Biometrics 69, 206–213
March 2013

DOI: 10.1111/j.1541-0420.2012.01823.x

Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models

Jeremy M. G. Taylor,[1,*] Yongseok Park,[1] Donna P. Ankerst,[2,3] Cecile Proust-Lima,[4] Scott Williams,[5] Larry Kestin,[6] Kyoungwha Bae,[7] Tom Pickles,[8] and Howard Sandler[9]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
[2]Zentrum Mathematik, Technische Universitaet Muenchen, Munich 85748, Germany
[3]Department of Epidemiology/Biostatistics and Urology, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, U.S.A.
[4]INSERM, U897, Epidemiology and Biostatistics Research Center, 33076 Bordeaux, France
[5]Division of Radiation Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria 3002, Australia
[6]Michigan Healthcare Professionals/21st Century Oncology, Farmington Hills, Michigan 48334, U.S.A.
[7]Novartis Pharmaceuticals Corporation, Florham Park, NJ 07932, U.S.A.
[8]British Columbia Cancer Agency, Vancouver, British Columbia V5Z 4E6, Canada
[9]Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, California 90048, U.S.A.
[*]*email*: jmgt@umich.edu

Summary. Patients who were previously treated for prostate cancer with radiation therapy are monitored at regular intervals using a laboratory test called Prostate Specific Antigen (PSA). If the value of the PSA test starts to rise, this is an indication that the prostate cancer is more likely to recur, and the patient may wish to initiate new treatments. Such patients could be helped in making medical decisions by an accurate estimate of the probability of recurrence of the cancer in the next few years. In this article, we describe the methodology for giving the probability of recurrence for a new patient, as implemented on a web-based calculator. The methods use a joint longitudinal survival model. The model is developed on a training dataset of 2386 patients and tested on a dataset of 846 patients. Bayesian estimation methods are used with one Markov chain Monte Carlo (MCMC) algorithm developed for estimation of the parameters from the training dataset and a second quick MCMC developed for prediction of the risk of recurrence that uses the longitudinal PSA measures from a new patient.

Key words: Joint longitudinal-survival model; Online calculator; Predicted probability; Prostate cancer; PSA.

1. Introduction

Joint models of longitudinal and survival data, which have been much researched in recent years (Henderson, Diggle, and Dobson, 2000; Lin et al., 2002; Yu et al., 2004), are applicable in situations where both the longitudinal data and the survival time data are considered as response variables. Many different model formulations have been suggested and different estimation methods proposed. The most common formulation is a random effects model for continuous longitudinal data, and a proportional hazards model for the event time data. The covariates in the hazard model typically include the random effects or functions of the random effects, thus linking the two models. A representative set of publications in this area is Henderson et al. (2000), Xu and Zeger (2001), and Wang and Taylor (2001), with a review given in Yu et al. (2004). Joint models can be used for different purposes, one is estimation of the parameters in the longitudinal model and another is estimation of the parameters in the survival model. A different use of these models is for prediction purposes (Taylor, Yu, and Sandler, 2005; Yu, Taylor, and Sandler, 2008; Proust-Lima and Taylor, 2009; Rizopoulos, 2011; Proust-Lima, Sène, Taylor, and Jacqmin-Gadda, 2012) and that will be the focus in this article.

The application of the methods we will describe comes from the prostate cancer setting. After initial diagnosis of prostate cancer and subsequent treatment by radiation therapy, patients are typically monitored at regular intervals using the Prostate Specific Antigen (PSA) blood test. If during the follow-up the values of PSA start to increase, this may be an early indication that the cancer is growing or spreading within the patient's body, but has not grown or spread enough to be clinically detectable by other means, such as imaging methods or biopsy. If the cancer does grow enough to be detectable through these means, then the patient is said to have clinical recurrence with the time of recurrence noted as the time of detection. Thus higher levels of PSA or high rates of increase of PSA indicate an increased risk for clinical recurrence. If PSA is starting to rise, the patient may wish to start a treatment

which is known to slow down the growth of the cancer. Specifically they may elect to initiate salvage androgen deprivation therapy, also called hormone therapy, in order to prevent or delay the recurrence of cancer. However, the hormone therapy has some side effects, so if the risk of recurrence is low, the patient may opt not to start hormone therapy. Thus providing the patient with an accurate estimate of the risk of clinical recurrence would aid in the decision of whether or not to start hormone therapy. The purpose of this article is to describe the method we used to calculate the risk of future recurrence for an individual patient, given the patient's history of PSA values up to the present time. The method will utilize a joint longitudinal survival model.

We have implemented this method in a calculator that is available on the web at `http://psacalc.sph.umich.edu`. This calculator requires as input the patient's baseline variables and their longitudinal sequence of PSA values, then it gives as output predictions of future PSA values and estimates of the probability of clinical recurrence of prostate cancer up to 3 years in the future assuming the person does not initiate hormone therapy. We choose a relatively short time frame of 3 years because a longer time frame is not necessary as the patient would be expected to have another PSA measurement within 3 years, at which point the risk calculation can be updated.

Developing a calculator for anonymous use by others raises some statistical issues beyond the usual ones that statisticians face when analyzing and interpreting data. One issue is that the calculation has to be fast if it is going to be used in a clinical setting. Another issue is that we need to be confident that the model is applicable more generally than just on the datasets that have been used to develop it. Another issue is that we felt it important to alert the user if he was trying to use the model outside of the range of values for the input data that had been used to develop the model, or if it had appeared that the user had input erroneous data.

In Section 2 of this article, we describe the joint model and the datasets. In Section 3, we present the Bayesian estimation method. In Section 4, we describe the method that is used for individual prediction and illustrate the predictions for different patients. In Section 5, we describe the warning messages that are built into the web-based calculator. In Section 6, we describe validation of the model, and we end with a discussion.

2. Datasets, Notation, and Model

We used four datasets, all consisting of patients treated with radiation therapy, without surgery, for localized prostate cancer. We combined three of the datasets as a training dataset: these were from the University of Michigan (503 patients), the Radiation Therapy Oncology Group (RTOG) (615 patients), and Beaumont Hospital (1268 patients). Data from Vancouver (846 patients) were held back as a separate testing dataset. For all except the Vancouver data the basic patient characteristics and response variable summaries are given in Proust-Lima et al. (2008) and will not be repeated here. A general description of the Vancouver data is given in Pickles et al. (2003). Overall in the four datasets there were a total of 3232 patients of which 458 had a clinical recurrence. Only data up to the time of clinical recurrence is used. Of the 3232 patients 391 did receive hormone therapy prior to clinical recurrence. The average number of postradiation PSA measurements was 9.2 and the average follow-up time was 5.8 years, with a maximum of 15.4 years. The distribution of the patient characteristics and follow-up time is similar in the training and testing datasets.

The datasets have the following structure. All patients are diagnosed with localized prostate cancer and initially treated with external beam radiation therapy only. Patients have pretreatment characteristics. The three variables used were T-stage, Gleason grade, and baseline PSA. Each patient has a sequence of values of PSA after the radiation therapy and these are used to monitor the patient. These are denoted by $(Y_{i1}, \ldots, Y_{im_i})$ measured at times $(t_{i1}, \ldots, t_{im_i})$ for subject i. Time is measured in years from the end of radiation therapy. The typical pattern of PSA after radiation therapy is well known. It decreases in everyone for about a year and then may or may not start to rise; if it does rise, it increases approximately exponentially over time. Rising values of PSA are indicative of tumor cells growing and dividing, but the tumor may not have yet grown to such a size that it is clinically detectable. The time of clinical recurrence is the time when the tumor is clinically detected, denoted by R_i, and that is the event of interest. The recurrence can happen locally, regionally, or as distant metastases, and R_i is the first occurrence of any of these. Let δ_i be the associated censoring indicator. Excluded from the datasets are patients with R_i less than one year, because such patients very likely had latent metastatic disease at the time of diagnosis and radiation therapy will not be effective. If the values of PSA start to rise, and particularly if they rise steeply, some patients do begin hormone therapy prior to any clinical recurrence; we denote the time of initiating hormone therapy as H_i. Hormone therapy quickly reduces the values of PSA in just about all patients, and to near zero in most patients, but later PSA may rise and the patient may experience clinical recurrence. In none of the modeling or analysis we undertake do we consider the observed values of PSA after H_i, but we do consider the clinical recurrences after H_i.

The joint model has two components, a longitudinal model and a survival model. The longitudinal model is a random effects model with three random components with a multivariate normal distribution, fixed effects, and a T distribution with 5 degrees of freedom for the measurement error. A T distribution is used to accommodate possible outliers. The survival model is a time-dependent proportional hazards model where the covariates are the baseline variables, time-dependent PSA and rate of change of PSA, and an indicator for hormone therapy. These models have been developed over a number of years, and they are similar to, but not identical to, models we have used in the past for similar data (Yu et al., 2004; Proust-Lima and Taylor, 2009). The main structure of the model and choice of covariates in different parts of the model closely follow those developed in Proust-Lima et al. (2008). In that paper five datasets were considered, which included our three training datasets. Separate longitudinal and survival models were fit using maximum likelihood methods to those data, both as individual datasets and as pooled data. The exact details of the model used in the current article are given below.

The assumed model for PSA in the absence of treatment by hormone therapy is:

$$\mathrm{Y}_i(t) = \mathrm{EY}_i(t) + \epsilon_{it},$$

where $Y_i(t)$ are the observed values of $\log(\mathrm{PSA} + 0.1)$ for subject i at time t, and where $\epsilon_{it} \sim T(0, \sigma^2, 5)$, the central T distribution with scale parameter σ and 5 degrees of freedom. The expected value of Y is given by

$$\mathrm{EY}_i(t) = \beta_{0i} + \beta_{1i}\{(1+t)^{-1.5} - 1\} + \beta_{2i}t, \qquad (1)$$

where $\boldsymbol{\beta}_i^T = (\beta_{0i}, \beta_{1i}, \beta_{2i}) \sim N_3(\boldsymbol{x}_i^T\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{3\times 3})$ are subject-specific random effects, $\boldsymbol{\alpha}$ are fixed effect parameters, $\boldsymbol{x}_i$ is a design matrix including baseline covariates. The choice of covariates is described below.

In equation (1), $(1+t)^{-1.5} - 1$, a Box–Cox transformation, captures the short-term evolution of PSA, while t captures the long-term evolution. The justification for the linear term in t on the log scale is the following. In addition to the empirical observation that log PSA values increase linearly with time, the PSA value also approximately reflects the number of tumor cells, and cancer cells divide approximately exponentially, leading to a linear term on the log scale. For the β_{1i} term, the initial drop in PSA following radiation therapy, Ankerst and Finkelstein (2006) have used $\log(1+t)$. We considered generalizations of this and using just the longitudinal data and mixed models from the five datasets pooled as described in Proust-Lima et al. (2008), we considered Box–Cox transformations of the form $(1+t)^{\lambda} - 1$. We estimated λ in this model using profile likelihood methods and found that the MLE was close to -1.5.

The specific covariates included for the three random effects in the longitudinal model are (i) an intercept and Baseline PSA expressed as $\log(\mathrm{basePSA} + 0.1)$ for the β_{0i} term, (ii) an intercept, $\log(\mathrm{basePSA} + 0.1)$, I(T-Stage $= 2$) and I(T-Stage ≥ 3) for the β_{1i} term, and (iii) an intercept, $\log(\mathrm{basePSA} + 0.1)$, I(T-Stage $= 2$), I(T-Stage ≥ 3), I(Gleason $= 7$) ,and I(Gleason ≥ 8) for the β_{2i} term. The choice of which covariates to include in each of the three terms was based on the findings in Proust-Lima et al. (2008), where separate analyses of each of the datasets found consistent patterns of parameter estimates. The covariates for which the estimates were consistently close to zero were excluded in the current model.

The model for the hazard of clinical recurrence at time t for the ith person is:

$$\lambda_i(t) = \lambda_0(t)\exp[\boldsymbol{Z}_i(t)^T\boldsymbol{\theta}], \qquad (2)$$

where

$$\begin{aligned}\boldsymbol{Z}_i(t)^T\boldsymbol{\theta} = [&\boldsymbol{\theta}_0^T\boldsymbol{W}_i + \theta_6 \mathrm{logit}^{-1}(\mathrm{EY}_i(t) - 0.7)/0.45 \\ &+ \theta_7 \mathrm{EY}_i'(t) + \theta_8 I(t \geq H_i)],\end{aligned} \qquad (3)$$

where H_i is the time of hormone therapy and $\mathrm{EY}_i'(t)$ is the derivative of $\mathrm{EY}_i(t)$. A piecewise constant function is assumed for $\lambda_0(t)$ with jumps at $t = 0.95, 2, 3, 5, 7$, and 30 years. There is zero hazard prior to $t = 0.95$ and the five levels of the baseline hazard are denoted by $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{05})$. The knot locations were chosen to give similar numbers of events in each interval. The covariates $\boldsymbol{W}_i$ in the survival model are $\log(\mathrm{basePSA} + 0.1)$, I(T-Stage $= 2$), I(T-Stage ≥ 3), I(Gleason $= 7$), and I(Gleason ≥ 8), and $\boldsymbol{\theta}_0$ is the associated five-dimensional vector of hazard ratios. We did not consider interactions between t and any of $\boldsymbol{W}_i$, $\mathrm{EY}(t)$, or $\mathrm{EY}_i'(t)$, although in principle that would be possible. We note in this model that $\mathrm{EY}(t)$ and $\mathrm{EY}_i'(t)$ are the expected values of PSA and the slope of PSA under the assumption that hormone therapy had not been given. The term $I(t \geq H_i)$ is included to account for the fact that the hazard after hormone therapy is given is less than it would have been had the hormone therapy not been given.

The specific form of the covariates in equation (3) was again arrived at based on the work in Proust-Lima et al. (2008). In that paper we used the longitudinal mixed model to impute values of PSA and the slope of PSA, then using a standard Cox model with partial likelihood estimation, it was deemed that all the baseline covariates W_i were important. Possible transformations of the covariates PSA and slope of PSA were considered. The initial step was to discretize the covariates into ranges and consider step functions, and compare the likelihood with a standard model. No transformation was deemed necessary for slope of PSA, but a sigmoid transformation gave a better fit for PSA. Then using profile likelihood we found that coefficients of 0.7 and 0.45 as shown in equation (3) gave a good fit.

The set of all population parameters is denoted by $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \sigma^2, \lambda_0)$, where $\boldsymbol{\theta}' = (\boldsymbol{\theta}_0', \theta_6, \theta_7, \theta_8)$.

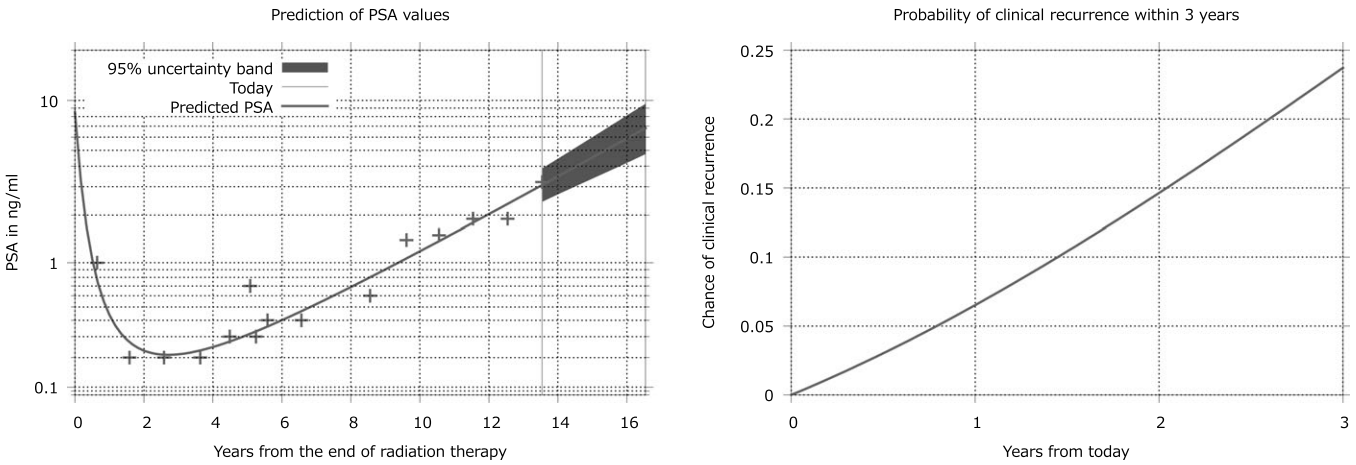
3. Estimation Methods

Estimation was performed using a Bayesian approach via implementation of a Markov chain Monte Carlo (MCMC) algorithm. Upon convergence, the chain delivers draws from the posterior distribution of all the parameters and of the three random effects for all subjects. Details of the algorithm and the priors used are given in the Supplementary Material. The program was written in `C` and needed to be run for many hours to be confident that the chain had converged. Most of the parameters converged very quickly, the only exceptions were the θ parameters which could be slow to converge. Upon convergence, we saved 1000 draws for each of the parameters, which will be used later for individual predictions. Table 1 in the online Supplementary material gives estimates of the main parameters in the model.

4. Individual Predictions

Predictions of future PSA values and distributions of time to recurrence depend on the parameters and the random effects for that person. Consider a subject N who has available longitudinal PSA measurements ($\mathbf{Y}$) up to time c, who has not had hormone therapy or a recurrence prior to c, and let $\boldsymbol{W}_N$ denote his baseline covariates. If we know the random effects ($\boldsymbol{\beta}_N$) for this person and the population parameters ($\boldsymbol{\phi}$), then the future PSA value at time t is predicted by $E(\mathrm{Y}(t)) = \boldsymbol{\beta}_{N0} + \boldsymbol{\beta}_{N1}\{(1+t)^{-1.5} - 1\} + \boldsymbol{\beta}_{N2}t$. Similarly the hazard is $\lambda_N(t) = \lambda_0(t)\exp\{\boldsymbol{Z}(t)^T\boldsymbol{\theta}\}$ with the term $I(t \geq H)$ set to zero. From this we calculate the residual time distribution

$$P(R > t \,|\, \boldsymbol{W}_N, \boldsymbol{\beta}_N, \boldsymbol{\phi}, R > c) = \exp\left\{-\int_c^t \lambda_N(u)du\right\}, \qquad (4)$$



(a) Prediction of PSA values

(b) Probability of clincial recurrence

Figure 1. (a) Observed and predicted PSA values (on log scale) for years from the end of radiation therapy and (b) probability of clinical recurrence for years starting at the end of PSA follow-up.

where $\lambda_N(u)$ is given by equation (2). In this expression the integral is calculated numerically using Gaussian quadrature. If we have M draws of $\boldsymbol{\beta}_N$ and $\boldsymbol{\phi}$ from the posterior distribution using the training data set, then the final estimated residual time distribution $P(R > t|\boldsymbol{W}_N, R > c)$ is the average of the M separate $P(R > t|\boldsymbol{W}_N, \boldsymbol{\beta}_N, \boldsymbol{\phi}, R > c)$ estimates. Alternatively, one could draw one value of R from each distribution $P(R|\boldsymbol{W}_N, \boldsymbol{\beta}_N, \boldsymbol{\phi}, R > c)$ for each draw of $(\boldsymbol{\beta}_N, \boldsymbol{\phi})$ to construct an empirical histogram of the residual time distribution.

If person N is part of the training dataset, then the saved values for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}_N$ can be used. If the person is not part of the training dataset, one strategy would be to add this person to the dataset and rerun the program, but this would not be feasible for an online calculator. Instead as an approximation we use the saved values of $\boldsymbol{\phi}$, and estimate $\boldsymbol{\beta}_N$ from this new person's data. Estimation of the random effects for new person N is achieved by running a quick MCMC, where the likelihood is for the data for this new person, and the prior distribution for the population parameters comes from the posterior distribution from the training datasets analysis. For each of the saved values of $\boldsymbol{\phi}$ the MCMC is run for 50 iterations and the final value of the three random effects saved. This is repeated for the 1000 saved values of $\boldsymbol{\phi}$. Because there are only three quantities to estimate for each person, we found that 50 iterations was sufficient for convergence. Complete details of the algorithm are given in the online Supplementary materials.

The prediction of future PSA values, given by $\mathrm{EY}(t) = \beta_{N0} + \beta_{N1}\{(1+t)^{-1.5} - 1\} + \beta_{N2}t$, similarly uses the drawn values of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}_N$. This leads to a set of 1000 curves, from which at each time we report the pointwise median and the 2.5th and 97.5th percentiles. The computational time to produce the predictions for a range of values of t is about 2 seconds.

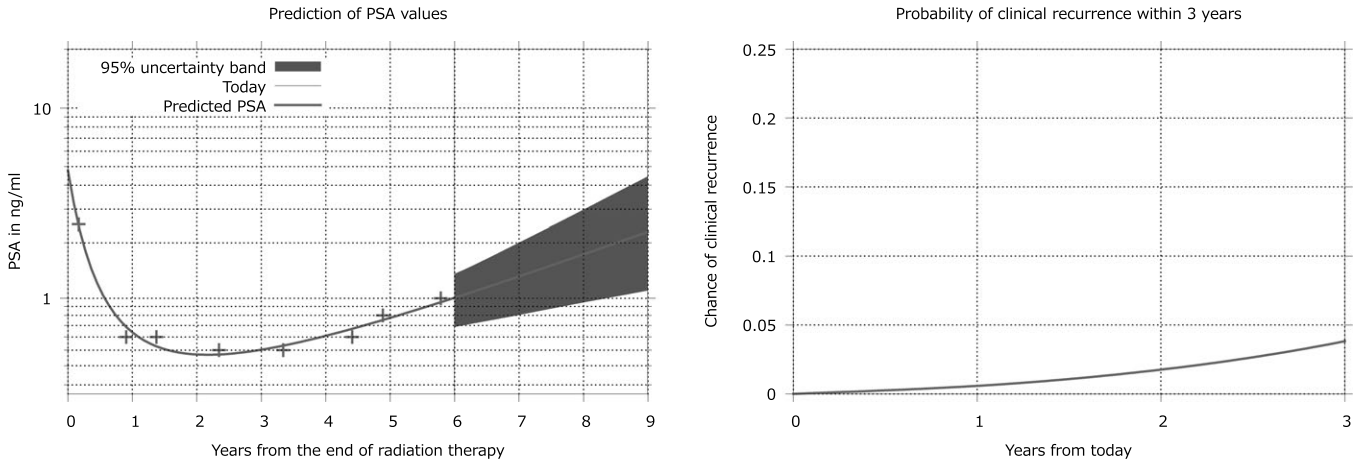
To illustrate the prediction method in Figures 1 and 2, we show the observed and predicted PSA values and the estimated risk of recurrence for two new patients, who have been followed for 13.5 and 6 years, respectively, after radiation therapy without clinical recurrence or hormone therapy. As expected the patient with the rising pattern of PSA has a greater risk of recurrence.

5. Warning Messages on the Internet Calculator

We built into the program on the website a number of checks to warn the user if we did not think the model was providing reliable estimates for the data they entered. The approach used was to calculate eight different statistics $(S_1, \ldots, S_8)$. These statistics, which are derived from the data and the results of the quick MCMC, capture different aspects of the data the user had entered. Each statistic is compared with either the 95th or 99th percentile of these statistics calculated from the training data samples. If the magnitude of the deviation is greater than either the 95th or the 99th percentile the program will either give a warning message or not give an estimate of the risk of recurrence.

Three statistics (S_1, S_2, S_3) reflect the goodness-of-fit of the longitudinal model. One statistic was an overall goodness-of-fit measure given by $S_1 = \frac{1}{m}\sum_{j=1}^{m}(Y_j - \hat{Y}_j)^2$, where Y_j is the PSA data input by the user, $\hat{Y}_j$ is the predicted value of PSA from the model, and m is the number of PSA values. Because of larger variability for small m compared to large m, separate percentiles were used for m less than or equal to 4 and m larger than 4. Another summary statistic considered the deviation from the predicted value of the most recent PSA, given by $S_2 = (Y_m - \hat{Y}_m)^2$. This criterion was used because it was felt to be very important that the model fits adequately at the most recent time. Another statistic considered the largest deviation of a PSA value from its prediction, given by $S_3 = \max_j (Y_j - \hat{Y}_j)^2$. This was included to detect single gross outliers, possibly due to a data entry error by the user.

Four statistics $(S_4, \ldots, S_7)$ relate to the estimated random effects. We considered the deviation of each of the three random effects separately from the population mean estimates given by $S_4 = \hat{\beta}_{N0} - \boldsymbol{x}_{N0}^T\hat{\boldsymbol{\alpha}}, S_5 = \hat{\beta}_{N1} - \boldsymbol{x}_{N1}^T\hat{\boldsymbol{\alpha}}$, and



(a) Prediction of PSA values

(b) Probability of clincial recurrence

Figure 2. (a) Observed and predicted PSA values (on log scale) for years from the end of radiation therapy and (b) probability of clinical recurrence for years starting at the end of PSA follow-up.

$S_6 = \hat{\beta}_{N2} - \boldsymbol{x}_{N2}^T \hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}}, \hat{\beta}_{N0}, \hat{\beta}_{N1}$, and $\hat{\beta}_{N2}$ are the mean of the posterior distributions. We also assessed whether the three random effects as a set were extreme, using the statistic $S_7 = (\hat{\boldsymbol{\beta}}_N - \boldsymbol{x}_N^T \hat{\boldsymbol{\alpha}})^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\beta}}_N - \boldsymbol{x}_N^T \hat{\boldsymbol{\alpha}})$.

We also considered whether the uncertainty in the predicted PSA values 3 years after the current time was too large. The statistic used was S_8, the variance of the predicted PSA 3 years in the future.

The 95th and 99th percentiles or the range of S_1 to S_8 is given in Table 2 in the Supplementary material. The user is warned if any of S_1 to S_8 are outside the 95th percentile range, and the estimated risk of recurrence is not provided if any of S_1, S_3, or S_6 are outside the 99th percentile range. Since the predictions are strongly impacted by the most recent PSA and the calculator is really designed to be used with up-to-date PSA values, we also warn the user if their latest PSA value is more than 3 years old.

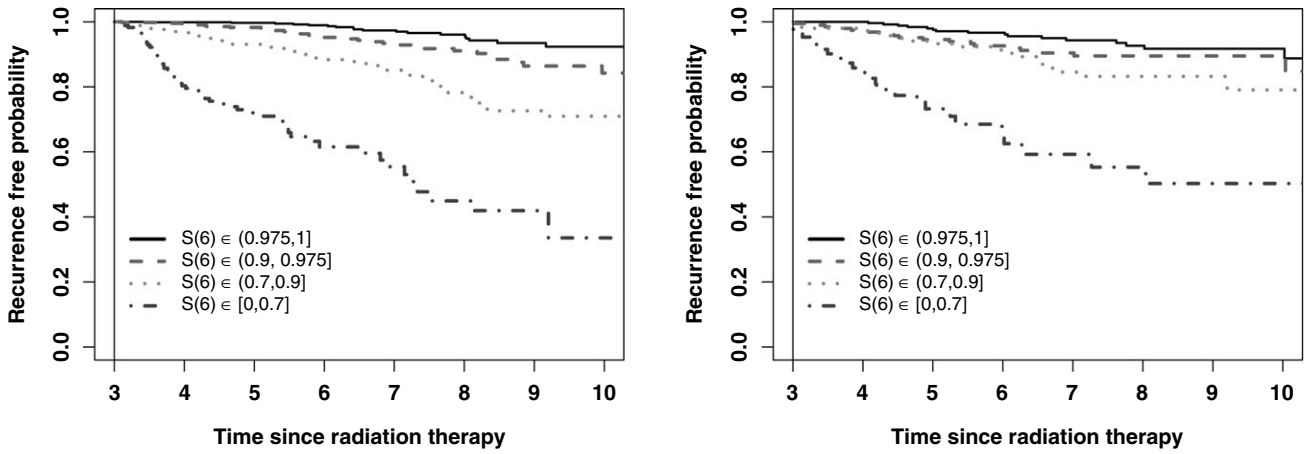
6. Validation of the Model

If a model such as this is to be used it needs to be externally validated. It was for this reason that a dataset was held back to be used for validation. There is a large literature on validating models with many different approaches described in Taylor, Ankerst, and Andridge (2008) and Steyerberg et al. (2010), including receiver operating characteristic curves, Brier scores and other metrics for calibration, discrimination and decision analysis. Most of these measures are for binary predictions. Some have been developed for survival analysis data, but we are not aware of any that can handle time-dependent treatments in the validation data. Because of this complicating issue in our data, we opt for a simple graphical way to assess the predictions.

The complicating issue is that the predictions we make are under the assumption that the person will not take hormone therapy; however, in the testing data, some of the patients do take hormone therapy prior to recurrence during their follow up. Because hormone therapy tends to be taken by those at higher risk of recurrence, and it delays recurrence, we would expect to predict more recurrences than are actually seen in the data. We attempt to address this in two ways. In one approach we censor individuals in the testing data at the time of hormone therapy, which results in dependent censoring. In the other approach we call the initiation of hormone therapy an event.

The graphical approach we use is to fix a time c years after baseline in the testing dataset, and consider all people who are at risk at c. For these people, for the purpose of making an individual prediction, we discard all their information after c. Then for each person separately we estimate the risk of recurrence using the calculator for times after c, and repeat this for all people. We now categorize the people into four groups based on the estimated risk at $c + 3$. The four groups are defined by the ranges for the risk of 0–0.025, 0.025–0.1, 0.1–0.3, and 0.3–1.0. The exact choice of the groups is arbitrary but designed to have similar numbers in each group. We then consider what actually happened to the patients in the four groups after c by constructing Kaplan–Meier plots of time to recurrence. The Kaplan–Meier point estimates at c plus 3 years should be in the appropriate range if the model is valid, and if there were no dependent censoring and the correct definition of the event is used.

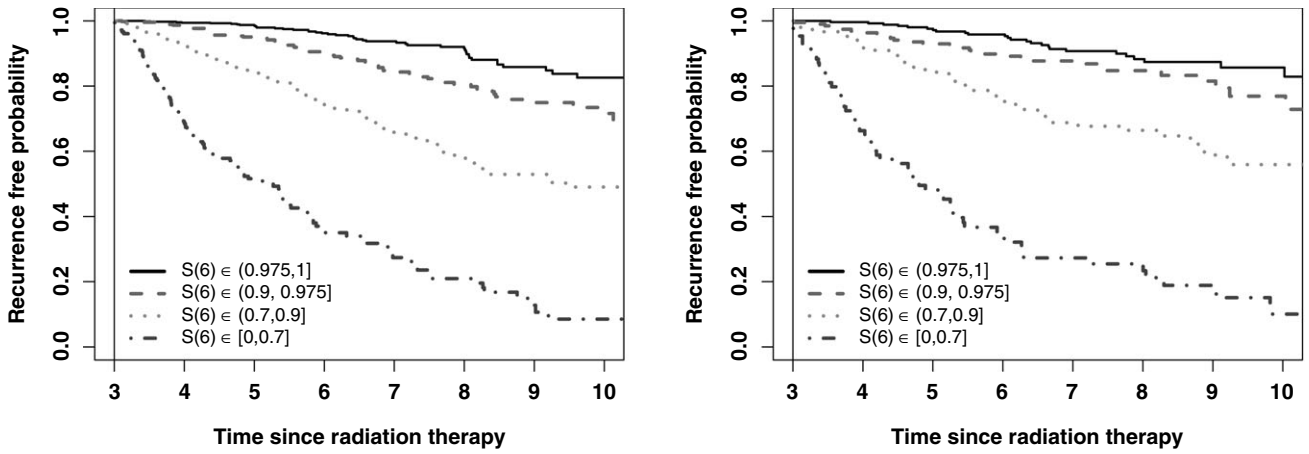
Figure 3(a) shows these Kaplan–Meier plots when the model is reapplied to the individuals in the training dataset and Figure 3(b) shows the results when applied to the individuals in the independent testing dataset, using $c = 3$, when we censor at the time of hormone therapy. The plots are similar and both show that predictions give the correct ordering of the groups. In both plots the curves are slightly higher than would have been expected from the predictions, e.g. for the group with predicted risk at the range of $(0.1, 0.3)$, or equivalently $S(6) \in (0.7, 0.9)$, the Kaplan–Meier estimate is close to 0.9, whereas we might have liked it to be closer to the middle of the 0.7 to 0.9 range. However, this is exactly the direction of the bias we would have expected because of the dependent censoring. The alternative strategy of counting hormone therapy as an event is expected to cause bias in the other



(a) Training Data Set

(b) Testing Data Set

Figure 3. Kaplan–Meier plots of the patients who are at risk 3 years after radiation therapy. Groups are based on a range of S(6), the predicted 3-year recurrence-free probability. Observations are censored at the time of hormone therapy.



(a) Training Data Set

(b) Testing Data Set

Figure 4. Kaplan–Meier plots of the patients who are at risk 3 years after radiation therapy. Groups are based on a range of S(6), the predicted 3-year recurrence-free probability. Hormone therapy is treated as an event.

direction. The Kaplan–Meier plots from doing this are shown in Figure 4. As expected the estimates from the Kaplan–Meier curves are lower than the predictions.

Another consideration is the range of risks in which it is most important for the predictions to be well calibrated. In particular if the 3 year prediction is less than about 2.5% hormone therapy would not be recommended, and thus it would not be important that the predictions be highly accurate in that low range. Similarly if the 3 year risk is larger than about 50%, that person should get hormone therapy, and very accurate predictions of such high risk is not needed. Thus, based on the ranges of the four groups it would be important that the middle two lines in each of Figures 3 and 4 are accurate. In Figure 3 these lines are within or near the high end of the valid range at 3 years, and in Figure 4 they are within or near the low end of the valid range at 3 years.

While different ad hoc strategies or more technical approaches to validation could be developed, the graphical approaches suggested that the model was adequate based on the fact that the estimates from the training and testing datasets were similar, that we had bracketed the predictions at 3 years by treating the hormone therapy in different ways, and that we were close to the valid range within the range of risk that matters.

7. Discussion

There are numerous calculators available on the Internet for prediction of prostate cancer recurrence, and they calculate the risk of various types of recurrence, for differing groups of patients, using different sets of input variables. To the best of our knowledge, these calculators use only baseline variables. The calculator in this article is unique in that it additionally

incorporates a longitudinal series of PSA values, the strongest predictor of recurrence. Thus it can be used to give real time predictions and a patient can update his predictions as new PSA values are obtained.

Radiation oncologists are very much aware that patients with higher values of Gleason grade, T-stage, and pretreatment PSA are more likely to recur and also that patients with high and fast rising values of PSA after treatment are at higher risk of recurrence. Various methods of summarizing the pattern of PSA have been developed, including such things as PSA doubling time and definitions of biochemical recurrence based on the pattern of PSA changes (Roach et al., 2006). Although somewhat ad hoc, these are very useful aids in medical decision making. In a broad sense the model-based predictions in this article are quite similar in that they take the available data for the person and summarize them. The advantage of the model-based approach is that it summarizes the available data in a principled way, and it gives a summary in terms of an interpretable number, the probability of clinical recurrence. In contrast, biochemical recurrence is a binary variable with varying definitions and PSA doubling time is a continuous variable, neither are explicitly calibrated to risk of clinical recurrence. The baseline variables also provide some information about the future risk of recurrence; these are incorporated into the model-based predictions, but neither the doubling time nor the biochemical recurrence definitions use the baseline information.

The data that were used to develop the calculator is quite old, and it may be important to update the model using newer data to account for changes in clinical practice. However, it will always be necessary to include old data so that there is long follow-up in the training dataset.

There are a number of ways in which we can extend this calculator. We have described in this article the predicted risk of recurrence if the person does not start hormone therapy, we could also give the predicted risk of recurrence if the person were to start hormone therapy today. A common treatment for more advanced stage of disease patients is to give a short course of hormone therapy at the same time that radiation therapy is given. We could also provide the risk of recurrence for such patients, using the joint modeling approach. For these patients the pattern of PSA is quite different from those who do not receive hormone therapy initially, and much more heterogeneous, thus a different longitudinal model would be needed.

For the goal of predicting the risk of recurrence without hormone therapy, the presence of hormone therapy in the training datasets is a nuisance. Our method for handling this is to simply include it as a time-dependent covariate in the hazard model. If the goal is to estimate the effect of hormone therapy, experts in causal inference would prefer other strategies than the regression adjustment in equation (3). However, we demonstrate in Kennedy et al. (2010) that the modeling approach in equation (3) is effective and justified. A more important issue is whether the effect of hormone therapy differs from one person to the next and may depend on factors such as the patient's age, T-stage, or current PSA value. Equation (3) could be easily modified to accommodate such effects by including interaction terms with $I(t \geq H_i)$. This would be very useful for calculation of the predicted risk of recurrence if the person were to start hormone therapy today.

The model we presented is very parametric in its structure. While there is a lot of prior analysis of other datasets by the authors and others which would support the general form for the models, it is certainly possible that they could be extended and modified in a number of different ways. For example one might consider splines and correlated errors in the longitudinal model, interactions in the longitudinal or survival model, other ways of linking the longitudinal and survival models, different knot locations for the baseline hazard, smooth baseline hazards, time-dependent parameters in the Cox model, to name a few possibilities. Recent work (Proust-Lima et al., 2012) has also suggested that a latent class structure for the joint models can be very useful for predictions.

There is a large literature on methods of validation of prediction models. While these methods have been extended to handle censoring, an open question is how to extend them to handle time-dependent covariates and dependent censoring.

An interesting, but nonstatistical question, is how to communicate risk estimates from a web calculator. While the graphs such as those in Figures 1(b) and 2(b) might be very familiar and understandable to statisticians, they are probably less well understood by clinicians, other health care providers, and prostate cancer patients, the target users. We are currently in the process of pilot testing a number of different graphical displays of the numerical results from this calculator to assess which displays are better understood.

8. Supplementary Materials

Web Appendices and Tables referenced in Sections 3, 4, and 5 are available with this article at the Biometrics website on Wiley Online Library.

Acknowledgements

This work was partially funded by US National Institutes of Health grants CA 129102, CA 110518, CA054174, and CA 69568, and the French National Institute of Cancer grant 2010-059.

References

Ankerst, D. P. and Finkelstein, D. M. (2006). Clinical monitoring based on joint models for longitudinal biomarkers and event times. *Handbook of Statistics in Clinical Oncology*, 2nd ed., *JC Crowley and DP Ankerst* (eds), 397–418. Boca Raton, CRC Taylor.

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

Kennedy, E. H., Taylor, J. M. G., Schaubel, D. E., and Williams, S. (2010). The effect of salvage therapy on survival in a longitudinal study with treatment by indication. *Statistics in Medicine* **29,** 2569–2580.

Lin, H., Turnbull, B. W., McCulloch, C. E., and Slate, E. H. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data. *Journal of the American Statistical Association* **97,** 53–65.

Pickles, T., Kim-Sing, C., Morris, W., Tyldesley, S., and Paltiel, C. (2003). Evaluation of the Houston biochemical relapse definition

in men treated with prolonged neoadjuvant and adjuvant androgen ablation and assessment of follow-up lead-time bias. *International Journal of Radiation Oncology, Biology, Physics* **57,** 11–18.

Proust-Lima, C., Sène, M., Taylor, J. M. G., and Jacqmin-Gadda, H. (2012). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research* to appear.

Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach. *Biostatistics* **10,** 535–549.

Proust-Lima, C., Taylor, J. M. G., Williams, S. G., Ankerst, D. P., Liu, N., Kestin, L. L., Bae, K., and Sandler, H. M. (2008). Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *International Journal of Radiation Oncology, Biology, Physics* **72,** 782–791.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67,** 819–829.

Roach, 3rd, M., Hanks, G., Thames, Jr., H., Schellhammer, P., Shipley, W. U., Sokol, G. H., and Sandler, H. (2006). Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: Recommendations of the RTOG-ASTRO Phoenix Consensus Conference. *International Journal of Radiation Oncology, Biology, Physics* **65,** 965–974.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **21,** 128–138.

Taylor, J. M. G., Ankerst, D. P., and Andridge, R. R. (2008). Validation of biomarker-based risk prediction models. *Clinical Cancer Research* **14,** 5977–5983.

Taylor, J. M. G., Yu, M., and Sandler, H. M. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of Clinical Oncology* **23,** 816–825.

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.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **50,** 375–387.

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,** 835–862.

Yu, M., Taylor, J. M. G., and Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association* **103,** 178–187.

Received December 2011. *Revised May* 2012.
Accepted July 2012.