TO : John Doe, M.D.

FROM : Scott Emerson, M.D., Ph.D.

DATE : March 29, 1995

SUBJECT : PSA nadir vs time in remission

The following are the results of my analysis of your data on PSA's in treated prostatic carcinoma patients. Some of the sections that I have written below cover material which you know better than I do (e.g., Background, Description of the Data). However, I would ask that you carefully read those sections to make sure that I understand the relevance of the data and the clinical importance of the questions. Some of the statistical techniques I used may have been directed by my understanding of the problem. Therefore, any errors in my understanding may invalidate some of the analyses. Please bring any such errors to my attention, and we can then explore the impact they may have had on the analysis results.

After you have read this memo, please give me a call so we can discuss any parts which remain unclear.

Summary

Fifty patients with hormonally treated carcinoma of the prostate were followed prospectively to determine whether post-treatment serum prostate specific antigen (PSA) levels were predictive of length of remission. Each patient had PSA levels measured on a regular basis following treatment, and the lowest value observed was recorded as the PSA nadir. Patients with lower PSA nadirs were found to have significantly longer times in remission ($P < .0001$). A patient having double the nadir PSA as another is predicted to have a 27% shorter median time to progression (95% confidence interval: 18% to 35% decrease). This association was found to persist (adjusted $P < .0001$) even when comparisons are made between patients having identical performance status and tumor mass as measured by extent of disease on bone scan (predicted decrease in median survival is 23%, 95% confidence interval: 18% to 35% decrease). No statistically significant effects of age or tumor grade on length of remission were noted. The interpretation of these results is complicated somewhat by the fact that our ability to measure a PSA nadir may depend upon the patient continuing in remission. Exploratory analyses directed toward assessing this possibility suggest that the association between lower PSA nadirs and longer remission times persists even when attention is restricted to only those patients having remission times 18 months or longer. These results then suggest that serial determination of serum PSA levels may prove useful in monitoring the clinical course of prostatic carcinoma patients post-treatment.

Background

Prostate cancer is a common form of cancer in males with a striking increase in incidence with advanced age. It is estimated, for instance, that well over 50% of men over age 70 have small carcinomas in their prostate. Many of these cancers pose no serious risk to the patients. However, in some patients (particularly those diagnosed at an early age), prostate cancer proves very aggressive, spreading to the bone marrow, lungs, brain, and other vital organs. Untreated, these aggressive tumors often lead to early death. As many of the prostate cancers are stimulated by male hormones, many therapies for aggressive disease is hormonal in nature (e.g., DES or orchiectomy).

Recently, great advances have been made in the diagnosis of prostate cancers. Serum prostate specific antigen (PSA) levels have been found to be markedly elevated in patients with prostate cancer relative to those without. Less clear is whether PSA levels are predictive of worsening disease either pre- or post-treatment. This study was designed to follow hormonally treated prostatic carcinoma patients to determine whether PSA levels were associated with length of cancer remission.

Questions of Interest

1. Are serum PSA levels predictive of length of time in remission?
2. Are any associations between post-treatment PSA levels and length of remission independent of other prognostic indicators such as performance status, extent of disease, aggressiveness of tumor, and patient age?

Description of the Data

You have data available on 50 patients (identified prospectively) who received some sort of hormonal treatment of their prostatic carcinoma (e.g., orchiectomy, DES). These patients were then followed (for a minimum of 24 months) with serial measurements of their prostate specific antigen (PSA). For each patient, you then recorded the time to progression of disease (in months), and you identified the lowest value of PSA observed post-therapy (the PSA nadir). In addition, information is available on patients' age (in years), pre-treatment PSA level, performance status (measuring the ability of the patient to perform usual tasks: most impaired=0, least impaired=100), tumor grade (measuring the aggressiveness of the tumor as assessed by microscopic examination: least aggressive=1, most aggressive=3), and bone scan score (measuring extent of disease: least extensive bony metastases=1, most extensive bony metastases=3).

We do not have some pertinent data available:

1. Some patients were still in remission at the time of study closure. Thus, for those patients we have what is termed *censored* observations in the statistical literature. That is, we know that the disease has not progressed by the time of last follow up, but we do not know the exact time in remission.
2. I do not have available the time following treatment at which the PSA nadir was observed. Because those nadirs may have been reached at markedly different times, the measurements for different patients may not be directly comparable. For instance, a patient who died within 3 months post-treatment may not have had the opportunity to reach as low a nadir as another patient who reached his PSA nadir, say, 9 months post-treatment.
3. We are missing measurements for some covariates on some patients. For example, we are missing pre-treatment PSA measurements on 7 patients and tumor grade data on 9 patients.

Statistical Methods

Because of the presence of censored observations, statistical methods for the analysis of times to failure were used. Graphical assessment of associations between the predictor variables and time

to progression was made using Kaplan-Meier estimates (Kaplan, E.L. and Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457, 1958.). For the purposes of these graphs, continuous variables were categorized to obtain groups with roughly comparable numbers of patients with progression of disease. Determination of statistically significant associations between the predictor variables and time to progression were done in the context of a log linear regression model for censored observations from a Weibull distribution (Kalbfleisch, J.D. and Prentice, R.L. *The Statistical Analysis of Failure Time Data.* New York: John Wiley and Sons, Inc., 1980.). The validity of the Weibull assumption was assessed using graphical means and hypothesis tests based on more general distributions. Stepwise model building was used to determine the most important variables to be included in a regression model. Variables were included on the basis of the strength of their association with time to progression, as well as the presence of confounding among the predictors. Logarithmic transformations (base 2) of the pre-treatment and nadir PSA levels were performed to minimize the skewness present in these measurements. Cases were excluded from a particular analysis only if they were missing data for a variable included in that analysis.

The Kaplan-Meier curves estimate the percentage of patients continuing in remission at each point in time. These methods were derived to handle the problem of censored observations. For the categorical variables of bone scan score and tumor grade, I can compute curves for each of the categories. For age, PSA levels, and performance status, I found it useful to combine the patients into groups having somewhat similar, though not identical, measurements. These groupings are for descriptive purposes only, and they allow us to observe trends in the probability of remaining in remission. It should be noted that there is some statistical variability inherent in the estimation of these curves, thus they should not be interpreted too concretely.

In the log linear regression models, the logarithm of the time to progression is modeled as a function of the predictor variables. Such methods are easily adapted to handle censored observations. Under these models, we can think of the predictor variables acting to increase or decrease the median time to progression in a multiplicative manner. We most often quantify this effect by estimating the proportionate change in the median time to progression associated with a one unit change in a specific predictor. However, if a base 2 logarithmic transformation of the predictor has been used, we estimate the proportionate change in the median survival time associated with a doubling of the predictor. The logarithmic transformation is appropriate when we would expect that the change in time to progression would depend more on the proportionate change in the predictor than the absolute change. That is, we make logarithmic transformations when we expect to see the same percentage change in median survival time when comparing predictor variables having values of 10 and 100 as when comparing those with values of 100 and 1000. In both of these comparisons, we are considering a 10 fold increase in the measurements.

By including additional predictors in the model, we effectively make comparisons within risk groups sharing common values for the other variables. For instance, in this analysis we are primarily interested in the association between PSA nadir and time to progression. Due to the highly skewed nature of the PSA measurements, we first transform the PSA nadir measurements using a base 2 logarithmic transformation. Using a Weibull regression model we can estimate the proportionate change in median time to progression associated with a doubling of the PSA nadir. If no other variables are included in the model, there exists the possibility that our estimate is confounded by the fact that patients with lower PSA nadirs might also have lower bone scan scores and higher performance status. Thus we might consider a model including bone scan score and performance status. In this way, our estimates of the effect of PSA nadir on time to progression are interpretable as the proportionate change in median time to progression associated with a doubling of the PSA nadir when all other variables in the model (performance status and bone scan score) are unchanged. By doing this adjustment for the potentially confounding performance status and bone scan score, we are effectively making comparisons within groups of patients who are similar with respect to these two variables.

In order to assess whether the observed associations between the predictors and the time to progression are more than would be expected by random chance, we can perform a hypothesis test of the null hypothesis of no change, which would correspond to a proportionate change equal to 1.00. The

results of such a hypothesis test can be expressed as a P value, which, when smaller than, say, .05, is generally taken as evidence of a significant association. In addition, we often estimate a range of reasonable values for the true proportionate change in median time to progression using a 95% confidence interval. It should be noted that whenever the 95% confidence interval includes 1.0, we conclude that we have insufficient evidence to say that there was a significant association between the median time to progression and the specific predictor.

In the results that follow, I present estimates of the effect of each predictor on time to progression both when modeled individually (univariately) and when adjusted for other important predictors. In deciding which variables to adjust for in a multivariate model, significance tests are not always the answer. Instead, you should include those variables which logically seem appropriate. This is a point that is not understood by a lot of statisticians, much less non-statisticians, and thus in publishing your results you might get some flak. In your case, the variables which should be included in the model are the performance status and the bone scan score, since those are the variables which are widely regarded as highly predictive of time in remission. As it turns out, the PSA nadir is also a strong predictor, and thus it should also be included in each model. Thus, in addition to the univariate hazard ratios, I include for each covariate estimates adjusted for these three variables.

An additional comment should be made about my treatment of missing data for the covariates. Any case missing data for a variable needed for a particular analysis was omitted only from that analysis. This approach is tantamount to the assumption that the missing data is "missing at random" conditional on the modeled covariates. There is no way to verify that this assumption is valid. I will note, however, that this mattered the most when modeling tumor grade and pre-treatment PSA. Hence, analyses performed with those variables are necessarily using a subset of the full dataset, and it is possible that the subjects missing data for those variables might also have different associations between nadir PSA and time in remission. Hence, I also did some exploratory analyses using only those subjects who were not missing data for any variables (a "complete case" analysis). I should note, however, that with extremely small numbers of subjects deleted, we do not have very good precision to be able to tell whether the subgroup with missing data behaves substantially differently. This is detailed further in the Discussion.

Results

Descriptive Statistics

Table 1 presents the descriptive statistics for each of the variables used in the model. In this table I have included the PSA measurements untransformed. From the table you can see that the PSA measurements are quite highly skewed as evidenced by the standard deviation being much larger than the mean for this data which can only be positive. This fact alone is usually reason to consider logarithmic transformations, because in such settings it is often the case that constant multiplicative changes in the laboratory measurement, rather than additive differences, are more reproducibly indicative of higher risk. It is also apparent from the data that we have relatively few patients with bone scan score of 1.

I have included in the table some statistics on the observed months in remission. The most important numbers to have included is the count of how many were censored. Since there was censoring, the means and standard deviations are not truly interpretable, and we would never include them in a manuscript. However, I include these numbers just to demonstrate that the observation times for the censored observations tend to be longer than those for the observed failures. This highlights the need for special statistical techniques to handle the censored data: If we merely discarded these incomplete observations, we would be ignoring our healthiest patients.

There tended to be associations in the sample among some of the predictor variables. In particular, there was somewhat of a tendency for patients with higher performance status to have lower PSA nadirs, and patients with higher bone scan scores to have higher PSA nadirs. Table 2 presents the distribution

of PSA nadir measurements for groups defined by performance status and bones scan scores. There did not seem to be much of an association between PSA nadirs and pre-treatment PSA, age, or tumor grade.

Time to Progression

Figure 1 displays Kaplan-Meier curves for groups defined by each of the predictor variables. Readily apparent is the wide separation between the curves for groups defined by PSA nadir, bone scan score, and performance status. The curves for pre-treatment PSA suggest little difference between the two highest groups. There does not appear to be substantial differences among the groups defined by age or tumor grade. These findings are corroborated by the estimates presented in Table 3. In that table we see that significant univariate associations are observed between time to progression and PSA nadir ($P < .0001$), bone scan score ($P = .005$), performance status ($P = .025$), and pre-treatment PSA ($P = .035$). From the estimates presented in that table, we would expect a 27% decrease in median time to progression when comparing two populations representing a doubling of the PSA nadir (e.g., from 1.0 to 2.0 or from 4.0 to 8.0), a 53% decrease in median time to progression when comparing populations which differ by a one unit decrease in bone scan score (from 1 to 2 or from 2 to 3), and a 3% difference for each one unit difference in performance status between two patients (e.g., from 70 to 71 or from 80 to 81). Based on this last estimate, we would expect a patient to have a median time to progression of $1.03^{10} = 1.34$ that of a patient having a ten unit lower performance status (e.g., from 70 versus 60 or 90 versus 80).

Also given in Table 3 are estimated effects on median time to progression for each predictor when adjusted for PSA nadir, bone scan score, and performance status. From this we can see that there is still a statistically significant association between PSA nadir and time to progression even after adjusting for performance status and bone scan score. Such adjustment does weaken the effect slightly as evidenced by the fact that we would estimate only a 23% lower median time to progression in patients having double the PSA nadir but the same performance status and bone scan score as some reference group. It should be noted that the effects due to bone scan score and pre-treatment PSA are no longer statistically significant at the .05 level. We include the bone scan score in our final model, however, to remove the possibility of confounding. I note that there was no evidence against our assumption of a Weibull distribution for the time to progression.

The results presented here suggest that longer remission times are associated with lower PSA nadirs. In order to assess whether this finding is perhaps due to an inability to measure PSA nadir in patients with early progression of disease, I also examined the regression model including only those patients who continued in remission for at least 18 months. In this analysis, the statistical significance of the association between time to progression and PSA nadir remained. Thus we can conclude that the observed association was not due solely to the delay in reaching the nadir for some patients.

Discussion

As in all observational studies, the results of the above analyses need to be interpreted cautiously. We have found a tendency for patients with lower PSA nadirs to experience longer remissions. Furthermore, we found that this effect persisted even after adjusting for the potential confounding by performance status and bone scan score. While we can never be certain that such confounding has been totally removed, our results certainly suggest that PSA nadir is a stronger predictor of remission length than either performance status or bone scan score. This conclusion is supported both by the greater degree of statistical evidence (the lower P value for PSA nadir), as well as the magnitude of the estimate. Though a one-unit difference in bone scan score is associated with a greater proportionate difference in median time in remission than a doubling of the PSA nadir is (32% decrease for bone scan score versus 23% for PSA nadir), our range of bone scan scores is only two units, while we have patients with a 1000 fold (ten doublings) difference in PSA nadirs. Thus the estimated changes in median time to progression vary much more as we consider the range of PSA nadirs than when we consider the range of performance status (50 units) or bone scan score (2 units).

In Table 3, lower numbers of subjects are used in the analyses that include tumor grade or pretreatment PSA, because some subjects were missing data for these variables. As noted in the Statistical Methods, it is possible that the relationship between nadir PSA and time in remission was different for subjects who were missing data. To explore this, I also fit an analysis of nadir PSA adjusted for bone scan score and performance status when only the 36 subjects with observed pretreatment PSA and tumor grade are included. In these analyses, a slightly lower median ratio was observed for the log nadir PSA (0.64, 95% CI 0.53 to 0.77 compared to 0.77, 95% CI 0.69 to 0.80 in Table 3), though the statistical significance of that estimate is largely unchanged from that in the 48 subjects with available data for nadir PSA, bone scan score, and performance status. I think it best to highlight the analysis on the larger sample size, because it will likely tend to be more generalizable across all subjects. I did not, however, further explore, the impact of the missing data on the analysis results.

In the above analysis, we have not directly addressed the question of whether serum PSA levels are predictive of longer time in remission. We have, however, demonstrated that the patients whose PSA levels dropped more post-treatment had longer times in remission. It would be of interest to directly assess whether later (post nadir) changes in PSA levels were predictive of relapse. This could be effected if I had all PSA measurements on all patients. Should such an analysis be of interest to you, please let me know.

Table 1
Descriptive Statistics
$n = 50$ Patients with Hormonally Treated Prostate Cancer

Variable	Number Missing	Frequency (percent)	Mean	Std Dev	Minimum	Maximum
Age	0		67.4	5.8	58	86
Performance Status	2		80.8	11.1	50	100
Grade	9					
grade 1		10 (24%)				
grade 2		15 (37%)				
grade 3		16 (39%)				
Bone Scan Score	2					
bss 1		5 (10%)				
bss 2		13 (27%)				
bss 3		30 (62%)				
Pre-treatment PSA	7		670.8	1287.6	4.8	4797.0
Nadir PSA	0		16.4	39.2	0.1	183.0
Months in Remission	0		28.5	18.4	1	75
relapse		36	21.8	15.5	1	60
censored		14	45.7	13.7	24	75

Table 2
Measurements of PSA Nadir by Performance Status and Bone Scan Score

		PSA Nadir Levels			
Variable	Number	Mean	Std Dev	Minimum	Maximum
Performance Status					
50	2	0.7	0.6	0.2	1.1
60	2	94.1	125.7	5.2	183.0
70	6	29.0	40.3	0.2	104.0
80	21	18.0	41.0	0.2	169.0
90	14	4.8	10.3	0.2	38.0
100	3	0.3	0.2	0.1	0.5
Bone scan score					
1	5	0.2	0.0	0.2	0.2
2	13	4.8	10.3	0.2	38.0
3	30	24.9	48.7	0.1	183.0

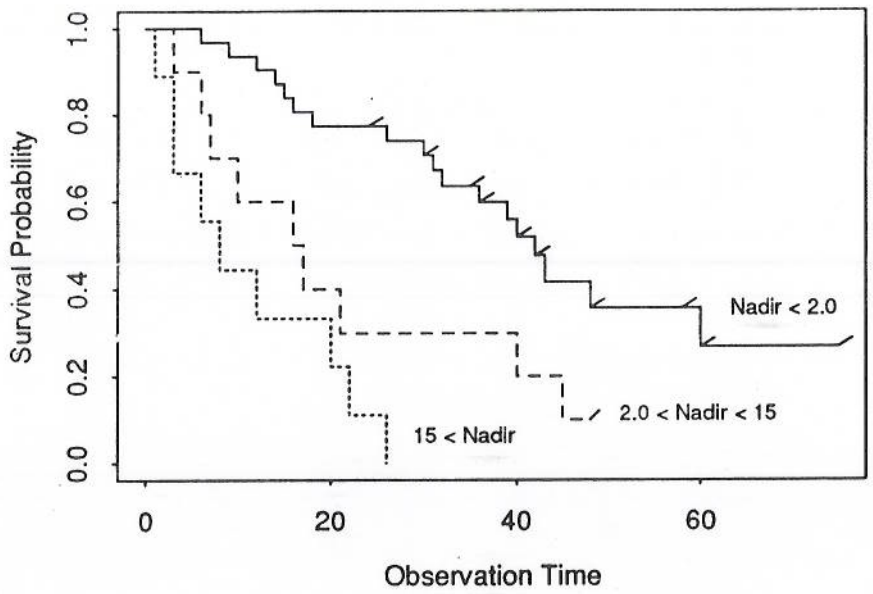
Table 3
Estimated Proportionate Change in Median Time to Progression

	Unadjusted				Adjusted†			
Variable	n	Ratio*	95% CI	P val	n	Ratio*	95% CI	P val
PSA nadir (doubling)	50	0.73	0.65, 0.82	.000	48	0.77	0.69, 0.86	.000
Perf Status	48	1.03	1.00, 1.05	.025	48	1.02	1.00, 1.04	.050
Bone scan score	48	0.47	0.28, 0.80	.005	48	0.68	0.43, 1.08	.104
PSA pretx (doubling)	43	0.86	0.74, 0.99	.035	42	0.98	0.86, 1.11	.775
Tumor grade	41	1.00	0.64, 1.57	.983	40	0.88	0.66, 1.18	.399
Age	50	1.00	0.95, 1.05	.940	48	1.02	0.98, 1.07	.268

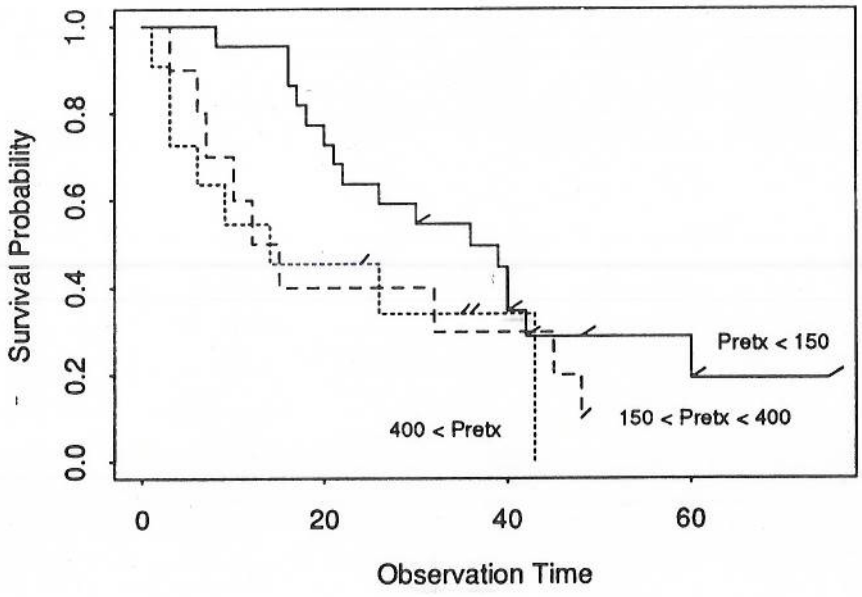
* Estimates are for the ratio of median time to progression comparing two populations the first of which has a variable measurement one unit higher than the second group, except for the PSA measurements when the comparison is being made for two populations the first of which has a nadir or pre-treatment PSA double that of the reference group.

† All estimates adjusted for *nadir*, *ps*, and *bss*3, hence the populations compared in the ratio estimates for pre-treatment PSA, tumor grade and age are similar with respect to those variables. Ratio estimates for nadir PSA are based on populations which are similar with respect to performance status and bone scan score, ratio estimates for performance status are based on populations which are similar with respect to nadir PSA and bone scan score, and ratio estimates for bone scan score are based on populations which are similar with respect to nadir PSA and performance status.

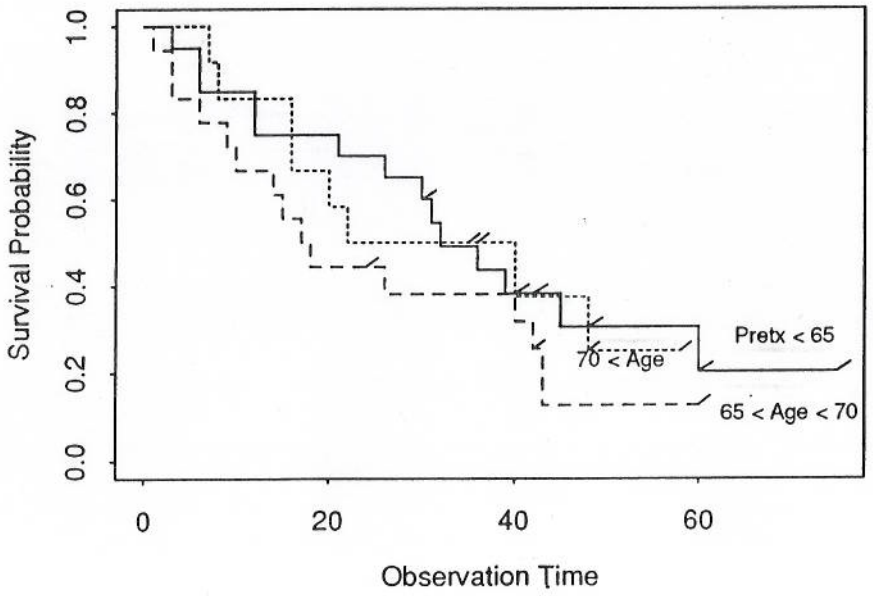
Progression by Nadir PSA



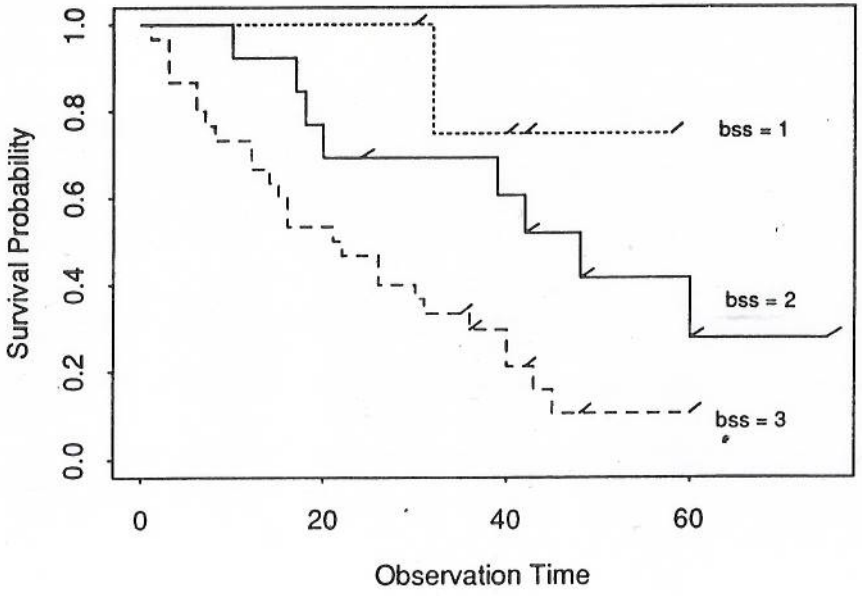
Progression by Pretx PSA



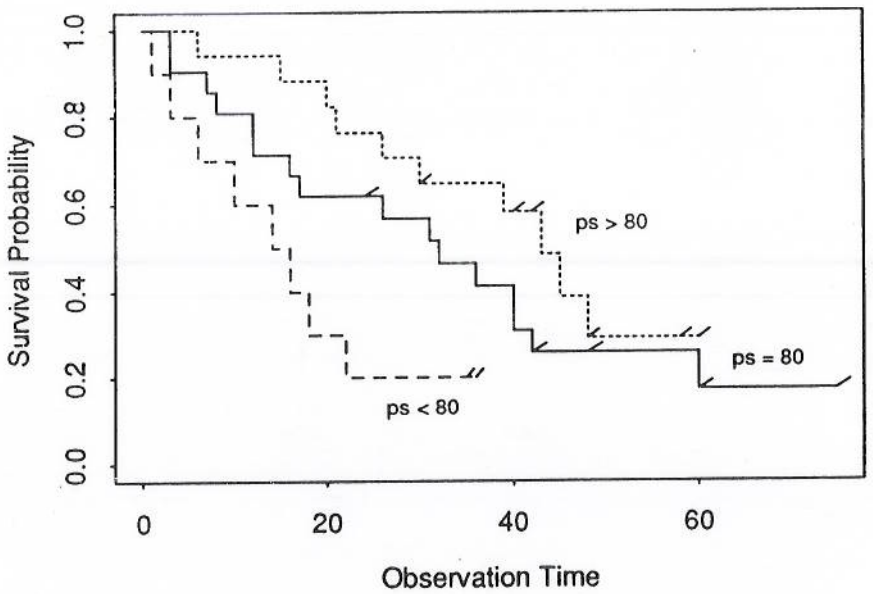
Progression by Age



Progression by BSS



Progression by PS



Progression by Grade