(D_i = 1 \,|\, \mathbf{b},\, Z_i) = \frac{\exp(\mathbf{b}' Z_i)}{1 + \exp(\mathbf{b}' Z_i)}. \tag{9}$$

**Longitudinal model**

The post-treatment PSA data are modeled by a hierarchical nonlinear mixed effects model. The response model of PSA is given by

$$\log(Y_{ij} + 1) = \log(Y_{ij}^* + 1) + \epsilon_{ij}, \quad j = 1, \cdots, m_i, \tag{10}$$

where $Y_{ij}^*$ is the "true" PSA process at time $t_{ij}$, $\epsilon_{ij}$ is the measurement error at time $t_{ij}$, and the $\epsilon_{ij}$ are i.i.d. $N(0, \sigma_e^2)$. The transformation for the PSA values is done to adjust for the skewness of the distribution and to minimize the influence of extremely low PSA values.

The true PSA marker process is modeled by a nonlinear exponential decay-exponential growth model (Kaplan, Cox and Bagshaw (1991), Zagars and Pollack (1993) and Cox, Kaplan and Bagshaw (1993)):

$$Y_i^*(t) = r_{i1}e^{-r_{i2}t} + r_{i3}e^{r_{i4}t}, \tag{11}$$

where $r_{i1}, r_{i2}, r_{i3}$ and $r_{i4}$ are the unobserved random effects for subject $i$ ($r_{i1}, r_{i2}$, $r_{i3}$ and $r_{i4} > 0$). The term $(r_{i1} + r_{i3})$ is the intercept of the post-treatment PSA profile, $r_{i2}$ is the rate of decline of the PSA following treatment, while $r_{i4}$ is the rate of rise following the initial decline. Although this model was empirically derived, it does have some nice biological interpretations, in particular the exponential growth part $e^{r_{i4}t}$ reflects the fact that in a recurring tumor, PSA is thought to be proportional to the volume of the tumor and tumors grow approximately exponentially.

Depending on the patient's cure status $D_i$, we use different mixed effect models for the true underlying marker profile. For the random effects of a subject $i$ in the susceptible group, we assume

$$[\,\mathbf{R}_i \,|\, D_i = 1, Z_i\,] \sim N(\mathbf{Z}_i^1\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \tag{12}$$

where $\mathbf{R}_i$ denotes the log random effects $(\log r_{i1}, \log r_{i2}, \log r_{i3}, \log r_{i4})$, and $\mathbf{Z}_i^1\boldsymbol{\mu}_1$ is the mean vectors of the random effects in the susceptible group, $\mathbf{Z}_i^1 = (I_4 \otimes Z_i)^T$, and $\boldsymbol{\Sigma}_1$ is the corresponding covariance matrix.

For the random effects of a subject $i$ in the cured group, we assume that the rate of rise denoted by $r_{i4}$ is zero:

$$\begin{cases} [\,\mathbf{R}_{i(-4)} \,|\, D_i = 2, Z_i\,] \sim N(\mathbf{Z}_i^2\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \\ [\,r_{i4} \,|\, D_i = 2\,] \equiv 0, \end{cases} \tag{13}$$

where $\mathbf{R}_{i(-4)} \equiv (\log r_{i1}, \log r_{i2}, \log r_{i3})$, $\mathbf{Z}_i^2\boldsymbol{\mu}_2$ is the mean vectors of these random effects in the cure group, $\mathbf{Z}_i^2 = (I_3 \otimes Z_i)^T$, and $\boldsymbol{\Sigma}_2$ is the corresponding covariance matrix.

**Conditional failure time model**

A time-dependent Cox proportional hazards regression model is used to model the time to endpoint for the subjects in the susceptible group. Conditional on the unobserved random effects, the relative hazard function of the event time $t$ is given by

$$\lambda(t \,|\, D_i = 1, \mathbf{R}_i, Z_i) = \lambda_0(t|\boldsymbol{\eta}) \exp[\,\gamma \log(Y_i^*(t) + 1) + \boldsymbol{\beta}' Z_i^*\,], \tag{14}$$

where $\lambda_0(t|\boldsymbol{\eta})$ is the unspecified baseline hazard function at time $t$ and $Z_i^*$ is a vector of $q$ baseline covariates without intercept.

Note that the parameter vector $\boldsymbol{\beta}$ 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 $\gamma$). Equation (14) can be generalized to include other aspects of the PSA marker process, for example, the initial rate of decline given by $r_{i2}$, or the subsequent rate of rise given by $r_{i4}$.

## 5. Estimation Methods

Two estimation methods, maximum likelihood and Bayesian, are described below. The maximum likelihood method was described in Law et al. (2002). We briefly describe it here giving some additional details which were not previously presented. The Bayesian approach was developed as an alternative with which to contrast the ML results. We note that since there are different longitudinal models in the cured and non-cured group, the use of a two-stage estimation method is not feasible.

### 5.1. EM algorithm

Let $\Omega = (\mathbf{b}, \sigma_e, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \boldsymbol{\beta}, \lambda_0)$ denote the parameters of the model, with $D$ and $R$ regarded as latent variables. Let $X_{i,obs} = (Z_i, Y_i, t_i, \delta_i)$ denote the observed data, and $X_i = (Z_i, Y_i, t_i, \delta_i, D_i, \mathbf{R}_i)$ denote the complete data. Maximum likelihood estimation is performed to obtain the parameter estimates. Briefly, an EM algorithm is used to implement the maximum likelihood estimation. In the E-step, the cure group indicator ($D_i$) of the censored subjects and the random effects ($\mathbf{R}_i$) are treated as missing data. The evaluation of the expected complete data log-likelihood requires Monte Carlo integration. This is performed using importance sampling. In the M-step the complete data log-likelihood partitions into five separate parts, each of which is maximized by Newton Raphson. Standard errors are based on the inverse observed information matrix, with the baseline hazard function treated as a vector of parameters at each distinct event time. Convergence of the parameters and the log-likelihood estimates is monitored graphically. The final estimate for each of the parameters is the average of the estimates from the last ten EM iterations. Computations are performed using MATLAB.

Predicting the cure status of a censored patients is interesting to both physicians and patients. The formula to estimate the probability that a censored subject is in the susceptible group is given by

$$P(D_i = 1 \mid X_{i,obs}, \hat{\Omega}) = \frac{\int \{\hat{P}_i \, \hat{S}_i \, \hat{g}_i \, \hat{h}_{1i}\} \, \mathrm{d}\mathbf{R}_i}{\int \{\hat{P}_i \, \hat{S}_i \, \hat{g}_i \, \hat{h}_{1i}\} \, \mathrm{d}\mathbf{R}_i + \int \{(1 - \hat{P}_i) \hat{g}_i \, \hat{h}_{2i}\} \, \mathrm{d}\mathbf{R}_i}, \qquad (15)$$

where $\hat{P}_i = P(D_i = 1 \,|\, \hat{\mathbf{b}}, Z_i)$ is the incidence probability from (9) evaluated at $\hat{\mathbf{b}}$, $\hat{S}_i = S(t_i \,|\, \mathbf{R}_i, \hat{\boldsymbol{\beta}}, \hat{\gamma}, Z_i)$ is the conditional survival probability from (14) evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$, $\hat{g}_i = g(\log(Y_i + 1) \,|\, \mathbf{R}_i, \hat{\sigma_e}, Z_i)$ is the normal density for transformed longitudinal data from (10), and $\hat{h}_{1i} = h(\mathbf{R}_i \,|\, D_i = 1, \hat{\boldsymbol{\mu}}_1, Z_i)$ and $\hat{h}_{2i} = h(\mathbf{R}_i \,|\, D_i = 2, \hat{\boldsymbol{\mu}}_2, Z_i)$ are densities for random effects conditioning on their incidence group from (12) and (13). Detailed derivation is given in Law et al. (2002).

### 5.2. Markov chain Monte Carlo

We also fit the model using a MCMC technique. Further, due to the fact that a normal person's PSA level actually slowly increases with age, instead of (13), we assume that for a subject in the cure group,

$$\log r_{i4} \,|\, D_i = 2 \sim N(-6, \sigma_{44}), \tag{16}$$

where $-6$ is chosen from the fact that PSA level doubles on average in about 20 years for a healthy male. Hence the covariance matrix $\boldsymbol{\Sigma}_2^*$ of $\mathbf{R}_i$ for the cured group is block-diagonal with two blocks, $\boldsymbol{\Sigma}_2$ and $\sigma_{44}$. We assume a normal model for $\log r_{i4}$ to allow for variations of the growth in the cured group. An additional slight modification of the model is that we use a parametric baseline hazard function in (14) for the susceptible group. Specifically, we assume the hazard from a Weibull distribution, that is, $\lambda_0(t) = \alpha\lambda t^{\alpha-1}$. We make this change because a parametric form for $\lambda_0(t)$ is more convenient to handle in an MCMC approach than an unspecified form. Furthermore, it is scientifically reasonable that $\lambda_0(t)$ should be a smooth function.

Data-driven vague normal priors are taken for $\mathbf{b}, \gamma$ and $\boldsymbol{\beta}$ with the prior mean derived from estimates from separate analysis. Specifically, we treat all censored patients with censoring time $> 60$ months and last longitudinal PSA $< 4$ as cured. Then we fit a logistic model to all patients in order to get the mean of the normal prior for $\mathbf{b}$, we set prior variance of each component of $\mathbf{b}$ as 16, approximate 100 times the variance estimate from this simple method. Similarly, we obtain prior means of $\gamma$ and $\boldsymbol{\beta}$ by fitting a Cox proportional hazard model to 'susceptible' patients from the above simplified rule and using the nearest preceding value of the PSA as the current value. The prior variances are obtained by inflating the variance estimate from the simpler method approximately 100 times. Vague conjugate priors are used for other parameters. Multivariate normal distributions are used as prior distributions for each row of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and these prior distributions are assumed to be independent. That is, the prior distributions for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are products of multivariate normal distributions. The prior of $\sigma_e^2$ has an inverse Gamma distribution and priors for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ have inverse Wishart distributions. An inverse Gamma distribution is used as the

prior for $\sigma_{44}$, the variance of $\log r_{i4}$ for a subject $i$ in the cure group. For the scale parameter $\lambda$ of the baseline hazard, the prior is taken from a gamma distribution. For the shape parameter $\alpha$ of $\lambda_0(t)$, for computational ease we assume that it has a discrete distribution with support points on the set $0.5, 0.55, 0.6, \ldots, 2.5$. The prior is then taken to be uniform on these points.

The posterior distributions for all the parameters can be obtained from the product of full complete data likelihood and prior distributions. The full complete data likelihood is

$$\begin{aligned} L = \prod_{i=1}^{n} \Big\{ & g\Big(\log(Y_i+1) \,|\, D_i = 1, \mathbf{R}_i, \sigma_e^2\Big) f(t \,|\, \mathbf{R}_i, \boldsymbol{\beta}, \gamma, \alpha, \lambda, Z_i)^{I(\delta_i=1)} \\ & S(t \,|\, \mathbf{R}_i, \boldsymbol{\beta}, \gamma, \alpha, \lambda, Z_i)^{I(\delta_i=0)} h(\mathbf{R}_i |\, \mathbf{Z}_i^1 \boldsymbol{\mu}_1, \boldsymbol{\Sigma_1}) P(D_i = 1 \,|\, \mathbf{b},\, Z_i) \Big\}^{I(D_i=1)} \\ & \cdot \Big\{ g\Big(\log(Y_i+1) \,|\, D_i = 2, \mathbf{R}_i, \sigma_e^2\Big) h(\mathbf{R}_{i(-4)} |\, \mathbf{Z}_i^2 \boldsymbol{\mu}_2, \boldsymbol{\Sigma_2}) h(R_{i4} \mid -6, \sigma_{44}^2) \\ & \qquad P(D_i = 2 \,|\, \mathbf{b},\, Z_i) \Big\}^{I(D_i=2)}, \end{aligned}$$

where $f(t \,|\, \cdot) = \lambda(t \,|\, \cdot) S(t \,|\, \cdot)$ is the density function of the conditional failure time model for subjects in the susceptible group.

Adaptive Rejection Sampling (Gilks and Wild (1992)) is used for $\mathbf{b}, \gamma$ and $\boldsymbol{\beta}$ since the posteriors are log-concave. Adaptive Rejection Metropolis Sampling (Gilks, Best and Tan (1995)) is used for random effects $\mathbf{R}_i$ since their posterior distributions are approximately log-concave. Explicit forms exist for other parameters since conjugate priors are used.

We use multiple sequences (Gelman and Rubin (1992)) to check for convergence of the Gibbs sampler. Overdispersed starting values are used in 25 different chains. We eliminate a total of 3,000 iterations as burn-in and then generate an additional 10,000 samples for summarization. Computations are performed using C++. For each censored patient, we estimate the conditional recurrence probability $P(D_i \,|\, X_{i,obs})$ from the draws of $D_i$, that is, the average of the number of 1's among all drawn $D_i$'s. Note that this conditional probability $P(D_i \,|\, X_{i,obs})$ is different from (15) since we do not fix the parameters at their estimated values.

## 6. Data Analysis

### 6.1. Comparison of parameter estimates

Tables 1, 2 and 3 give the results of fit for some of the major regression parameters. For the most part, the parameter estimates between the EM and MCMC methods are similar, but not identical. The SD's from the MCMC method are usually similar to the SE's from the EM method, although for a few parameters they are a bit smaller. In the incidence model shown in Table 1, all the baseline covariates have significant effects on the probability of recurrence using the

EM algorithm. Hence measures of increased tumor size and aggressiveness are associated with decreased probability of cure. For the effect of Gleason score, although it is not significant from the MCMC method, it has the same sign as the estimate using the EM algorithm.

Table 1. Parameters in the incidence model: **b**.

| | EM | | | MCMC | | |
|---|---|---|---|---|---|---|
| Parameter | Estimates | S.E. | Est/S.E. | Mean | S.D. | Mean/S.D. |
| intercept | 1.06 | 0.48 | 2.24 | 1.15 | 0.43 | 2.70 |
| T1 | -1.79 | 0.62 | -2.86 | -1.29 | 0.54 | -2.41 |
| T2 | -1.27 | 0.50 | -2.52 | -1.20 | 0.39 | -3.07 |
| log(1+bPSA) | 1.19 | 0.20 | 5.82 | 0.85 | 0.18 | 4.57 |
| Gleason | 0.25 | 0.11 | 2.19 | 0.14 | 0.11 | 1.26 |

Table 2. Parameters in the failure time model: $\gamma, \boldsymbol{\beta}$.

| | EM | | | MCMC | | |
|---|---|---|---|---|---|---|
| Parameter | Estimates | S.E. | Est/S.E. | Mean | S.D. | Mean/S.D. |
| log(1+PSA) | 1.07 | 0.09 | 12.54 | 0.70 | 0.06 | 11.67 |
| T1 | -1.82 | 0.49 | -3.71 | -2.27 | 0.55 | -4.09 |
| T2 | -0.84 | 0.25 | -3.30 | -0.71 | 0.24 | -2.92 |
| log(1+bPSA) | -0.55 | 0.14 | -3.86 | -0.07 | 0.14 | -0.48 |
| Gleason | 0.27 | 0.09 | 2.89 | 0.35 | 0.09 | 3.99 |

Table 3. Random effects $\log(r_4)$ in the longitudinal model for the susceptible group.

| | EM | | | MCMC | | |
|---|---|---|---|---|---|---|
| Parameter | Estimates | S.E. | Est/S.E. | Mean | S.D. | Mean/S.D. |
| intercept | -2.42 | 0.17 | -14.24 | -2.53 | 0.23 | -10.98 |
| T1 | -1.00 | 0.26 | -3.85 | -0.93 | 0.18 | -5.23 |
| T2 | -0.49 | 0.17 | -2.88 | -0.46 | 0.10 | -4.72 |
| log(1+bPSA) | 0.04 | 0.09 | 0.44 | 0.15 | 0.07 | 2.07 |
| Gleason | 0.21 | 0.05 | 4.20 | 0.22 | 0.04 | 5.25 |

Both T stage and Gleason score are also significant in the conditional failure time model, see Table 2. The current PSA value is highly significant in the failure time model. We note that baseline PSA has a negative sign rather than the expected positive sign. It is significant from the EM algorithm. This may be due to the high correlation between 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. 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. The estimated effect of baseline PSA is nonsignificant and close to 0 using the MCMC method. We can also see from the table that the magnitude of the estimated effect of current PSA is smaller from MCMC than that from the EM algorithm. The differences between the EM and MCMC results are consistent with a high correlation between the baseline PSA and the current PSA.

The parameters estimates for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, which describe the association between the baseline covariates and the longitudinal pattern of PSA are given in Law et al. (2002) for the EM method. The effects of covariates on the rate of rise ($r_{i4}$) is especially interesting in that it determines how badly a susceptible's PSA deteriorates. Table 3 shows that the rate of rise is associated with both T stage and Gleason score, but not with baseline PSA from the EM algorithm, while it is associated with all covariates from the MCMC results.

### 6.2. Comparison of marginal survival distributions

A second way to compare the results from the two estimation methods is through other derived interpretable quantities. We examined the marginal survival distributions for fixed values of the covariates. For a patient with baseline covariate values $Z_i$, the marginal survival function $S(t \,|\, Z_i)$ can be written as

$$S(t \,|\, Z_i) = P(T_i > t \,|\, Z_i) = P(D_i = 1 \,|\, Z_i) \times S(t \,|\, D_i = 1, Z_i) + 1 - P(D_i = 1 \,|\, Z_i). \tag{17}$$

In the EM algorithm, the estimated marginal survival function for a patient with baseline covariate values $Z_i$ is then approximated by

$$\hat{S}(t \,|\, Z_i) \approx \hat{P}(D_i = 1 \,|\, Z_i) \times \hat{S}(t \,|\, D_i = 1, Z_i) + 1 - \hat{P}(D_i = 1 \,|\, Z_i), \tag{18}$$

where $\hat{P}(D_i = 1 \,|\, Z_i)$ is calculated from (9) using MLE of $\mathbf{b}$. To estimate $\hat{S}(t \,|\, D_i = 1, Z_i) = \exp[-\int_0^t \hat{\lambda}(v|D_i = 1, \mathbf{R}_i, Z_i)\, \mathrm{d}v\,]$ using (14), we need to obtain the estimated PSA value at $t$. We use the mean estimate for $\mathbf{R}_i$ from (12) with MLE $\hat{\boldsymbol{\mu}}_1$ from the EM, that is,

$$\hat{Y}_i^*(t) = e^{E[R_{i1}]} \exp\Big\{-e^{E[R_{i2}]}t\Big\} + e^{E[R_{i3}]} \exp\Big\{e^{E[R_{i4}]}t\Big\}.$$

In MCMC, we can also use (17) to estimate the marginal survival function. We estimate $P(D_i = 1 \,|\, Z_i)$ also from (9) using the posterior mean of $\mathbf{b}$. But we estimate the conditional survival function $S(t \,|\, D_i = 1, Z_i)$ differently. We draw $K$ vectors of $\mathbf{R}_i$ from (12) with estimates of $\boldsymbol{\mu}_1$ from MCMC output. After obtaining the estimate of $Y_i^*(t)$ from each draw of $\mathbf{R}_i^{(k)}$, we calculate an estimate of $S^{(k)}(t \,|\, D_i = 1, Z_i)$ with the other parameters $\boldsymbol{\beta}, \gamma, \lambda$ and $\alpha$ fixed at the posterior mean from MCMC output. Finally, we average over $k$ and obtain the estimate of $S(t \,|\, D_i = 1, Z_i)$ as

$$\tilde{S}(t \,|\, D_i = 1, Z_i) = \frac{1}{K} \sum_1^K S^{(k)}(t \,|\, D_i = 1, Z_i).$$

The estimated marginal survival function $\tilde{S}(t \mid Z_i)$ for a patient with baseline covariate values $Z_i$ is approximated by

$$\tilde{S}(t \mid Z_i) \approx \tilde{P}(D_i = 1 \mid Z_i) \times \tilde{S}(t \mid D_i = 1, Z_i) + 1 - \tilde{P}(D_i = 1 \mid Z_i). \tag{19}$$

Figure 2 shows the substantial effect of the baseline covariates on the marginal survival probability from both the EM algorithm and MCMC. We do not plot after 73 months since the last event happened at 72 months. Note that, from Table 2, the baseline PSA has a negative effect in the hazard model, however the overall effect is positive due to the substantial effect of baseline PSA in the incidence and longitudinal model.

(a) T stage (T1, T2, T3/T4) at Gleason score = 6, baseline PSA = 14;
(b) Gleason score (3, 6, 9) at T2, baseline PSA = 14;
(c) baseline PSA(1, 14, 79) at T2, Gleason score =6;

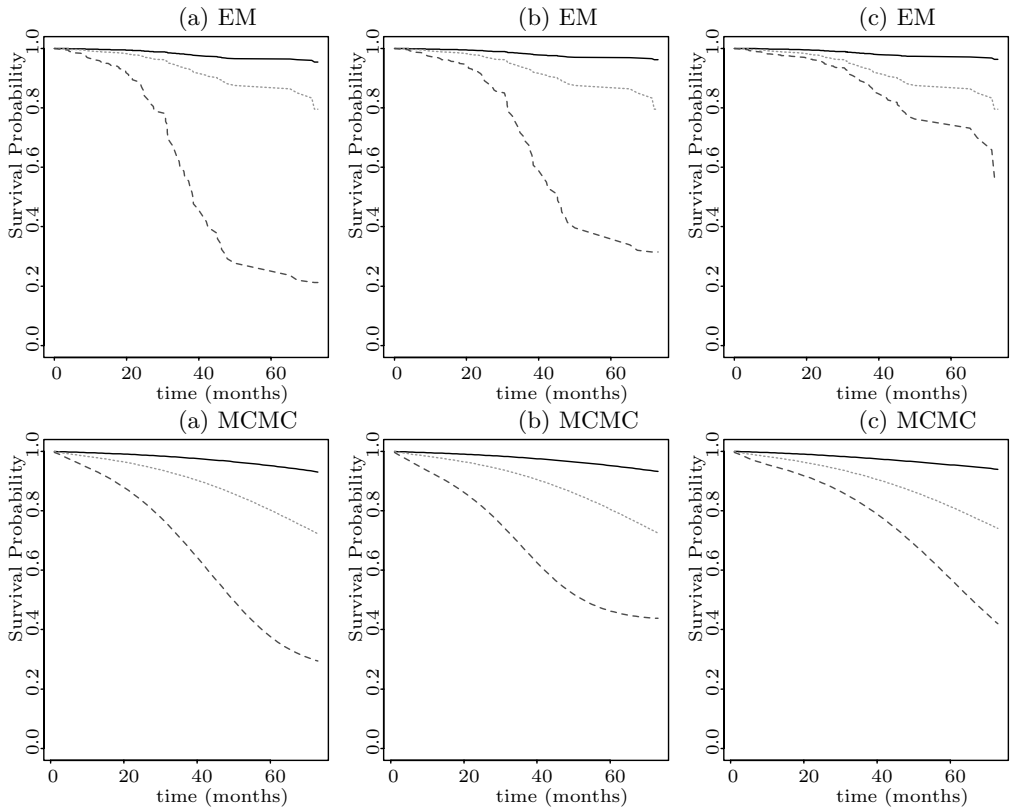


Figure 2. Marginal survival probability.

The marginal survival distribution is a quantity which could also be estimated from a more standard approach, such as a proportional hazards model with time-independent baseline covariates. In principle, the joint cure model

approach is preferred, because it uses the information in the longitudinal PSA data, which would lead to higher efficiency and a reduction in the bias due to dependent censoring.

### 6.3. Comparison of estimated recurrence probabilities

With the MLE $\hat{\mathbf{\Omega}}$ of the parameters in $\mathbf{\Omega}$ from the EM algorithm, the conditional recurrence probability for a censored subject, $p(D_i = 1 \mid X_{i,obs}, \hat{\mathbf{\Omega}})$, can be estimated using (15). In the MCMC method, however, we look at the conditional recurrence probability $P(D_i \mid X_{i,obs})$ without conditioning on the parameters in $\mathbf{\Omega}$. Calculation of $P(D_i \mid X_{i,obs})$ is straightforward since we have draws of $D_i$ conditional on observed data $X_{i,obs}$ from the MCMC output, and $P(D_i \mid X_{i,obs})$ is the proportion of 1's among the draws of $D_i$.

Figure 3 gives the plots of the estimated conditional recurrence probability with the censored time and baseline PSA for the censored subjects from both the EM algorithm and MCMC. We can see that the conditional recurrence probability generally increases with baseline PSA. Patients who were censored at a later time tend to have a lower conditional recurrence probability. Note that estimates of the recurrence probabilities $P(D_i \mid X_{i,obs})$ from the MCMC output allocate more patients in the middle region between zero and one than estimates of $P(D_i \mid X_{i,obs}, \hat{\mathbf{\Omega}})$ from the EM algorithm.

(a) censored time (with spline), (b) baseline PSA (with spline).

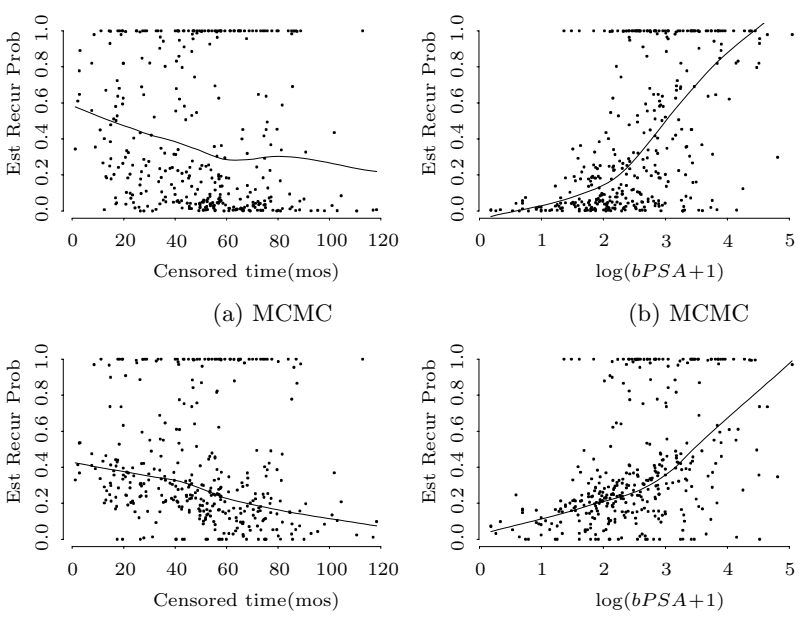


Figure 3. Plots of the estimated conditional recurrence probability.

### 6.4. Comparison of individual profiles

Figure 4 shows the estimated profiles of PSA from the EM algorithm for selected censored subjects. There are two estimated PSA profiles conditional on the subject being in the cure group or the susceptible group. The final estimates of the conditional recurrence probability are also 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 for these individuals are 1. The estimated probabilities are lower for the two subjects of plots (c) and (d), and the plots show good fit of both the susceptible group profile and the cure group profile. The subject in (c) is probably from the cure group. The susceptible group PSA profile fits reasonably well, because the exponential growth part of the model does allow for a slow rate of increase of PSA. However, such a small rate of increase is not typical of others in the susceptible group and hence the probability of this subject being in the susceptible group is very low. The subject in (d) has only a few PSA measurements, making it difficult to classify him into the cure/susceptible group based on the observed PSA data. Thus the estimated recurrence probability of 0.58 is derived mainly from the baseline covariates.

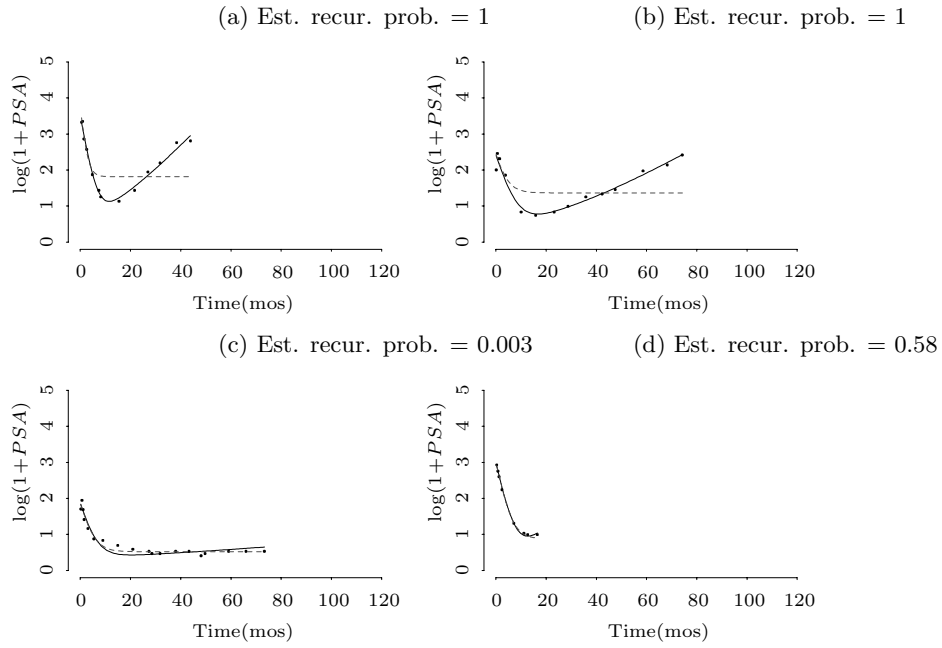


Figure 4. Individual estimated PSA profiles for selected censored subjects − from the EM algorithm†.

† The solid line is the PSA profile conditional on the subject being in the susceptible group, and the dotted line is the profile conditional on the cure group. The observed PSA value are denoted by •.

Figure 5 plots the estimated PSA profiles derived from the MCMC method for the same set of patients as those in Figure 4. We can see a very similar fit to the longitudinal data. The estimated recurrence probabilities for the patients in (c) and (d) are slightly different from the two methods.

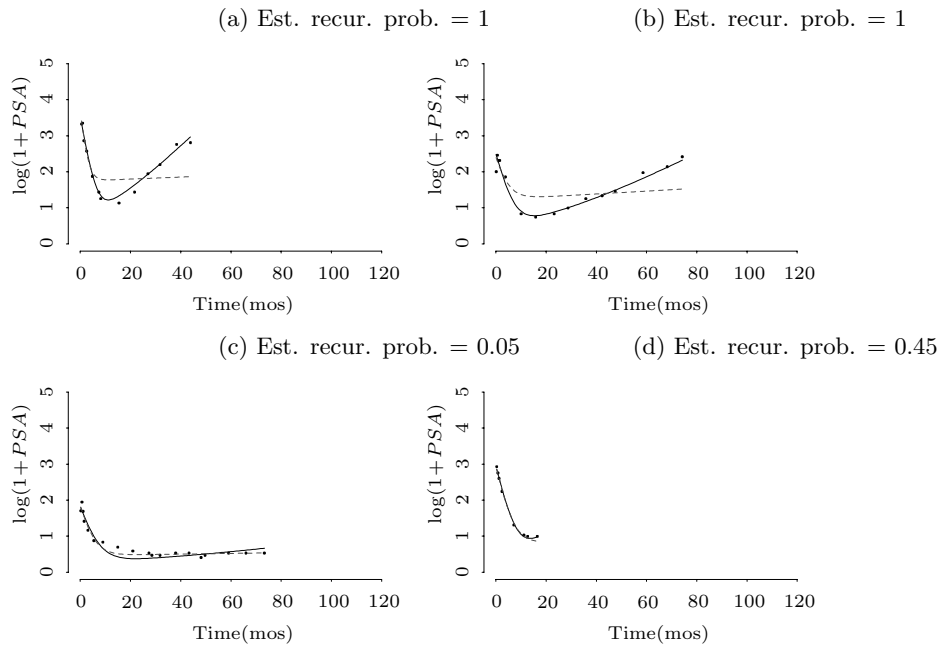


Figure 5. Individual estimated PSA profiles for selected censored subjects − from the MCMC†.

† The solid line is the PSA profile conditional on the subject being in the susceptible group, and the dotted line is the profile conditional on the cure group. The observed PSA value are denoted by •. Same set of patients as in Figure 4 are used.

In general, we find that both methods of estimation work well for this data set. The size and the direction of the parameter estimates in different components of the model are consistent with expectation. Results from both methods are usually close. The estimated longitudinal PSA profiles fit well to the observed PSA measurements. The difference in the results from the two methods may result from using different forms of the baseline hazard function and different probability assumptions for the rate of rise for cured patients. Another difference is that the MCMC approach requires specification of a prior distribution. While we have not performed a full sensitivity analysis, we have considered more diffuse priors for regression parameters $\mathbf{b}$, $\gamma$ and $\boldsymbol{\beta}$ and found this has little effect on the final results of estimates. Increasing the range of the support points for $\alpha$

of the baseline hazard $\lambda_0(t)$ does not seem to be necessary since the posterior probability of $\alpha$ lying outside $[1.3, 2]$ is almost null. The results of the MCMC method can depend on the assumption concerning $r_{i4}$ given in (16). By setting the mean of $r_{i4}$ closer to 0 and using a prior with less variation, we find that the parameter estimates of the MCMC method become closer to the estimates of the EM algorithm. However, the marginal survival distribution estimated by (19) is very robust to different distributional assumptions of $r_{i4}$.

## 7.Discussion

In this paper, a joint-cure model is developed, in which the longitudinal disease progression marker and the failure time process are modeled jointly, when a fraction of subjects are immune from the endpoint. We have performed simulation studies to evaluate the performance of the EM algorithm to fit this model, the results are described elsewhere (Law et al. 2002). The simulation studies show that in the presence of the disease progression marker, including such longitudinal information in the cure model improves the classification of censored subjects into the correct group. The cure model without the longitudinal component gives estimated conditional recurrence probabilities less concentrated at zero or one.

Within the three components in the joint-cure model, other model specifications can be assumed. For instance, in the current model, we include the current PSA value only in the conditional failure time model. Other aspects of the longitudinal profile can be considered, 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. Also we can include other baseline covariates such as age and total dose of radiation in the incidence model.

Whether a maximum likelihood or a Bayesian method is preferred is largely a matter of personal choice and computational convenience. Both methods can be generalized easily to include additional covariates and the computational burden will not increase much when we incorporate such expanded models. The MCMC method has the advantage that we have draws available for all parameters including the latent variables $D_i$ and $R_i$. These draws can be utilized when we consider certain applications of the model. For example, we are currently considering using the model for medical monitoring and updating of individual predictions for censored patients when we have additional follow-up data or data for new patients. For this application the MCMC method provides a more natural framework for making individual predictions than is provided by the EM algorithm.

Regarding the longitudinal data model, the exponential decay-exponential growth model is appropriate for the prostate cancer data set that we analyze.

However, the statistical approach developed in this paper can be generalized to other studies with context-specific models used for each of the three component models.

A usual characteristic of a cure model is a marginal survival distribution which levels off at long times. In Figure 2 the follow-up times are not long enough to clearly see a plateau. The slow progressive nature of prostate cancer also means that recurrences are possible many years after the initial treatment. Thus despite the strong scientific rationale for a cure component, it may be possible to fit these data without using a cure model. From a pragmatic point of view the addition of the cure model component could just be viewed as a means of formulating a richer and more flexible class of models. From this viewpoint, the specific parameters of the cure model are of less interest, whereas quantities such as the marginal survival distribution are inherently interpretable irrespective of the choice of model.

A natural question which arises when there are several models available to fit a data set is model selection. Within a Bayesian framework, possible approaches include BIC and Bayes factors and prediction based criteria, such as the Conditional Predictive Ordinate (CPO) (Geisser (1993) and Gelfand, Day and Chang (1992)). We are currently investigating the use of these measures to choose between non-nested models in the context of joint longitudinal, survival and cure models.

## References

Berman, S. M. (1990). A stochastic model for the distribution of HIV latency time based on T4 counts. *Biometrika* **77**, 733-741.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *J. Amer. Statist. Assoc.* **47**, 501-515.

Brown, E. R. and Ibrahim, J. G. (2004). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*. To appear.

Bycott, P. and Taylor, J. M. G. (1998). A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statist. Medicine* **17**, 2061-2077.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a survival fraction. *J. Amer. Statist. Assoc.* **94**, 909-919.

Cox, R. S., Kaplan, I. D. and Bagshaw, M. A. (1993). Prostate-specific antigen kinetics after external beam irradiation for carcinoma of the prostate. *Internat. J. Radiation Oncology, Biology, Physics* **28**, 23-31.

Dafni, U. G. and Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics* **54**, 1445-1462.

Doksum, K. A. and Normand, S. L. T. (1995). Gaussian models for degradation processes - part 1: methods for the analysis of biomarker data. *Lifetime Data Anal.* **1**, 131-144.

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

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

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *Canad. J. Statist.* **14**, 257-262.

Faucett, C. L., Schenker, N. and Elashoff, R. M. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *J. Amer. Statist. Assoc.* **93**, 427-437.

Faucett, C. L., Schenker, N. and Taylor, J. M. G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58**, 37-47.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statist. Medicine* **15**, 1663-1685.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Gail, M. (1981). Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics* **37**, 67-78.

Geisser, S. (1993). *Predictive Inference: An Introduction.* Chapman & Hall, London.

Gelfand, A. E., Day, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with Discussion). In *Bayesian Statistics* **4** (Edited by J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith), 147-167. Oxford University Press, Oxford.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457-472.

Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. *Bayesian Statist.* **4**, 1-9.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

Hogan, J. W. and Laird, N. M. (1997a). Mixture models for the joint distribution of repeated measures and event times. *Statist. Medicine* **16**, 239-257.

Hogan, J. W. and Laird, N. M. (1997b). Model-based approaches to analyzing incomplete longitudinal and failure time data. *Statist. Medicine* **16**, 259-272.

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.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal model. *Biometrics* **38**, 963-974.

Law, N. J., Taylor, J. M. G. and Sandler, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* **3**, 547-563.

Lee, M. L. T., DeGruttola, V. and Schoenfeld, D. (2000). A model for markers and latent health status. *J. Roy. Statist. Soc. Ser. B* **62**, 747-762.

Li, C. S. and Taylor, J. M. G. (2002). A semi-parametric accelerated failure time cure model. *Statist. Medicine* **21**, 3235-3247.

Li, C. S., Taylor, J. M. G. and Sy, J. P. (2001). Identifiability of cure models. *Statist. Probab. Lett.* **54**, 389-395.

Little, R. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.* **88**, 125-134.

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

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-342.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-590.

Song, X, Davidian, M. and Tsiatis, A. A. (2001). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* **1**, 1-25.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227-236.

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

Tsodikov, A. (1998). A proportional hazard model taking account of long-term survivors. *Biometrics* **54**, 1508-1516.

Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27-37.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447-458.

Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Amer. Statist. Assoc.* **96**, 895-905.

Whitmore, G. A., Crowder, M. J. and Lawless, J. F. (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Anal.* **4**, 229-251.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl. Statist.* **50**, 375-387.

Yamaguchi, K. (1992). Accelerated failure time regression models with a regression time model of surviving fraction: an application to the analysis of 'permanent employment'. *J. Amer. Statist. Assoc.* **87**, 284-292.

Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and their Biostatistical Applications*. World Scientific, River Edge, New Jersey.

Zagars, G. K. and Pollack, A. (1993). The fall and rise of prostate-specific antigen. *Cancer* **72**, 832-842.

Department of Medicine / Biostatistics, Indiana University, 1050 Wishard Boulevard, RG4101, Indianapolis, IN 46202, U.S.A.

E-mail: meyu@iupui.edu

Roche Molecular Systems, Inc., 4300 Hacienda Drive, Pleasanton, CA 94588, U.S.A.

E-mail: Jacqueline.law@roche.com

Department of Biostatistics, University of Michigan, Ann Arbor, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.

E-mail: jmgt@umich.edu

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

E-mail: hsandler@umich.edu

(Received January 2003; accepted July 2003)