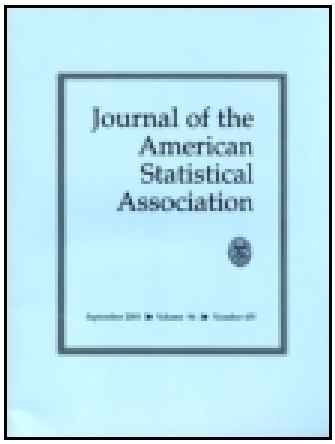

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival-Cure Model

Menggang Yu[a], Jeremy M. G Taylor[a] & Howard M Sandler[a]
[a] Menggang Yu is Assistant Professor, Department of Medicine, Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202 . Jeremy M. G. Taylor is Professor, Department of Biostatistics, and Howard M. Sandler is Professor, Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109.
Published online: 01 Jan 2012.

To cite this article: Menggang Yu, Jeremy M. G Taylor & Howard M Sandler (2008) Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival-Cure Model, Journal of the American Statistical Association, 103:481, 178-187, DOI: 10.1198/016214507000000400

To link to this article: http://dx.doi.org/10.1198/016214507000000400

Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival–Cure Model

Menggang YU, Jeremy M. G. TAYLOR, and Howard M. SANDLER

Patients treated for prostate cancer are monitored by periodically measuring prostate-specific antigen (PSA) after treatment. Increases in PSA are suggestive of cancer recurrence and are used in making decisions about possible new treatments. The data from studies of such patients typically consist of longitudinal PSA measurements, censored event times, and baseline covariates. Methods for the combined analysis of both longitudinal and survival data have been developed in recent years, with the main emphasis on modeling and estimation. We analyze data from a prostate cancer study in which the patients are treated with radiation therapy, using a joint model extended by adding a mixture structure to the model. Here we focus on using this model to make individualized predictions of disease progression for censored and alive patients. In this model, each patient is assumed to be either cured by the treatment or susceptible to clinical recurrence. The cured fraction is modeled as a logistic function of baseline covariates, measured before the end of the radiation therapy period. The longitudinal PSA data is modeled as a nonlinear hierarchical mixed model, with different models for the cured and susceptible groups. To accommodate the heavy tail manifested by the data and possible outliers, a t distribution is used for the measurement error. The clinical recurrences are modeled using a time-dependent proportional hazards model for those in the susceptible group, where the time-dependent covariates include both the current value and the slope the of posttreatment PSA profile. The baseline hazard is assumed to have a generalized Weibull form. Estimates of the parameters in the model are obtained using a Markov chain Monte Carlo method. The model is used to give individual predictions of both future PSA values and the predicted probability of recurrence up to four years in the future. These predictions are compared with observed data from a validation data set consisting of further follow-up of the subjects in the study. There is good correspondence between the predictions and the validation data.

KEY WORDS: Cure models; Joint model; Longitudinal model; Markov chain Monte Carlo; Prostate cancer data; Survival model.

1. INTRODUCTION

In many circumstances, both a repeatedly measured biomarker outcome and the elapsed time to an event are collected on each individual in a medical study. These biomarkers are frequently important health indicators that represent the progression of a disease. Such data typically will have additional associated features and complications, 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. Joint models for both the marker process and survival data have been developed in recent years to analyze such data. Estimation of the parameters can be done through a two-stage approach (e.g., Tsiatis, DeGruttola, and Wolfsohn 1995; Bycott and Taylor 1998) and a likelihood- or Bayesian-based approach (e.g., Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997; Henderson, Diggle, and Dobson 2000; Wang and Taylor 2001; Xu and Zeger 2001; Pauler and Finkelstein 2002; Law, Taylor, and Sandler 2002; Brown and Ibrahim 2003). A review of these approaches has been given by Yu, Law, Taylor, and Sandler (2004).

Prostate-specific antigen (PSA) is a well-known biomarker for prostate cancer used both for screening and for monitoring response to treatment. It is a routine laboratory assay obtained in a blood sample and thus is easy to acquire. Common treatments for patients with local prostate cancer include radiation therapy and surgery. After treatment, clinical recurrence of disease may occur after a period of time. Clinicians and patients monitor the outcome of the treatment by measuring PSA regularly; slight changes or increased values can be a source of great concern and anxiety. In patients undergoing 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, although it can be many years before clinical manifestations of the recurrence appear. If the PSA remains low and stable, this is an indication that the tumor is not regrowing in the patient and the patient may be cured. Thus the longitudinal PSA can 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 regarding of disease progression and the hazard of clinical recurrence. If the pattern of PSA is suggestive of an increased risk of clinical recurrence, the patient may be put on new therapy based solely on this pattern of PSA (typically hormone therapy with substantial potential side effects), to slow progression of the disease. Thus methods that enhance early detection of recurrence and accurate prediction of future disease progression for an individual patient based on the pattern of PSA values can have great utility.

A feature in many cancer applications is the fact that some patients may have their tumor completely killed by the treatment and so will never experience clinical recurrence. These patients are considered "cured." We incorporate this aspect of the study into our joint modeling using mixture cure models (e.g., Farewell 1982; Kuk and Chen 1992; Taylor 1995). An alternative modeling strategy for data for which a cure appears plausible is to use the bounded cumulative hazard cure model (Yakovlev and Tsodikov 1996; Chen, Ibrahim, and Sinka 1999). This approach also has recently been extended to the joint model setting (Brown and Ibrahim 2003; Chen et al. 2004).

The longitudinal survival–cure model that we adopt here has been used by Law et al. (2002) and Yu et al. (2004). We extend the previously developed model in a number of ways. Besides including additional baseline covariates, we use a time-dependent proportional hazards model depending on hormonal therapy (HT), the current slope, as well as the current value of

Menggang Yu is Assistant Professor, Department of Medicine, Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202 (E-mail: *meyu@iupui.edu*). Jeremy M. G. Taylor is Professor, Department of Biostatistics, and Howard M. Sandler is Professor, Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109.

Journal of the American Statistical Association
March 2008, Vol. 103, No. 481, Applications and Case Studies
DOI 10.1198/016214507000000400

PSA. To accommodate heavy tails manifested in the longitudinal data and possible outliers, we use t distributions for the measurement error. This model allows estimation of a number of different aspects, including how PSA changes over time, how this change is influenced by other covariates, and how PSA influences the hazard of clinical recurrence. However, the main focus of this article is on using the model to make individualized predictions for disease progression. Specifically, we predict future PSA values and cancer recurrence probabilities for censored and alive patients. We evaluate the performance of the prediction using a validation data set obtained through further follow-up for these patients.

The task of using a series of biomarker values is considerably more complicated than using a single value for early detection of disease or monitoring disease progression. The hope is that using all of the data from the serial observations will lead to an earlier and more precise prediction of future disease progression. There are a few examples of using serial observations in the statistical literature, all involving fairly complicated models and considerable computation. These include using CA125 for early detection of ovarian cancer (Skates, Pauler, and Jacobs 2001), using PSA for early detection of prostate cancer (Slate and Cronin 1997), and using PSA for detection of disease recurrence in prostate cancer (Pauler and Finkelstein 2002).

The rest of this article is organized as follows. Section 2 describes two data sets: an analysis data set and a validation data set. Section 3 describes a joint cure model, and Section 4 presents the Bayesian estimation schemes for the model, and Section 5 gives the results. Section 6 uses the model to make individualized predictions, and Section 7 assesses the performance of the prediction through a comparison to a validation data set. Finally, Section 8 concludes the article with a discussion.

2. DATA

The joint model is developed and fit to one data set, called the analysis data set. Predictions derived from this data set are compared with observations in a second data set, called the validation data set, which consists of further follow-up on the subjects in the analysis data set.

2.1 Analysis Data Set

The data consist of 928 patients with localized prostate cancer who were treated with external beam radiation therapy at the University of Michigan between July 1987 and February 2001. Patients were excluded from this analysis if they received planned HT before the end of the radiation therapy regimen. The baseline variables were age, radiation dose and duration, T-stage (a measure of the size and location of the tumor), Gleason score (a measure of the aggressiveness of the tumor) and pretreatment PSA. Earlier versions of these data have been described elsewhere (Sandler et al. 2000). Posttreatment PSA was measured at approximately 6-month intervals. The median number of PSA values per patient was 6 (range, 1–29). The total number of PSA measurements was 6,150. The maximum time between treatment and a PSA measurement was 145 months. During the follow-up period, up to February 2001, 146 patients experienced clinical recurrences, including 70 with local recurrence, 72 with distant metastases, and 4 with regional failure as

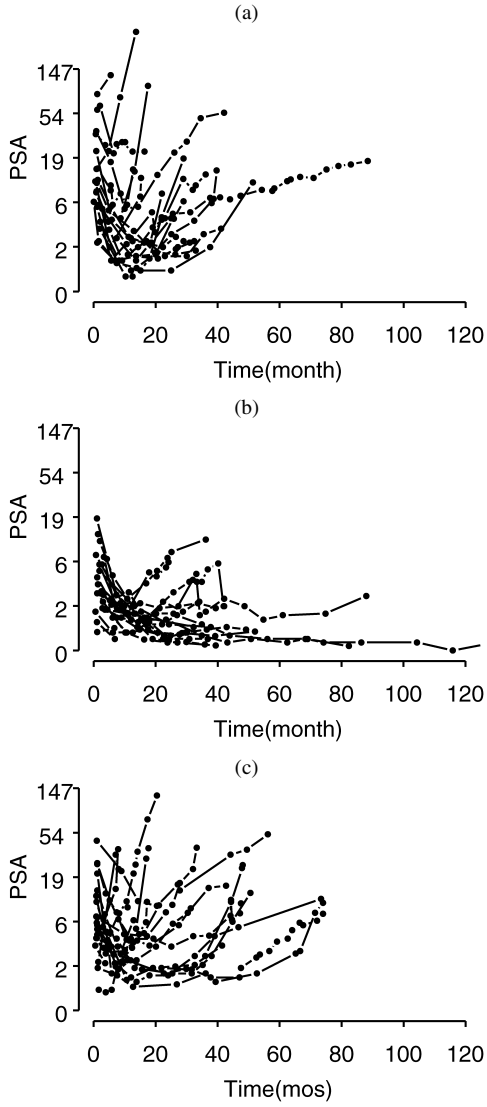

Figure 1. Observed posttreatment PSA measurements for (a) 20 patients who had clinical recurrence; (b) 20 censored patients with no HT; and (c) 20 HT patients.

their first clinical recurrence. Of the 782 censored patients, 143 died of other causes before any clinical recurrence and 639 were either censored at February 2001 or lost to follow-up before that date. A total 56 patients received HT before any clinical recurrence; of these, 15 had a later clinical recurrence.

For patients who received HT, only PSA measured before the HT are used. For patients who developed clinical recurrence, the PSA measurements before the endpoint are included; for the other (censored) patients, all PSA measurements are included.

Figure 1 provides plots of posttreatment PSA measurements for 20 randomly selected patients from each of three categories of patients: those who experienced a clinical recurrence, those who were censored and did not receive HT, and those who received HT before recurrence. We can see a clear pattern of decline in PSA after therapy, followed by a possible later 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, with much flatter curves for censored patients. For patients who received HT

as a salvage therapy, we also see a trend toward increasing PSA measurements. This is expected because HT is usually given because of rising PSA.

2.2 Validation Data Set

The validation data set consists of additional data collected on these 928 patients after February 2001 and available in September 2003. We restrict attention to the 612 patients who were alive at the last contact time in the analysis data set and were not known to have experienced a clinical recurrence or received HT before February 2001. There were 541 patients where new follow-up information was available. Among these 541 patients, 472 were alive at the end of the this new follow-up period, 63 died from other causes not related to prostate cancer, and 6 died from prostate cancer. The median additional follow-up time was 30 months. There were 329 patients with additional PSA values; these patients provided 992 PSA measurements within 3 years of the previous last follow-up date. Fifteen of the patients developed clinical recurrence in the new follow-up period; 6 of these 15 received HT before the recurrence. An additional 14 have received HT without any clinical recurrence.

3. NOTATION AND MODEL SPECIFICATION

Let $\mathbf{z}_i = \{z_{i1}, \ldots, z_{iq}\}$ be the q fixed baseline covariates for subject i. The n_i posttreatment PSA measurements of an individual are denoted by vector $\tilde{\mathbf{y}}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{in_i})$, with the corresponding measurement time vector $\tilde{\mathbf{t}}_i = (t_{i1}, \ldots, t_{in_i})$. Log-transformed posttreatment PSA measurements are denoted by $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i}) \equiv (\log(\tilde{y}_{i1} + 1), \ldots, \log(\tilde{y}_{in_i} + 1))$. This transformation is used so that residuals from the models better satisfy the assumptions of symmetry and homogeneity of variance and also to reduce the influence of extremely high PSA values.

Let t_i be the observed follow-up time, and let δ_i be the corresponding censoring indicator. The latent 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. D_i is not observed for censored subjects.

Let $\mathbf{x}_{i,obs} = \{\mathbf{y}_i, t_i, \delta_i\}$ be the observed response data for subject i and let $\mathbf{X}_{obs} = \{\mathbf{x}_{i,obs},\ i = 1, \ldots, n\}$. Denote $\mathbf{Z} = \{\mathbf{z}_i,\ i = 1, \ldots, n\}$.

Incidence Model

The probability that an individual i is in the susceptible group is given by the logistic function

$$P(D_i = 1|\mathbf{b}, \mathbf{z}_i) = \frac{\exp(b_0 + b_1 z_{i1} + \cdots + b_q z_{iq})}{1 + \exp(b_0 + b_1 z_{i1} + \cdots + b_q z_{iq})}. \quad (1)$$

Longitudinal Model

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

$$\begin{aligned} Y_{ij} &\equiv \log(\widetilde{Y}_{ij} + 1) \\ &= \log(PSA^*_{ij} + 1) + \epsilon_{ij}, \qquad j = 1, \ldots, n_i, \end{aligned} \quad (2)$$

where $PSA^*_{ij} \equiv PSA^*(t_{ij})$ is the "true" PSA process at time t_{ij} (the expression of which is defined later) and ϵ_{ij} is the measurement error at time t_{ij}. The measurement error terms ϵ_{ij} are assumed to follow a mean-0 t distribution with degree of freedom $\nu > 1$, and scale parameter σ_e for all $j = 1, \ldots, n_i$; $i = 1, \ldots, n$. Note that the t distribution can be written as a scale mixture of normal distributions (Lange, Little, and Taylor 1989); that is, a latent variable ζ_{ij} can be introduced such that

$$\epsilon_{ij}|\zeta_{ij} \sim \mathrm{N}(0, \sigma_e^2/\zeta_{ij}) \quad \text{with } \zeta_{ij} \sim \mathit{Gamma}(\nu/2, \nu/2).$$

The "true" PSA marker process is modeled by a nonlinear exponential decay and exponential growth model (Zagars and Pollack (1993)),

$$PSA^*_i(t) = r_{i1}\exp(-r_{i2}t) + r_{i3}\exp(r_{i4}t), \quad (3)$$

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 posttreatment PSA profile, r_{i2} is the rate of decline of PSA after treatment, whereas r_{i4} is the rate of rise after the initial decline.

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

$$[\mathbf{R}_i|D_i = 1, \mathbf{z}_i] \sim \mathrm{N}\big(\mathbf{Z}_i^{(1)}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\big), \quad (4)$$

where $\mathbf{R}_i$ denotes the log random effects ($\log r_{i1}, \log r_{i2}, \log r_{i3}$, and $\log r_{i4}$) and $\mathbf{Z}_i^{(1)}\boldsymbol{\mu}_1$ is the mean vectors of the random effects in the susceptible group. $\mathbf{Z}_i^{(1)} = (\mathbf{I}_4 \otimes \mathbf{z}_i^*)^T$ is a Kronecker product between $\mathbf{I}_4$ and $\mathbf{z}_i^*$, a vector of baseline covariates.

For the random effects of a subject i in the cured group, we assume that the expected rate of rise denoted by r_{i4} is close to 0,

$$\begin{cases} \big[\mathbf{R}_{i(-4)}|D_i = 2, \mathbf{z}_i\big] \sim \mathrm{N}\big(\mathbf{Z}_i^{(2)}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\big) \\ [R_{i4}|D_i = 2] \sim \mathrm{N}(-6, \sigma_{44}), \end{cases} \quad (5)$$

where $\mathbf{R}_{i(-4)} \equiv (\log r_{i1}, \log r_{i2}, \log r_{i3})$, $R_{i4} \equiv \log r_{i4}$, and $\mathbf{Z}_i^{(2)}\boldsymbol{\mu}_2$ is the mean vector of these random effects in the cure group, where $\mathbf{Z}_i^{(2)} = (\mathbf{I}_3 \otimes \mathbf{z}_i^*)^T$. The mean R_{i4} of -6 for a cured patient is chosen based on the fact that PSA level doubles in about 20 years on average for healthy males. Thus the covariance matrix $\boldsymbol{\Sigma}_2^*$ of $\mathbf{R}_i$ for the cured group is block-diagonal with two blocks, $\boldsymbol{\Sigma}_2$ and σ_{44}. We assume a normal model for R_{i4}, to allow for variations of the growth in the cured group.

Conditional Failure Time Model

Conditional on the unobserved random effects, the relative hazard function of the event time t is given by

$$\begin{aligned} \lambda(t|D_i = 1, \mathbf{R}_i, \mathbf{z}_i) = \lambda_0(t|\boldsymbol{\eta})\exp[\gamma PSA_i(t) + \omega sl_i(t)^g \\ + \kappa HT_i(t, a) + \boldsymbol{\beta}'\mathbf{z}_i], \end{aligned} \quad (6)$$

where $sl_i(t) = |\partial PSA_i(t)/\partial t|$ is the absolute value of the slope of $PSA_i(t) \equiv \log(PSA^*_i(t) + 1)$ at time t and is raised to power g, and $HT_i(t, a)$ is a given function of t and parameter a that is 0 before patient i receives HT and is >0 after the patient receives HT. Due to the fact that only a small number of patients received HT and an even smaller number of these patients had events, we take a simple function form for HT, $HT_i(t) = (1 - \frac{t-th_i}{a})I(th_i < t < th_i + a)$, where th_i is the time of initiation of HT for a patient receiving HT and $I(\cdot)$ is the indicator function. Thus $HT_i(t) = 0$ before th_i and $HT_i(t) = 1$

when $t = th_i$ and it decreases linearly to 0 at $t = th_i + a$. The reason for taking a decreasing function for HT is that the effect of HT is thought to diminish over time as the patient becomes resistant to the therapy. However, we can take a to be infinity and have a constant and nondiminishing effect of HT.

Our choice to include the slope of the underlying PSA curve is based on commonly used empirical criteria in characterizing prostate cancer progression, such as PSA doubling time and PSA velocity. We use a power transformation with parameter g restricted to be positive. Our choice for the baseline hazard is from the generalized Weibull family. The hazard has the form $\lambda_0(t|\boldsymbol{\eta}) = \theta\alpha\lambda t^{\alpha-1}(1+\lambda t^{\alpha})^{\theta-1}$. The hazard rate of a generalized Weibull is very flexible. It can be monotone increasing or decreasing; it also can be ∩-shaped or ∪-shaped depending on the different values for α, θ, and λ (Bagdonavicius and Nikulin 2001). We note that in (6) the terms $PSA_i(t)$, $sl_i(t)$, and $HT_i(t)$ are internal covariates (in the sense of Kalbfleisch and Prentice 2002), and thus the coefficients associated with them must be interpreted with care.

4. ESTIMATION USING MARKOV CHAIN MONTE CARLO

Whether a maximum likelihood or a Bayesian method is preferred is largely a matter of personal choice and computational convenience. Both methods are computationally intensive in our model setting. When fitting a simpler model, the two methods yield similar results (Yu et al. 2004). The MCMC method has the advantage of having draws available for all parameters, including the latent variables D_i and R_i. These draws can be used when making individual predictions, as we demonstrate in Section 6.

4.1 Prior Distributions

We take data-driven vague normal priors for $\mathbf{b}, \gamma, \omega, \kappa$, and $\boldsymbol{\beta}$. Specifically, to obtain the parameters of these prior distributions, 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 to get the mean of the normal prior for $\mathbf{b}$. We set the prior variance of each component of $\mathbf{b}$ as 16, which is approximately 100 times the variance estimate from this simple method. Similarly, we obtain prior means of γ, ω, κ, and $\boldsymbol{\beta}$ by fitting a Cox proportional hazard model to noncured patients from the foregoing simplified rule and using the nearest preceding value of PSA as the current value. We obtain the prior variances by inflating the variance estimate from the simpler method approximately 100 times.

We use vague conjugate priors for other parameters, taking independent multivariate normal distributions as prior distributions for each row of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. The prior of σ_e^2 has an inverse-gamma distribution with mean 1 and variance 10. The prior for $\boldsymbol{\Sigma}_1$ is from the inverse-Wishart distribution $Inv\text{-}Wishart_{\nu_1^0}(\mathbf{S}_1^0)$, with mean $\mathbf{S}_1^0/(\nu_1^0 - 5)$. We take the mean as the estimated $\hat{\boldsymbol{\Sigma}}_1$ from simplified analysis using the longitudinal data alone and degrees of freedom $\nu_1^0 = 20$, which is close to the dimension of $\boldsymbol{\Sigma}_1$. From the posterior distribution of $\boldsymbol{\Sigma}_1$, the prior specification has little impact on the posterior distribution. Similarly, the prior for $\boldsymbol{\Sigma}_2$ is from the inverse-Wishart distribution $Inv\text{-}Wishart_{\nu_2^0}(\mathbf{S}_2^0)$ with degrees of freedom $\nu_2^0 = 19$, which is close to the dimension of $\boldsymbol{\Sigma}_2$.

We used an inverse-gamma distribution with mean .1 and variance 2 as the prior for σ_{44}. We use .1 for the mean because there should be less variation for r_{i4} for the cured group. From the posterior distribution of σ_{44}, we observe that the posterior distribution is dominated by the data. For the parameter λ of the baseline hazard $\lambda_0(t)$, we take the prior from a gamma distribution with mean .01 and variance 100. We assume uniform priors U(.1, 10) for α and U(.1, 10) for θ of $\lambda_0(t)$.

We specify noninformative priors with certain restrictions for tuning parameters g and a in the conditional failure model component (6) and the degrees of freedom ν of the longitudinal model component (2). Specifically, we restrict the power index parameter g to be positive. We take the duration a of HT to be at least 3 years, discrete, and in half-year intervals. We also allow a to be infinity corresponding to a constant effect of HT; thus we allow a to be in the set $\{36 + 6i, i = 0, 1, \ldots, 20\} \cup \{\infty\}$. For degrees of freedom ν, we restrict it to be >1.

4.2 Posterior Distributions and Implementation Details

The posterior distributions for all of the parameters can be obtained from the product of full complete-data likelihood and prior distributions. Tsiatis and Davidian (2004) have discussed in detail the conditions for facilitating expression of likelihood in the joint modeling setting. With the addition of the cure model, we need to assume that censoring and timing of longitudinal measurements are noninformative given a patient's complete history including baseline covariates, PSA trajectory, and cure status. The full complete-data likelihood is then determined by model components (1), (2), (3), (4), (5), and (6) specified in Section 3 given the fully observed data $\mathbf{x}_{obs} = \{\mathbf{z}_i, \mathbf{y}_i, t_i, \delta_i,\ i = 1, \ldots, n\}$,

$$
\begin{aligned}
L = \prod_i \Bigg[\Bigg\{ \prod_j & N(y_{ij}|D_i = 1, \mathbf{R}_i, \sigma_e^2/\zeta_{ij})^{I(D_i=1)} \\
& \times N(y_{ij}|D_i = 2, \mathbf{R}_i, \sigma_e^2/\zeta_{ij})^{I(D_i=2)} \Bigg\} \\
& \times \prod_j [\zeta_{ij}] f(t|\mathbf{R}_i, \mathbf{z}_i, \boldsymbol{\beta}, \gamma, \omega, \kappa, \boldsymbol{\eta})^{I(\delta_i=1)} \\
& \times S(t|\mathbf{R}_i, \mathbf{z}_i, \boldsymbol{\beta}, \gamma, \omega, \kappa, \boldsymbol{\eta})^{I(D_i=1)I(\delta_i=0)} \Bigg] \\
& \times \prod_i \{ h(\mathbf{R}_i|D_i = 1, \boldsymbol{\mu}_1, \mathbf{z}_i)^{I(D_i=1)} \\
& \times h(\mathbf{R}_i|D_i = 2, \boldsymbol{\mu}_2, \mathbf{z}_i)^{I(D_i=2)} \} \\
& \times \prod_i \{ P(D_i = 1|\mathbf{b}, \mathbf{z}_i)^{I(D_i=1)} \\
& \times P(D_i = 2|\mathbf{b}, \mathbf{z}_i)^{I(D_i=2)} \}, \qquad (7)
\end{aligned}
$$

where $N(y_{ij}|D_i = 1, \mathbf{R}_i, \sigma_e^2/\zeta_{ij})$ and $N(y_{ij}|D_i = 2, \mathbf{R}_i, \sigma_e^2/\zeta_{ij})$ are the normal densities for transformed longitudinal data from (2) conditioning on their incidence group, $f(t|\cdot) = \lambda(t|\cdot)S(t|\cdot)$ is the density function of the conditional failure time model for subjects in the susceptible group, and $h(\mathbf{R}_i|D_i = 1, \boldsymbol{\mu}_1, \mathbf{z}_i)$ and $h(\mathbf{R}_i|D_i = 2, \boldsymbol{\mu}_2, \mathbf{z}_i)$ are densities for random effects conditioning on their incidence group from (4)

and (5). The expressions for all conditional posterior distributions have been given by Yu (2004).

We use adaptive rejection sampling (Gilks and Wild 1992) for $\mathbf{b}, \gamma, \boldsymbol{\beta}, \omega$, and κ because the posteriors are log-concave. We use a random-walk chain (Metropolis et al. 1953; Casella and George 1992) algorithm for random effects $\mathbf{R}_i$, and mean-0 normal distributions for perturbation to get candidate draws. The standard deviations are chosen so that the acceptance rate is about .20. Similarly, we use a random-walk chain is used for α, θ, λ, g, and ν. For discrete parameters such as a, the posteriors are multinomially distributed and can be drawn directly. The continuous covariates are centered to improve convergence and reduce correlation between them. For censored subjects, we draw two sets of values of $\mathbf{R}_i$, one under the assumption that $D_i = 1$ and one under the assumption $D_i = 2$. This facilitates drawing a new value of D_i in the chain.

The program is written in C++. It takes an average of 50 hours to run 15,000 iterations on a Sun workstation. We use both the multiple sequences (Gelman and Rubin 1992) method and traceplots for population parameters to check for convergence of the Gibbs sampler in the early exploratory stage of checking convergence. The final results are based on 10,000 draws after 5,000 iterations of burn-in.

5. RESULTS

The main parameter estimates from the model fitting are listed in Table 1. The table shows the posterior mean and standard deviation derived from the posterior draws of the parameters. It compares the ratio of posterior mean over posterior standard deviation of a parameter, then compares this value with 1.96 to assess the significance of covariate effects. For tuning parameters g and ν, the table reports confidence intervals. For the HT effect duration a, it reports the posterior mode, its relative frequency, and relative frequency ratio of the mode over the second most frequently occurring value.

Table 1 shows the results for the incidence model (1). We find that a patient's tumor stage is significantly related to the probability of cure in the expected direction. Baseline PSA also appears to be affecting the probability of cure, although not significantly at the .05 level. Total radiations dose, Gleason score, age, and duration of treatment are not significant. For the conditional failure time model, we see that the slope of PSA profile and current value of PSA affect the hazard of cancer recurrence; a large slope greatly increases the hazard. HT is associated with reduced risk, whereas higher Gleason score is associated with elevated risk for not-cured patients. Tumor stage is also associated with the hazard of cancer recurrence. Being at T-stage 1 significantly reduces the hazard.

There appears to be a need to transform the slope, because the posterior draws of power parameter g are significantly different from 1. Posterior draws seem to favor a constant and nondiminishing effect of HT. The degrees of freedom for the t distribution used to model measurement error are estimated as 2.2.

6. MODEL PREDICTIONS

Suppose that we wish to forecast the t_0 months recurrence-free probability $P\{T_i > t_i + t_0 | \mathbf{X}_{obs}, \mathbf{Z}\}$ for a censored patient i (still alive at the censored time t_i) based on the available data $\mathbf{X}_{obs}, \mathbf{Z}$. Let $\boldsymbol{\Omega} \equiv \{\mathbf{b}, \sigma_e, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \omega, \kappa, \alpha, \lambda, \theta, g, a, \nu\}$ denote the population parameters in the joint models described in Section 3. With K draws $\{\boldsymbol{\Omega}^{(k)}, D_i^{(k)}, R_i^{(k)}, k = 1, \ldots, K\}$ from the posterior distribution $[\boldsymbol{\Omega}, D_i, \mathbf{R}_i | \mathbf{X}_{obs}, \mathbf{Z}]$, we can approximate $P\{T_i > t_i + t_0 | \mathbf{X}_{obs}, \mathbf{Z}\}$ by

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K} P\left\{T_i > t_i + t_0 | \boldsymbol{\Omega}^{(k)}, D_i^{(k)}, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z}\right\} \\
&\quad = \frac{1}{K}\sum_{k=1}^{K}\left[I\left(D_i^{(k)} = 2\right) + I\left(D_i^{(k)} = 1\right)\right. \\
&\qquad \left. \times P\left\{T_i > t_i + t_0 | \boldsymbol{\Omega}^{(k)}, D_i^{(k)} = 1, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z}\right\}\right], \qquad (8)
\end{aligned}$$

where $P\{T_i > t_i + t_0 | \boldsymbol{\Omega}^{(k)}, D_i^{(k)} = 1, \mathbf{R}_i^{(k)}, \mathbf{X}_{obs}, \mathbf{Z}\}$ is calculated from (6).

We consider two conditional survival curves 4 years after the last contact time for patient i who is censored at time t_i: $P\{T_i > t_i + t | \mathbf{X}_{obs}, \mathbf{Z}$, no HT in $[t_i, t_i + t]\}$ and $P\{T_i > t_i + t | \mathbf{X}_{obs}, \mathbf{Z}$, HT at time $t_i\}$ for $t \in (0, 48]$. Putting these two prediction curves together shows the effect of HT on the patient's time to recurrence and aids the decision of whether to give him HT. Uncertainty of the prediction probability, say, $P\{T_i > t_i + t | \mathbf{X}_{obs}, \mathbf{Z}$, no HT in $[t_i, t_i + t]\}$ can be shown by evaluating $P\{T_i > t_i + t | \boldsymbol{\Omega}^{(k)}, D_i^{(k)}, \mathbf{R}_i^{(k)}, \mathbf{x}_{obs}, \mathbf{z}_i\}$ for different draws $\{\boldsymbol{\Omega}^{(k)}, D_i^{(k)}, \mathbf{R}_i^{(k)}\}$ and calculating the variability of the

Table 1. Parameter estimates in the joint models

	Estimate	SD	Estimate/SD
The incidence model			
Intercept	2.465	.607	4.062
$I(T\ stage = 1)$	−1.096	.527	−2.081
$I(T\ stage = 2)$	−1.071	.510	−2.102
$\ln(bPSA + 1)$	.429	.222	1.933
Gleason	.070	.099	.710
Age at RT	−.020	.020	−.994
Total Dose	−.063	.040	−1.568
Duration	−.021	.026	−.803
The failure time model			
$I(T\ stage = 1)$	−1.288	.340	−3.784
$I(T\ stage = 2)$	−.333	.213	−1.564
$\ln(bPSA + 1)$	.010	.105	.099
Gleason	.267	.075	3.552
Total dose	.034	.030	1.121
$PSA(t)$	.569	.070	8.077
Slope	3.878	.620	6.254
HT	−3.115	.487	−6.397
Baseline hazard			
α	1.828	.399	4.578
λ	.001	.001	1.335
θ	.526	.160	3.280
Measurement error			
σ_e	.080	.003	21.11
Tuning parameters			
g for PSA slope	.244	.038	(.17, .32)*
a for HT effect duration†	∞	27.67%	3.820
ν for t error	2.219	.078	(2.07, 2.37)*

*95% confidence intervals.

†Posterior mode, its relative frequency, and relative frequency ratio of the mode over the second most frequently occurring value.

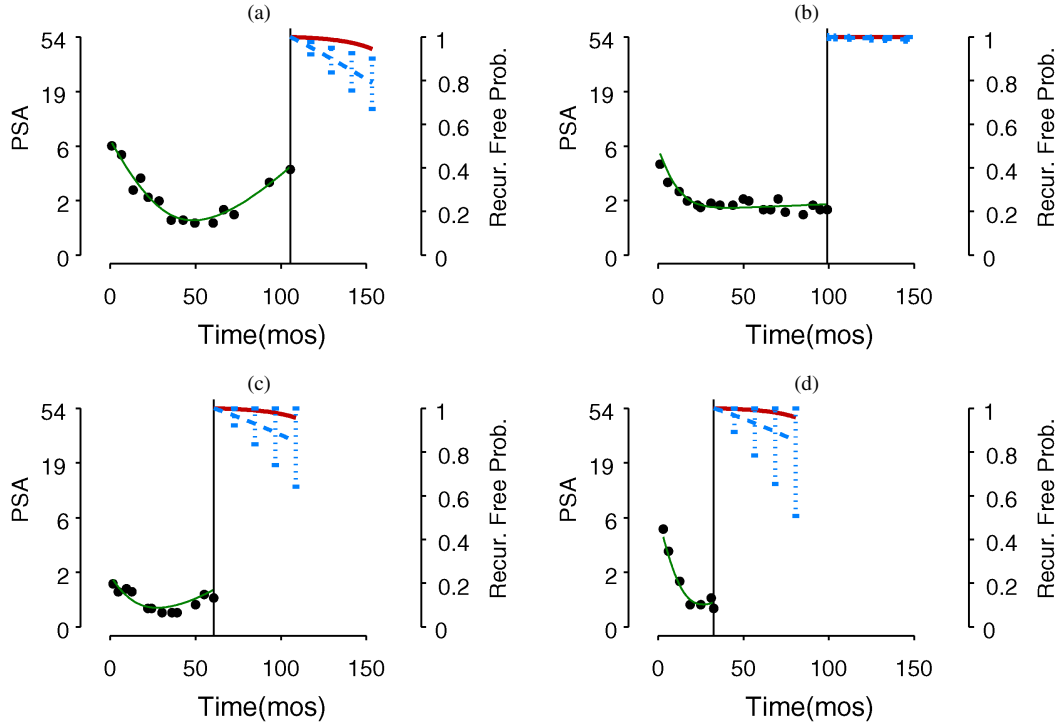

Figure 2. Individual prediction of distribution of time to clinical recurrence (up to 4 years) for four selected censored patients. (– – –, without HT given during the 4 years after the last contact time; ——, with HT given at the last contact time). The left-side vertical axis is shown on a $\log(PSA+1)$-transformed scale. The vertical line indicates the time of last contact. The right-side vertical axis shows the probability of being recurrence free from the date of last contact. The vertical dotted segments represent uncertainty involved with the prediction at years 1, 2, 3, and 4 after the time of last contact.

predicted probability. Figure 2 shows the pattern of PSA and the predicted recurrence-free probability from the date of last contact, with and without the addition of HT at the last contact time, for four selected patients. It also shows the uncertainty range of the prediction of clinical recurrence without HT for years 1, 2, 3, and 4. These uncertainty ranges are the maximum and minimum values of 50 draws from the MCMC.

Patients (a) and (b) have long follow-up times, patient (c) has a medium follow-up time, and patient (d) has a relatively short follow-up time. The patients were selected to illustrate a range of PSA patterns and predictions. The magnitude of the potential impact of HT can be seen. For patient (a), there is a clear pattern of increasing PSA, suggesting eventual clinical recurrence. Patient (a) has a steep rise that leads to a high probability of recurrence within 4 years. For patient (b), the favorable pattern of PSA posttreatment suggests cure, which corresponds to the almost horizontal predicted clinical recurrence curve. There is little probability of recurrence for this patient. Although the PSA values for patient (c) are relatively low, he has a clearly rising pattern, which leads to a moderate probability of recurrence within 4 years (.21). Patient (d) has very short follow-up, leading to considerable range of predicted probabilities of recurrence within 4 years.

The prediction of future PSA values for patients who are censored, are alive, and did not receive HT can be calculated from the draws in the Markov chain. For a posterior draw $\mathbf{R}_i^{(k)}$, the predicted (log-transformed) PSA at time t is $PSA_i^{(k)}(t) = \log(r_{i1}^{(k)} e^{-r_{i2}^{(k)} t} + r_{i3}^{(k)} e^{r_{i4}^{(k)} t} + 1)$. By adding corresponding measurement error $\epsilon_i^{(k)} \sim t_{\nu^{(k)}}\{0, (\sigma_e^2)^{(k)}\}$, a 95% pointwise predictive interval for log-transformed PSA is then formed using the 2.5% quantile to 97.5% quantile of $\{PSA_i^{(k)}(t) + \epsilon_i^{(k)}, k = 1, 2, \ldots, m\}$ for m draws. We use $m = 100$. Examples of these predictive intervals on the original PSA scale for four patients are shown as shaded regions in Figure 2.

Note that the construction of predictive intervals at each future time point is based on the assumption that HT is not given and that clinical recurrence events and death can be eliminated. This is not the same as the assumption that the patient is alive and is not given HT.

We note that patients (a) and (b), who have lots of data, have fairly narrow prediction intervals, whereas patients (c) and (d) have less follow-up and thus wider prediction intervals. We envision that a graph like this also can be useful in monitoring the progression of the patient; for example, a new PSA value measured that falls outside the shaded region is indicative that the patient is doing either worse or better than expected. After a new measurement is obtained, new graphs can be produced, thus giving real-time monitoring of a patient's progression.

7. VALIDATION

7.1 Validation of the Longitudinal Model

The + symbols in the graphs in Figure 3 represent PSA measurements obtained from the validation data set for these 4 patients. All of the values fall within the 95% prediction intervals. Patient (c) had a distant metastasis at 61 months after radiation therapy. Patients (a), (b), and (d) had no clinical recurrence in

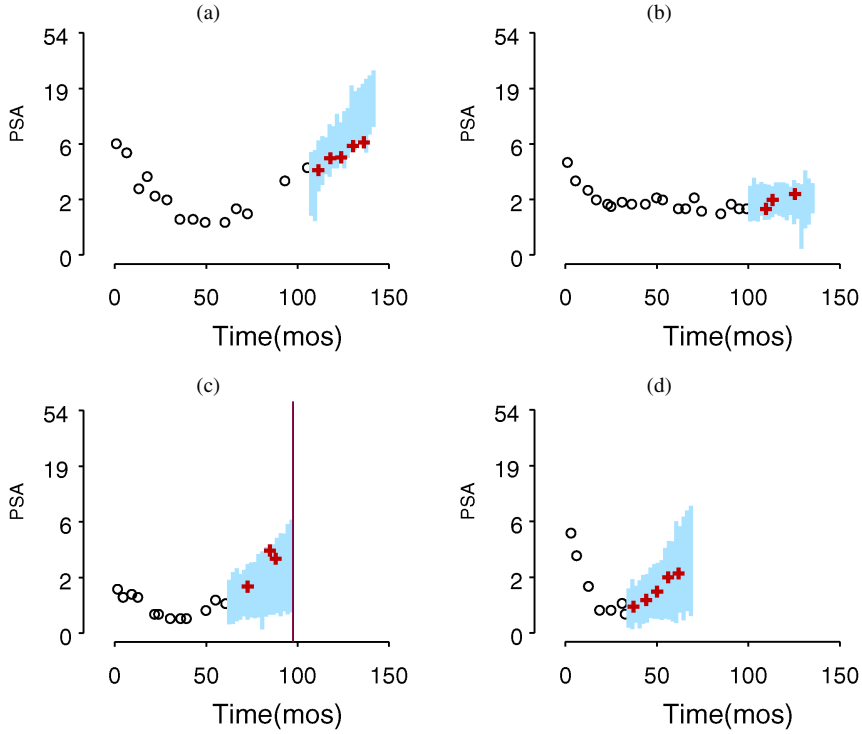

Figure 3. Prediction and validation of the longitudinal model (o, PSA measurements from the analysis data set; +, newly acquired PSA measurements; |, clinical event).

the validation data set in the follow-up period. Table 2 shows the proportion of future PSA values among all available future data that were within the 95% prediction intervals. We see good correspondence with the expected 95% level for all years, although the coverage quality does decrease slightly with time.

7.2 Validation of Conditional Failure Time Model

Table 3 compares the expected events and observed events. For a censored and alive subject i with survival information $(t_i, \delta_i = 0)$ and with follow-up information (t_i^*, δ_i^*), the expected number of events within $(t_i, t_i^*]$ is $P\{T_i < t_i^* | \mathbf{X}_{obs}, \mathbf{Z},$ no HT in $[t_i, t_i^*]\}$. Now to calculate the expected number for 12 months, we obtain the follow-up time for all censored $(\delta_i = 0)$ subjects, $t_i^* - t_i$. If $t_i^* - t_i > 12$, then we set $t_0 = 12$; otherwise, we use $t_0 = t_i^* - t_i$ in $P\{T_i > t_i + t_0 | \mathbf{X}_{obs}, \mathbf{Z},$ no HT in $[t_i, t_i^*]\}$. By summing over i, we get the expected number of events in $0 \sim 1$ year. Similarly, we can calculate expected number of events for $0 \sim 2$ years and $0 \sim 3$ years. Table 3 shows that the observed number of recurrence events is much lower than the expected number of events for all periods. One reason for this is that some patients get HT because of elevated PSA. But if we count HT as failure, then the numbers are much closer.

Another way to validate the survival model is to calculate

$$P\{T_i \le t_i + 36 | \mathbf{X}_{obs}, \mathbf{Z}, \text{ no HT in } [t_i, t_i + 36]\},$$

the probability of recurrence within 3 years after the last contact time in the analysis data for any censored and alive patient who had no HT before t_i, and then compare this value with observed recurrence or HT. Figure 4 shows the calculated Kaplan–Meier estimate of the 3-year recurrence or HT probability for five groups categorized by the estimated probability of recurrence within 3 years. The results show a larger proportion of recurrences or HT patients in the groups with the higher predicted probability, providing support for the validity of the model.

7.3 Sensitivity to Priors and Model Assumptions

Because of the large number of patients and longitudinal observations in this study, the posterior distributions of population parameters are dominated by data. Thus we expect our results to be quite robust to most prior specifications. Our limited experience with various prior specifications confirm that this is indeed the case for most parameters. In an attempt to accommodate the possibility of PSA increase in the cured group, we assumed a normal distribution for R_{i4} in the cured group. This

Table 2. Prediction and validation of the longitudinal model

Yearly intervals	Total number of PSAs	Above 97.5%	Between 2.5% and 97.5%	Below 2.5%
0–1 year	281	2.5%	96.5%	1.9%
1–2 years	407	4.3%	91.6%	3.8%
2–3 years	304	5.9%	90.9%	3.1%

Table 3. Comparison of the expected events and observed events

	0–1 year	0–2 years	0–3 years
Expected number of events	11	21	27
Observed number of recurrence	6	10	11
Observed number of recurrence or HT	10	20	24

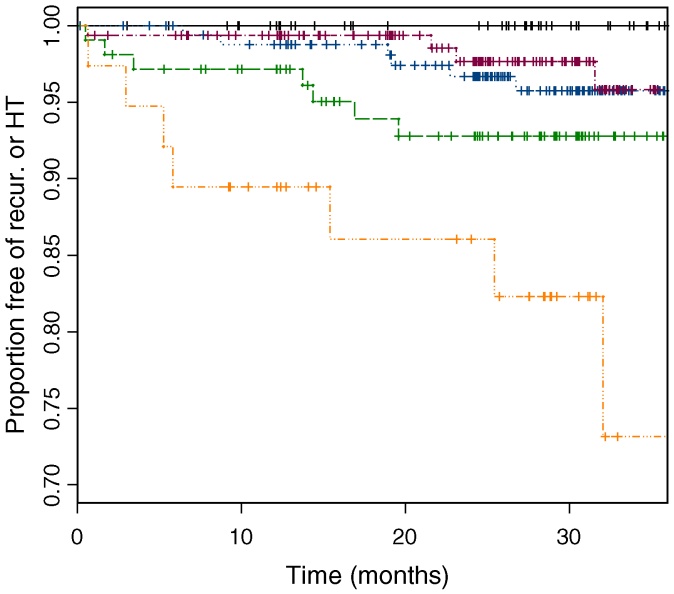

Figure 4. Kaplan–Meier estimate, from the validation data set of the probability of being free of recurrence or HT. Patients are categorized into five groups by the estimated probability of recurrence within 3 years from the last contact time in the analysis data [——, $0 \sim .01$ ($n = 64$); -·-·-, $.01 \sim .02$ ($n = 168$); ·····, $.02 \sim .05$ ($n = 160$); – – –, $.05 \sim .20$ ($n = 106$); -··-··-, $.20 \sim 1.00$ ($n = 38$)]. The number of patients in each group are listed.

extends the work of Law et al. (2002), where $r_{i4} = \exp(R_{i4})$ is set to be 0. However, we suspect that the variance term σ_{44} for the distribution of R_{i4} in the cured group, together with the prior mean -6, might have some effect on the classification of the cure status of patients and hence affect the estimates of population parameters. We set a fairly informative prior for σ_{44} to restrict the variation of R_{i4} for the random effects of the cured group. Also the mean -6 is chosen rather empirically. These are just some possible ways to separate into states (cured or not cured) for the patients. If we had made the prior for σ_{44} less restrictive and the prior mean of R_{i4} larger, then we would have had more patients with the probability of cure closer to .5, because the random-effects model under both groups would fit the data equally well. On the other hand, if we had set the prior mean of R_{i4} even smaller, then we would have had a more unequivocal estimate of probability of cure; that is, many patients would have had an estimated probability of cure either closer to 1 or 0. In our experience, various values for both the prior for σ_{44} and prior mean of R_{i4} had very little effect on the population parameters of the incidence and failure time models. The prior mean, but not the prior distribution, of σ_{44} affects individual prediction of cure status for some patients. For example, by setting -5 as the prior mean, we have 57 with predicted cure probability $\leq .1$, 137 with predicted cure probability $\geq .9$, and 103 with predicted cure probability between .45 and .55 based on posterior draws of D, whereas by setting -7 as the prior mean, we have 71 with predicted cure probability ≤ 0.1, 148 with predicted cure probability $\geq .95$, and 96 with predicted cure probability between .45 and .55. However $>86\%$ of the patients have a difference of predicted cure probability $<.1$ under the two different prior means. For the prediction distribution of cancer recurrence as described in Section 6, we find that prediction curves within 3 years are nearly always very similar under the assumption of different prior means of R_{i4}. We note that the normality assumption for these random effects may be restrictive and can be relaxed to a nonnormal distribution such as a t distribution or a skewed distribution. This has been proposed by Song, Davidian, and Tsaitis (2002) in the joint modeling setting without cure component.

The advantage of adding a cure model to the joint modeling setting is that it provides a way to model the heterogeneity due to the potential existence of long-term survivors and hence provide more accurate results. To assess the need for a cure model, we compare prediction results under the proposed model to one without a cured fraction. The results for the longitudinal model validation are comparable; however, for the conditional failure time model validation, the model without a cured fraction predicts 12, 24, and 32 events in 0–1 year, 0–2 years, and 0–3 years, respectively, compared with the values of 11, 21, and 27 in Table 3; that is, it overestimates the number of events. We also calculated two well-known Bayesian model selection criteria to assess whether or not the model needed a cure component for survival data. One of these is the conditional predictive ordinate (CPO). For a specific subject i under model M_r, the CPO is defined by $CPO_i^{(r)} = f_r(\mathbf{x}_{i,obs}|\mathbf{X}_{(i),obs})$, the conditional density of the observed data for subject i, $\mathbf{x}_{i,obs}$, given the observed data $\mathbf{X}_{(i),obs}$ for all subjects except i. A useful summary statistic of the $CPO_i^{(r)}$'s, termed the "pseudo-Bayes factor," is (Geisser and Eddy 1979; Gelfand, Dey, and Chang 1992)

$$B^* = \sum_{i=1}^{n} \log\bigl(CPO_i^{(r)}\bigr). \tag{9}$$

When comparing two models, the model with the larger value of B^* is the better-fitting model. The $CPO_i^{(r)}$ can be approximated from MCMC output by a harmonic mean formula (Gelfand 1995),

$$\widehat{CPO}_i^{(r)} = \left\{ \frac{1}{K} \sum_{k=1}^{K} \frac{1}{f_r(\mathbf{x}_{i,obs}|\boldsymbol{\Omega}_r^{(k)})} \right\}^{-1},$$

Computating this approximation involves evaluating $f_r(\mathbf{x}_{i,obs}|\boldsymbol{\Omega}_r^{(k)})$ for each draw of the population parameters $\boldsymbol{\Omega}_r$. To efficiently approximate these integrals, we used an iterative quadrature strategy similar to that proposed by Naylor and Smith (1982). Details have been given by Yu (2004). Applying the CPO criteria to our prostate cancer study, we found values of the summary statistic $\sum_{i=1}^{n} \log(CPO_i^{(r)})$ of 253.41 for the joint model with a cure component and $-2{,}125.10$ for the joint model without a cure component. This strongly supports the need for a cure component in the model.

Because harmonic mean formula can be unstable for calculation, due to the fact that it involves inverse of densities, we also used the Bayes information criterion (BIC) to compare the two models. For the BIC, we computed the observed likelihood at posterior modes of population parameters $\hat{\Omega}^{(r)}$ for each model. Due to the presence of random effects, we computed the observed likelihoods using the method proposed by Chib and Jeliazkov (2004). Basically, by treating random effects as parameters, the observed likelihood at $\hat{\Omega}^{(r)}$ follows the basic marginal

likelihood identity (Chib 1995; Chib and Jeliazkov 2004); for example, for the joint model with a cure component, we have

$$\begin{aligned}\log f(\mathbf{X}_{obs}|\mathbf{Z}, \hat{\Omega}) = \log f(\mathbf{X}_{obs}|\mathbf{Z}, \hat{\mathbf{D}}, \hat{\mathbf{R}}, \hat{\boldsymbol{\xi}}, \hat{\Omega}) \\ + \log f(\hat{\mathbf{D}}, \hat{\mathbf{R}}, \hat{\boldsymbol{\xi}}|\mathbf{Z}, \hat{\Omega}) \\ - \log f(\hat{\mathbf{D}}, \hat{\mathbf{R}}, \hat{\boldsymbol{\xi}}|\mathbf{X}_{obs}, \mathbf{Z}, \hat{\Omega}),\end{aligned}$$

where the random-effects parameters are taken to be at their posterior modes as well. Evaluation, of the first two terms is obvious, because these terms are readily available from model specifications. To calculate the third term, we used the approach of Chib and Jeliazkov (2004). We found observed log-likelihoods of 1,945.44 for the joint model with a cure component and 1,679.86 for the joint model without a cure component, with corresponding population parameters of 74 and 44. The resulting BIC is 326.17, favoring the need for a cure component in the model.

8. DISCUSSION

In prostate cancer, if the cancer cells are confined to the organ, there is high likelihood of killing these cancer cells by radiation and hence curing the patient of prostate cancer. These patients will not experience recurrence of cancer, and their probability of having recurrence is 0. However if cancer cells are not confined to the organ or not completely killed by radiation, then the patient is subject to recurrence. This is the biological reason for including a cure component in the model. The empirical evidence that a cure model may be appropriate is the relatively low number of recurrences (146 of 928 patients) despite the long follow-up time.

One issue that arises, because of patients who received HT due to elevated PSA, is dependent censoring. Had such patients not received HT, they very likely would have experienced cancer recurrence soon. The effect of HT postponed the time to recurrence. We included HT as a time-dependent covariate in a hazard model. This may not be quite correct from a causal inference standpoint without further assumptions (see, e.g., Robins 1997), but it may be satisfactory for predictions. The decision to give HT is usually based on the value and slope of PSA, and these two variables are already in the model. Thus adding HT to the model helps reduce the possible bias, because for those patients, the observed time is delayed by HT.

Our choice of a parametric form for the baseline hazard facilitates individual predictions. Smoothness of the hazard is a realistic and desirable feature for predictions that may require extrapolation to the maximum follow-up time. In contrast, with a totally nonparametric baseline hazard function (Law et al. 2002), the estimate would not be smooth and, we would need to assume a form for the hazard beyond the last failure time. However, we note that for estimation of parameters, a nonparametric assumption for the baseline hazard may be preferred. Law et al. (2002) took the baseline hazard to be nonparametric, and used a maximum likelihood estimation method. Their estimation results are very similar to those achieved when a Weibull baseline hazard was assumed (Yu et al. 2004). The extension from a Weibull baseline hazard to a generalized Weibull baseline hazard, as used in this article, allows more flexibility of the model. An alternative approach would be to consider the baseline hazard as a stochastic process, for example, as a gamma process (Sinha and Dey 1997).

With highly parameterized models, interpretation of the parameters can be difficult, and there can be identifiability problems with cure models (Farewell 1986; Li, Taylor, and Sy 2001). In addition, the slowly 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; however, prediction results and use of the BIC both suggest that models including a cured component are likely a better fit to the data. The addition of the cure model component also could just be viewed as a means of way to formulate a richer and more flexible class of models. From this standpoint, the specific parameters of the cure model are of less interest, whereas quantities such as the predictive survival distribution are inherently interpretable irrespective of the choice of model.

[Received May 2004. Revised December 2006.]

REFERENCES

Bagdonavicius, V., and Nikulin, M. (2001), *Accelerated Life Models: Modeling and Statistical Analysis*, London: Chapman & Hall.

Brown, E. R., and Ibrahim, J. G. (2003), "Bayesian Approaches to Joint Cure-Rate and Longitudinal Models With Application to Cancer Vaccine Trials," *Biometrics*, 59, 686–693.

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," *Statistics in Medicine*, 17, 2061–2077.

Casella, G., and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (1999), "A New Bayesian Model for Survival Data With a Survival Fraction," *Journal of the American Statistical Association*, 94, 909–919.

——— (2004), "A New Joint Model for Longitudinal and Survival Data With a Cure Fraction," *Journal of Multivariate Analysis*, 91, 18–34.

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–21.

Chib, S., and Jeliazkov, I. (2004), "Marginal Likelihood From the Metropolis–Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

Farewell, V. T. (1982), "The Use of Mixture Models for the Analysis of Survival Data With Long-Term Survivors," *Biometrics*, 38, 1041–1046.

——— (1986), "Mixture Models in Survival Analysis: Are They Worth the Risk?" *The Canadian Journal of Statistics*, 14, 257–262.

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.

Geisser, S., and Eddy, W. (1979), "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74, 153–160.

Gelfand, A. E. (1995), "Model Determination Using Sampling-Based Methods," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilds, S. Richardson, and D. J. Spiegelhalter, London: Chapman & Hall, pp. 145–161.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination Using Predictive Distributions With Implementation via Sampling-Based Methods," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 147–167.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Gilks, W., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Henderson, R., Diggle, P., and Dobson, A. (2000), "Joint Modeling of Longitudinal Measurements and Event Time Data," *Biostatistics*, 1, 465–480.

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), New Jersey: Wiley.

Kuk, A. Y. C., and Chen, C. H. (1992), "A Mixture Model Combining Logistic Regression With Proportional Hazards Regression," *Biometrika*, 79, 531–541.

Lange, K., Little, R. J. A., and Taylor, J. M. G. (1989), "Robust Statistical Modeling Using the *t* Distribution," *Journal of the American Statistical Association*, 84, 881–896.

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.

Li, C. S., Taylor, J. M. G., and Sy, J. P. (2001), "Identifiability of Cure Models," *Statistics and Probability Letters*, 54, 389–395.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machine," *Journal of Chemistry in Physics*, 21, 1087–1091.

Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics*, 31, 214–255.

Pauler, D. K., and Finkelstein, D. M. (2002), "Predicting Time to Prostate Cancer Recurrence Based on Joint Models for Non-Linear Longitudinal Biomarkers and Event Time Outcomes," *Statistics in Medicine*, 21, 3897–3911.

Robins, J. M. (1997), "Structural Nested Failure Time Models," in *The Encyclopedia of Biostatistics*, eds. P. Armitage and T. Colton, Chichester, U.K.: Wiley, pp. 4372–4389.

Sandler, H. M., Dunn, R. L., McLaughlin, P. W., Hayman, J. A., Sullivan, M. A., and Taylor, J. M. G. (2000), "Overall Survival After PSA-Detected Recurrence Following Conformal Radiation Therapy," *International Journal of Radiation Oncology, Biology and Physics*, 48, 629–633.

Sinha, D., and Dey, D. K. (1997), "Semiparametric Bayesian Analysis of Survival Data," *Journal of the American Statistical Association*, 92, 1195–1211.

Skates, S. J., Pauler, D. K., and Jacobs, I. J. (2001), "Screening Based on the Risk of Cancer Calculation From Bayesian Hierarchical Change-Point Models of Longitudinal Markers," *Journal of the American Statistical Association*, 96, 429–439.

Slate, E. H., and Cronin, K. A. (1997), "Changepoint Modeling of Longitudinal PSA as a Biomarker for Prostate Cancer," in *Case Studies in Bayesian Statistics III*, New York: Springer-Verlag, pp. 435–456.

Song, X., Davidian, M., and Tsiatis, A. A. (2002), "A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data," *Biometrics*, 58, 742–753.

Taylor, J. M. G. (1995), "Semi-Parametric Estimation in Failure Time Mixture Models," *Biometrics*, 51, 899–907.

Tsiatis, A. A., and Davidian, M. (2004), "An Overview of Joint Modeling of Longitudinal and Time-to-Event Data," *Statistica Sinica*, 14, 809–834.

Tsiatis, A. A., DeGruttola, V., and Wolfsohn, 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.

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*, 96, 895–905.

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," *Applied Statistics*, 50, 375–387.

Yakovlev, A. Y., and Tsodikov, A. D. (1996), *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, River Edge, NJ: World Scientific.

Yu, M. (2004), "Joint Longitudinal-Survival-Cure Model With Application in Prostate Cancer Studies," unpublished doctoral thesis, University of Michigan.

Yu, M., Law, N. J., Taylor, J. M. G., and Sandler, H. M. (2004), "Joint Longitudinal-Survival-Cure Models and Their Application to Prostate Cancer," *Statistica Sinica*, 14, 809–834.

Zagars, G. K., and Pollack, A. (1993), "The Fall and Rise of Prostate-Specific Antigen," *Cancer*, 72, 832–842.