University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working
Paper Series

Year 2004

Paper 25

Individualized predictions of disease progression following radiation therapy for prostate cancer.

Jeremy Taylor*

Menggang Yu[†]

Howard M. Sandler[‡]

*University of Michigan, jmgmt@umich.edu

[†]University of Michigan, menggang@umich.edu

[‡]University of Michigan Radiation Oncology, hsandler@umich.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/umichbiostat/paper25>

Copyright ©2004 by the authors.

Conclusions: This statistical model presented accurately predicts future PSA values and risk of clinical relapse. This predictive information for each individual patient, which can be updated with each additional PSA value, may prove useful to patients and physicians in determining what post-treatment salvage should be employed.

2 Material and Methods.

2.1 Patients.

2.1.1 Analysis dataset.

The data consist of 934 patients with localized prostate cancer, who were treated with external beam radiation therapy at the University of Michigan and affiliated institutions between July 1987 and February 2000. This patient series has been previously described (Sandler et al 2000, Symon et al 2003). The collection of these data were approved by the University of Michigan Institutional Review Board. Patients were excluded from this analysis if they received planned hormonal therapy before the end of the radiation therapy regimen, or if they had missing information on T-stage, pretreatment PSA or Gleason grade. No patient had a prostatectomy. The percentage of tumor stages 1, 2 and 3-4 were 29%, 62% and 9% respectively. The percentage in Gleason score 5 or less, 6, 7 and 8 or more were 27%, 32%, 32% and 9% respectively. Pretreatment PSA ranged from 0.3 to 229 with a median of 8.2. The patients ages at time of radiation ranged from 41 to 90 with a median of 71, the total dose ranged from 61Gy to 80Gy with a median of 69Gy and the overall treatment time ranged from 43 days to 94 days with a median of 54 days. The radiation was delivered using 3D conformal techniques as described elsewhere (Sandler et al 2000). Post treatment 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 measures is 6150. The maximum time between treatment and a PSA measurement is 145 months. During the follow-up period, up to February 2001, 140 patients experienced clinical recurrences, 65 had local recurrence, 70 had distant metastases and 5 had regional failure as their first clinical recurrence. Of the 794 censored patients, 144 died of other causes before any clinical recurrence and 650 were censored at February 2001 or were lost to follow-up prior to that date. The median follow-up time was 44 months. Fifty nine patients received hormonal therapy prior to any clinical recurrence, of these 15 had a later clinical recurrence.

2.1.2 Validation dataset.

The validation dataset consists of all data collected on these 934 patients after February 2001 and available in September 2003. We focus on the 612 patients who were alive at the last contact time in the analysis dataset, and were not known to have experienced a clinical recurrence or received hormonal therapy prior to February 2001. There were 541 patients where new follow-up information was available. Amongst these 541 patients, 472 were alive at the end of the this new follow-up, 63 died from other causes not related to prostate cancer and 6 died from prostate cancer. The median additional follow-up time is 30 months. There were 329 patients with additional PSA values, these patients provided 999 PSA measurements within three 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 had HT before the recurrence, an additional 14 have received HT without any clinical recurrence.

2.2 Statistical Methods.

We use a comprehensive statistical model to fit the clinical recurrence events and pattern of PSA data. The model used in this paper is an adaptation and enhancement of the model previously described in Law et al (2001), and Yu et al (2003). In this paper the model is used to fit the data in the analysis dataset and to make predictions of future clinical events and future PSA values for individual patients. These predictions are then compared with the real data from the validation dataset.

The statistical model requires three components, one to model the possibility of cure, one to model the serial PSA measurements and one to model the hazard of a clinical event. A brief description is given below, with a fuller description in the Appendix. A logistic regression model was used for the probability of cure. We assume that the log-odds of the probability of cure is a linear combination of the baseline covariates, T-stage, Gleason, PSA, total dose, treatment duration and Age. Non-linear hierarchical mixed models were used for serial PSA measurements. The longitudinal model for PSA for person i at time t is based on $PSA_i(t) =$

three groups as shown in the table. Within each of these groups the observed number of recurrences and salvage hormonal therapies in the validation dataset is shown, as well as the calculated Kaplan-Meier estimate of the three year recurrence or hormonal therapy probability. The results show that a larger proportion of recurrences in the groups with the higher predicted probability, this provides support for the validity of the model. It is also clear that the model is predicting a slightly greater proportion of recurrences than were actually observed, however if one includes hormonal therapy, then the prediction are close to the observed data.

4 Discussion

In this paper we present the results of fitting a complex statistical model to PSA and clinical recurrence data following radiation therapy, and then show how this model can be used for individual monitoring of a patient's disease progression. The model predicts, based upon baseline clinical features, as well as all post-treatment PSA values, the probability of future clinical recurrences with and without the use of hormonal therapy, and predicts future PSA values. After potential further enhancements and validation we hope to be able to make the program available for others to use through a web-based interface, for practical post treatment clinical predictive use.

As with any model it contains a number of assumptions. We cannot hope to incorporate in the model all aspects of the known biology of prostate cancer progression, and in this sense it is too simple. However, compared to most statistical models this one would be regarded as complex, but this level of complexity was considered necessary to capture the variability in disease progression. An assessment of whether the assumptions are justified can be made empirically, by considering whether the model fits the current data, whether it can be validated using new data and whether the parameter estimates are plausible. On all three counts the model appears to be adequate.

One of the clearest findings from the parameter estimates is the strength of the association of the slope of PSA with the hazard of clinical recurrence. This is consistent with the well known observation that PSA doubling time is an important determinant of disease progression (Hanks

17. Sartor, C.I., Strawderman M.H., Lin X.H., Kish K.E., McLaughlin P.W., Sandler H.M.: Rate of PSA rise predicts metastatic versus local recurrence after definitive radiation therapy. *International Journal of Radiation Oncology, Biology and Physics*, 1997 **38**:941-947.
18. Skates SJ, Pauler DK and Jacobs IJ. Screening Based on the Risk of Cancer Calculation from Bayesian Hierarchical Change-Point Models of Longitudinal Markers. *Journal of the American Statistical Association*, 2001 **96**:429-39.
19. Slate, EH, Cronin, KA: Changepoint modeling of longitudinal PSA as a biomarker for prostate cancer. *In Case Studies in Bayesian Statistics III*, Springer-Verlag, 1997 435-456.
20. Symon, Z, Griffith KA, McLaughlin, PW, Sullivan, MA, Sandler, HM. Dose escalation for localized prostate cancer: Substantial benefit observed with 3D conformal therapy. *International Journal of Radiation Oncology, Biology and Physics*, 2003 **57**:384-390.
21. Taylor, J.M.G., Griffith, K., Sandler, H.: Definitions of Biochemical Failure in Prostate Cancer Following Radiation Therapy. *Int. J. Radiation Oncology Biology Physics*, 2001 **50**: 1219-9.
22. Thames H.D., Kuban, DA., et al. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995. *Int J Radiat Oncol Biol Phys*, 2003 **57**:929-943.
23. Vicini F.A., Kestin L.L., Martinez A.A. The importance of adequate follow-up in defining treatment success after external beam irradiation for prostate cancer. *Int. J. Radiation Oncology Biology Physics*, 1999 **45**: 553-561.
24. Yu M., Law N.J., Taylor J.M.G. and Sandler H.M. Joint Longitudinal-Survival-Cure Models and Their Application to Prostate Cancer. *Statistica Sinica*, 2003 In Press.
25. Zagars, G. K. and Pollack, A. The fall and rise of prostate-specific antigen. *Cancer*, 1993 **72**, 832-842.

Table 1: Odds-ratio estimates from cure model.

Parameter	odds-ratio	95% interval
$I(T_{stage} = 1)$	0.19	(0.08,0.46)*
$I(T_{stage} = 2)$	0.22	(0.09,0.52)*
$I(T_{stage} = 3 - 4)$	1.00	
$\ln(1+\text{basePSA})$	2.66	(1.83,3.86)*
Gleason	1.18	(0.98,1.42)
Age at RT (yrs)	0.99	(0.96,1.02)
Total Dose (Gy)	0.89	(0.83,0.96)*
Duration (days)	0.97	(0.93,1.02)

* indicates 95% interval excludes 1.0

Table 2: Relative hazard estimates from clinical recurrence model.

Parameter	Relative hazard	95% interval
$I(Tstage = 1)$	0.46	(0.23,0.90)*
$I(Tstage = 2)$	0.83	(0.52, 1.34)
$I(Tstage = 3 - 4)$	1.00	
$\ln(1+basePSA)$	1.32	(1.04,1.66)*
Gleason	1.21	(1.04,1.41)*
Total Dose (Gy)	1.05	(0.99,1.12)
Current PSA ¹	1.11	(0.98,1.25)
Current PSA Slope ²	691	(156,3047)*
Hormonal Therapy	0.056	(0.02,0.20)*

* indicates 95% interval excludes 1.0.

¹ $\ln(1 + PSA)$.

² square root(slope of $\ln(1 + PSA)$).

Table 3: Validation of the longitudinal model for predicting future PSA values. Comparison of model predicted PSA ranges and observed PSA data from the validation dataset.

	0-1 Year	1-2 Years	2-3 Years
Total #PSA measurements	287	406	306
Percentage > upper 97.5%	1%	3.6%	3.9%
Percentage between 2.5% and 97.5%	95.5%	92.8%	91.2%
Percentage < lower 2.5%	3.5%	3.6%	4.9%

Table 4: Validation of the survival model for predicting clinical recurrences. Comparison of model predicted and observed recurrences within 3 years.

Prob. of recur in 0-3 yrs	No. patients	No. recur	No. recur or HT	3yr KM estimate of recur or HT
0.00-0.02	191	0 (0.0%)	4 (2.1%)	0.02
0.02-0.10	257	5 (2.0%)	8 (3.1%)	0.04
0.10-1.00	93	6 (6.5%)	12 (12.9%)	0.18

Figure 1: Individual prediction of distribution of time to clinical recurrence (up to 4 years) for 3 selected censored patients with long follow-up. The left-side vertical axis is shown on a $\log(\text{PSA}+1)$ 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 right hand graph shows uncertainty estimates of the predictions.

— Without HT given during the 4 years after the last contact time
 - - - With HT given at the last contact time

Figure 2: Individual prediction of distribution of time to clinical recurrence (up to 4 years) for 3 selected censored patients with short follow-up. The left-side vertical axis is shown on a $\log(\text{PSA}+1)$ 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 right hand graph shows uncertainty estimates of the predictions.

(d) estimated probability of recurrence = 0.37

(e) estimated probability of recurrence = 1

(f) estimated probability of recurrence = 0.24

— Without HT given during the 4 years after the last contact time
 - - - With HT given at the last contact time

Figure 3: Validation of the model. Comparison of future observed PSA values with predicted PSA 95% ranges (shaded region) derived from statistical model for 6 selected patients.

● : PSA measurements from the analysis dataset
 + : PSA measurements from the validation dataset

5 Appendix

The statistical model has three components, one to model the possibility of cure, one to model the serial PSA measurements and one to model the hazard of a clinical event. The data available on patient i is $(PSA_i(U_1), PSA_i(U_2), \dots, PSA_i(U_m), X_i, t_i, \delta_i)$ where $PSA_i(U_1)$ is the value of PSA at time U_1 , X_i are baseline covariates, t_i is the time of clinical recurrence or last follow-up and δ_i is the indicator for clinical recurrence. A logistic regression model was used for the probability of cure. Let D_i be the latent cure indicator, with $D_i = 1$ denoting cure and $D_i = 2$ denoting not cure. Thus each patient is in one of two categories, either they are susceptible to recurrence or non-susceptible to recurrence. The value of D_i is unknown except for patients who have had clinical recurrence. It is reasonable to assume that D_i only depends on baseline and treatment variables, thus we assume $\log(P(D_i = 1)/(1 - P(D_i = 1))) = \beta_0 + \sum X_{ij}\beta_j$. Non-linear hierarchical mixed models were used for serial PSA measurements. Let $\log(1 + PSA_i(t_{ij})) = R_i(t_{ij}) + e_{ij} = \log(1 + Z_i(t_{ij})) + e_{ij}$ where e_{ij} is measurement error, which we assume has a student's T distribution with 5 degrees of freedom. The longitudinal model for PSA is $Z_i(t) = a_i \exp(-b_i t) + c_i \exp(d_i t)$. For cured patients we assume d_i is close to zero. The set (a_i, b_i, c_i, d_i) are latent variable random effects, which have their own joint multivariate normal distribution for each value of D_i , and are restricted to be positive, they are also allowed to depend on baseline covariates, through a regression model. The proportional hazards model for a clinical event for susceptible patients ($D_i = 2$) is given by $\lambda_0(t) \exp(\alpha_1 R_i(t) + \alpha_2 sl_i(t) + \kappa HT(t - \tau_i) + \omega X_i)$. In this model $sl_i(t)$ is the square root of the slope of $R_i(t)$, and $HT(t)$ is a function which is non-zero on 0 to 60 months, it linearly decreases from one at $t=0$ to zero at $t=60$, and τ_i is the initiation time of hormonal therapy for those who received it.

PSA data after hormonal therapy is not used in fitting the model, but clinical recurrences after hormonal therapy are used. The model is fit using Markov Chain Monte Carlo (MCMC) methods, as described in Yu et al (2003). The MCMC gives the posterior distribution of all the parameters and the latent variables.

The prediction of a future PSA value for patient i at time t is given by the formula $PSA_i(t) = \exp(\log(1 + Z_i(t)) + e) - 1$, where $Z_i(t) = a_i \exp(-b_i t) + c_i \exp(d_i t)$, with (a_i, b_i, c_i, d_i) produced by the MCMC method and e is a random draw from a T_5 distribution. Consecutive iterates of the MCMC give different values of the parameters and hence of $PSA_i(t)$, thus giving a range of possible future PSA values. We use 2000 draws from the MCMC and present the 2.5th to 97.5th percentile range.

For the prediction of future clinical events the probability that patient i is event free T months into the future can be calculated from the equations describing the model. Different iterates from the MCMC will give different probability, both the average probability and the range of probabilities (from 40 draws) are presented graphically.

6 Acknowledgement.

This research is supported by the National Institutes of Health grants P30 CA46592 and P50 CA069568.