Biological Determinants of Cancer Progression in Men With Prostate Cancer

Thomas A. Stamey, MD
John E. McNeal, MD
Cheryl M. Yemoto
Bronislava M. Sigal, PhD
Iain M. Johnstone, PhD

The dilemma of prostate cancer is that only a small portion of men with untreated carcinoma will die from it, yet the prevalence of this disease is so high that the annual mortality is still second only to lung cancer. Thus, there is a great need to find better ways to distinguish patients who have clinically innocuous cancers from those who have significant disease that can be eradicated by radical prostatectomy or other treatment and those for whom therapy is destined to fail. The dilemma has been sharpened by important advances in diagnostic techniques.

It is known that for most carcinomas, the finding of poor differentiation (high histological grade) offers a reliable and powerful index of aggressive biological behavior, including high risk of metastasis. In practice, histological grading systems are not derived directly from general biological principles of cell differentiation. They are largely empirical and individualized to each cancer by criteria that are tailored to fit observations on factors that are associated with survival.

The challenge of devising an effective grading system for prostatic adenocarcinoma was unique because of its great diversity of histological patterns, with several patterns often appearing together in the same cancer. The special contribution of Donald Gleason, MD, PhD, was to combine or merge all these patterns into a grading scale with only 5 grades, which were distinguished by relatively simple and straightforward criteria. The problem of multiple patterns (grades) in the same cancer was addressed by basing the final prognostic rating on a score, which was the sum of the 2 most prevalent grades in each case. The 2 most common scores in many series are 6 (usually 3 + 3) and 7 (usually 3 + 4). The question arises as to whether traditional use of the Gleason scoring system gives the best estimate of potential therapeutic failure. We have addressed this question for the first time in a quantitative way by using a Cox proportional hazards model in 379 consecutive men treated only by radical prostatectomy.
BIOLOGICAL DETERMINANTS OF PROSTATE CANCER PROGRESSION

prostatectomy. In addition to our study of the Gleason grading system, we have examined 7 other morphologic variables of prostate cancer behavior, as well as serum prostate-specific antigen (PSA) level and prostate size.

METHODS

Six hundred consecutive radical prostatectomies were performed by the full-time faculty at Stanford University Medical Center, Stanford, Calif, between August 1983 and July 1992. Two hundred twenty-one cases were excluded from this analysis—124 received hormones, irradiation, or transurethral resection before surgery, 46 had transition zone cancers (cancers in the anterior part of the prostate, where benign prostatic hyperplasia develops), 33 had incomplete data (serum PSA level or prostate weight), and 18 additional cases could not be classified with regard to failure because of a borderline positive serum PSA finding that was stable and failed to rise with time (median follow-up, 5.73 years). These were the only exclusions, leaving 379 consecutive men with peripheral zone cancers treated only by radical prostatectomy. Data from these 379 men were analyzed using the proportional hazards model, with presence or absence of biochemical failure as a dichotomous response. All patients were followed up at approximately 6-month intervals with serial serum PSA determinations.

Specimens were fixed for 24 hours in 10% formalin, accelerated by internal and external perfusion, serially blocked at 3-mm intervals in planes perpendicular to the rectal surface, and processed routinely into paraffin. Slides cut at 5 µm were examined microscopically; boundaries of prostate capsule, normal landmarks, and all cancer areas on each slide were traced with fine-point marking pens. Outlines were traced sequentially onto a sheet of paper. Study of the paper “map” identified the largest cancer, which was almost always the clinically detected cancer as well as the most poorly differentiated cancer. The largest cancer alone was selected for further study and its volume was determined by computer planimetry, as previously described. Extent of capsule penetration and positive margins were also marked on the slide and quantified in linear centimeters. Percentage of seminal vesicle invasion was estimated from a single coronal section of each side. Lymph nodes from the radical dissection were the only dichotomous variable (+ or −). Percentage of intraductal cancer area within the invasive cancer as well as the number of foci of

<table>
<thead>
<tr>
<th>Table 1. Characteristics and Distribution of 379 Peripheral Zone Cancers*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Gleason grade 4/5, %</td>
</tr>
<tr>
<td>Cancer volume, cm²</td>
</tr>
<tr>
<td>Vascular invasion, No. of foci</td>
</tr>
<tr>
<td>Lymph node involvement, yes/no</td>
</tr>
<tr>
<td>Intraductal cancer, %</td>
</tr>
<tr>
<td>Seminal vesicle invasion, %</td>
</tr>
<tr>
<td>Capsular penetration, cm</td>
</tr>
<tr>
<td>Surgical margin, cm</td>
</tr>
<tr>
<td>Serum prostate-specific antigen level, ng/mL</td>
</tr>
<tr>
<td>Prostate weight, g</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
<tr>
<td>Gleason score</td>
</tr>
<tr>
<td>Follow-up as of January 1997, y</td>
</tr>
</tbody>
</table>

*Ellipses indicate data not applicable.
†Data are percentage of involvement of 379 peripheral zone cancers with 7 morphologic variables.

<table>
<thead>
<tr>
<th>Table 2. Spearman Rank Correlations Among Potential Predictors (N = 379)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Gleason grade 4/5, %</td>
</tr>
<tr>
<td>Cancer volume, cm²</td>
</tr>
<tr>
<td>Vascular invasion, No. of foci</td>
</tr>
<tr>
<td>Lymph node involvement, yes/no</td>
</tr>
<tr>
<td>Intraductal cancer, %</td>
</tr>
<tr>
<td>Seminal vesicle invasion, %</td>
</tr>
<tr>
<td>Capsular penetration, cm</td>
</tr>
<tr>
<td>Surgical margin, cm</td>
</tr>
<tr>
<td>Serum PSA level, ng/mL</td>
</tr>
<tr>
<td>Prostate weight, g</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
</tbody>
</table>

*The cluster of cancer volume with capsular penetration, serum prostate-specific antigen (PSA), and seminal vesicle invasion are the only correlations of more than 0.50 among the 8 morphologic variables; percentage of Gleason grade 4/5 and positive surgical margins had correlations with cancer volume of 0.43 and 0.40. Ellipses indicate data already reported in table.
†Data are statistically significant at P<.001.

©1999 American Medical Association. All rights reserved.
vascular invasion were quantified as described in recent studies.\textsuperscript{4,5}

The Stanford modified Gleason scale was used to estimate the proportion of each cancer that was poorly differentiated in all 379 radical prostatectomies.\textsuperscript{6} Briefly, from the Gleason scale of 5 grades,\textsuperscript{1} grades 4 and 5 (poorly differentiated) constitute a variety of histological patterns having the common feature that they do not form intact gland units mimicking normal architecture. Grades 4 and 5 contrast sharply with Gleason grades 1, 2, and 3, all of which form complete gland units and appear to have similar favorable prognostic significance.\textsuperscript{7} Accordingly, we combined grades 1, 2, and 3 into a single well-differentiated category. The percentage of each cancer occupied by Gleason grades 4 and 5 (\% Gleason grade 4/5) was estimated by a single pathologist (J.E.M.) and performed prospectively before biochemical failure was detected by serum PSA findings. These \% Gleason grade 4/5 data were compared with the Gleason scores, in which the 2 most prevalent grades in each cancer are added together and used as a sum or score. Most commonly, the score reflects 2 adjacent grades (ie, 3 + 4 = 7 or 4 + 3 = 7) or cancers of pure grade (ie, 3 + 3 = 6). For a score of 7, the proportion of Gleason grade 4 cancer may vary between 5\% and 95\% without altering the score (sum). To use the traditional Gleason scoring system, fractional areas of tumor less than 5\% of the total tumor area should be ignored; if the secondary grade is less than 5\%, the primary grade (most common) is simply doubled to obtain the score.

Serum PSA levels were determined by the equimolar automated Tosoh A1A-600 assay (Foster City, Calif), which was run in the ultrasensitive mode, a technique in which a level of 0.07 ng/mL has a 99\% chance of indicating biochemical persistence of cancer cells after radical prostatectomy.\textsuperscript{8,9} Our definition of biochemical failure was a PSA level of at least 0.07 ng/mL and rising in subsequent serum samples (a single serum sample >0.07 ng/mL was sufficient to indicate failure if confirmed by a second serum sample). Fifty-seven of the 235 cured patients among the 379 men did not have a recent serum PSA sample available for the Tosoh assay; in these 57 men, 18 of whom were followed up for more than 5 years, we accepted the available Hybritech value of less than 0.2 ng/mL as evidence of an undetectable PSA level.

The distribution of all variables for the 379 men is shown in Table 1. Spearman rank correlations were calculated for each pair of potential predictors using all 379 cases (Table 2). Univariate Cox regression was used to fit the dependence of failure on all predictors individually (data not shown). The Cox proportional hazards model was used to identify those predictors with the most significant independent influence on prognosis.\textsuperscript{10} Ties at 0 were handled by the Efron approximation.\textsuperscript{11}

Logarithmic transformations of the variables were allowed for more flexible modeling of relationships between predictors and failure. Forward stepwise selection was used on all variables in Table 1 (except follow-up) as the initial model selection approach, supplemented with careful examination of particular models.\textsuperscript{12} We used the statistical software S-PLUS, Version 3.4 (CSIRO, North Ryde, Australia).

Two stopping criteria were used: when no candidate variable for admission had a \( P \) value of less than .05 and when no candidate variable for admission had a \( P \) value of less than .01. The model selected by the second criterion necessarily contains fewer variables than the first. While the \( P < .05 \) criterion is traditional, the relatively large sample size (\( n = 379 \)) argues for a more stringent test of statistical significance (the \( P < .01 \) level). Comparison of the \( P < .05 \) and \( P < .01 \) models is a valuable and cautionary tool in drawing conclusions from regression modeling with correlated predictors. Both models are shown in Table 3.

Variables are listed in Table 3 in the order in which they entered the model in the forward stepwise procedure. Cancer volume entered first because it was the most significant in the univariate Cox analysis. However, the \( P \) values attached to each variable are approximations to the significance of the likelihood ratio test comparing the model shown with a model in which that variable is dropped, and so represent measures of the independent significance of that variable, adjusting for other variables in the model.

In addition, backward stepwise selection was also used. For the \( P < .05 \) model, this gave the same model as in Table 3, while for the \( P < .01 \) criterion, the backward stepwise model had more variables than the forward stepwise model (data not shown).

It must be emphasized that when several predictors present in a model are

\begin{table}[h]
\centering
\begin{tabular}{cccccc}
\hline
Seminal Vesicle Invasion, \% & Capsular Penetration, cm & Surgical Margin, cm & Serum PSA Level, ng/mL & Prostate Weight, g & Age, y \\
\hline
... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... \\
... & ... & ... & ... & ... & ... \\
1.00 & ... & ... & ... & ... & ... \\
0.50† & 1.00 & ... & ... & ... & ... \\
0.31† & 0.42† & 1.00 & ... & ... & ... \\
0.44† & 0.39† & 0.27† & 1.00 & ... & ... \\
0.08 & 0.04 & -0.04 & 0.39† & 1.00 & ... \\
0.13 & 0.19 & 0.08 & 0.17 & 0.24† & 1.00 \\
\hline
\end{tabular}
\end{table}
correlated, considerable caution is needed in interpreting individual regression coefficients and P values, and regression analysis can only uncover associations between predictors and response and suggest hypotheses about causal mechanisms.

RESULTS

Series Characteristics

Ninety-six men had clinical stage T1c cancers (impalpable), 121 had stage T2a (small nodule <1/2 of 1 lobe), 130 had stage T2b (nodule >1/2 of 1 lobe but unilobar), and 32 had stage T2c (bilaterally palpable cancer). Thus, 217 men (57%) had favorable T1c and T2a clinical stages. The median age was 65 years (Table 1). Sixty of 379 men were older than 71 years; 60% had no biochemical evidence of disease recurrence, which was the same as that for the whole group of 379 men (62%).

For the first 8 morphologic predictive variables, there was a wide range of variation (Table 1) among cases, which was somewhat skewed toward smaller, better differentiated cancers, as expected. The percentage of involvement of the 379 peripheral zone cancers with 7 of the 8 morphologic variables is shown in the first column of Table 1. The Spearman rank correlations in Table 2 show that cancer volume, capsular penetration, seminal vesicle invasion, and positive surgical margins are highly interrelated.

Almost all predictors were statistically significant for failure by univariate Cox regression, except prostate weight (data not shown). Kaplan-Meier curves for all 379 men gave 3-, 5-, and 7-year biochemical failure-free probabilities of 66%, 61%, and 59%, respectively. This slowly declining profile indicates that increasing the follow-up period would be unlikely to alter the findings of this study. Both median and mean follow-up time for the 379 men was 5 years.

Table 3 shows the strong independent predictive value of % Gleason grade 4/5 cancer in both models and for cancer volume in the P<.01 model. Note high z scores for percentage of Gleason grade 4/5 cancer in both models and for cancer volume in the P<.01 model.

BIOLOGICAL DETERMINANTS OF PROSTATE CANCER PROGRESSION

Figure. Increase in Failure Rates as a Function of the Percentage of Gleason Grade 4/5 Cancer in 379 Prostatectomy Patients

Numbers above the bars are number of patients. The 19 men on the left side of the 1% to 10% bar had less than 5% Gleason grade 4 cancer. The 71 men on the right had 5% to 10% Gleason grade 4 cancer.

The P<.05 model was arrived at by both forward and backward stepwise procedures; the P<.01 model, by forward stepwise procedure only. Ellipses indicate data not applicable.

Table 3. Cox Proportional Hazards Models for 379 Men With Peripheral Zone Cancers*
shallower slope, ultimately approaching 87%. The 30% of men who had no evidence of disease with at least 41% Gleason grade 4/5 cancer had much smaller cancer volumes than men who experienced failure (P < .001). Table 3 shows that % Gleason grade 4/5 cancer has greater significance for prognosis than any other variable in the Cox proportional hazards model.

Independent Predictive Value of Cancer Volume

Table 3 also shows that prostate cancer volume is a highly significant and independent determinant of biochemical failure by radical prostatectomy in the P < .01 model. Among the 379 cases, the no-evidence-of-disease rate was 86% for men with a cancer volume of 0.5 to 2.0 cm³, 61% with 2.0 to 6.0 cm³, 33% with 6.0 to 12.0 cm³, and only 3% for men with more than 12.0 cm³ of cancer. These steeply declining, volume-based no-evidence-of-disease rates serve to emphasize that volume, like grade, is a powerful predictor of failure.

However, results concerning volume are more subtle because of its significant correlations with serum PSA level, capsular penetration, seminal vesicle invasion, and positive surgical margins, shown in Table 2. While the P < .05 model in Table 3 rejects the latter 2, capsular penetration and PSA level both enter the P < .05 model.

Volume enters the forward stepwise P < .05 model first in Table 3 as the most powerful univariate predictor. Its correlation with PSA level is sufficiently high that the independent effect of PSA level is not strong enough to allow PSA level to enter at the P < .01 level. Prostate-specific antigen level does enter at the P < .05 level, as does capsular penetration, but with 3 correlated variables present in the P < .05 model, it is essentially impossible to interpret their individual coefficients in the P < .05 model.

Further evidence of the importance of volume as a predictor comes from 2 other analyses. First, if PSA level is forced into the P < .01 model, then under forward stepwise selection, volume enters at the P < .01 level and, in the resulting model (data not shown), has a z score of 4.9 compared with 2.1 for PSA level. Thus, at the P < .01 level, volume “shuts out” PSA level, but PSA level does not shut out volume, which might be cautiously interpreted as suggesting that volume has greater significance than PSA level as an independent predictor.

Second, we examined the independent significance of volume vs the cluster of capsular penetration, seminal vesicle invasion, and positive surgical margins. We tested the independent significance, by likelihood ratio test, of volume when added to the P < .05 model, augmented with seminal vesicle invasion and positive surgical margins. The resulting P value was .01, suggesting that volume adds information beyond these 3 correlated predictors.

Finally, all 12 variables in Table 1 have also been examined by univariate and multivariate logistic analysis on 320 of the 379 men who had minimum follow-up of 3 years. The logistic analysis essentially confirmed the results of the Cox proportional hazards model.

COMMENT

Generations of academic debate about the natural history of prostatic adenocarcinoma has led to a tradition of uncertainty among clinicians about how (or whether) to treat any individual patient with this disease. In this study, we have tried to improve the quality of our information and the validity of its interpretation by 4 innovations that have, to our knowledge, never been used together. First, we have analyzed the grading system and discovered a technique for using it that yields greater information. Second, we have subjected each of the histological variables to precise quantitation, rather than the commonly used broad and subjectively determined categories. Third, we have considered all variables together in each patient, using 2 different multivariate analyses to learn which variables are independently responsible for the pathogenesis of disease progression. Last, we have been aggressive in obtaining and testing serum PSA samples in our laboratory, using an ultrasensitive assay, to include as many failures as possible.

We have shown previously that high-grade prostate cancer (% Gleason grade 4/5) and cancer volume are predictive of regional pelvic lymph node metastases. We now find that these same variables are the primary predictors of failure to eradicate prostate cancer, regardless of lymph node metastases. These findings have important research and clinical consequences. For example, research efforts at the genetic and enzymatic molecular levels clearly should be directed at Gleason grade 4 cancer. The clinical consequences of Gleason grade 4/5 cancer are shown in the Figure, in which biochemical failure increases almost linearly with each 10% increase in Gleason grade 4/5 cancer within the prostate. It is important that both the Cox proportional hazards and the multivariate logistic models rejected Gleason sum (score) in the presence of Gleason grade 4/5, clearly indicating that we should move away from the scoring system and simply estimate the % Gleason grade 4/5. This is not difficult because the traditional Gleason score of 7 already requires accurate recognition of grade 4 cancer. Pathologists and clinicians should measure the area of cancer in each core biopsy specimen and simply indicate the percentage of the cancer composed of Gleason grade 4/5 cancer.

How well do biopsies represent the % Gleason grade 4/5 of the cancer within the prostate? We reported in 1995 on the relationship between % Gleason grade 4/5 in 120 men who had 6 systematic biopsies followed by a radical prostatectomy. Although the correlation was not perfect, it was very useful; the R² was 0.63 and most of the nonlinearity was due to cancers in the prostate with less than 20% Gleason grade 4/5. In the current study, most of the biopsies were performed outside Stanford University and only 224 were available for review. Of these, 89 men had sextant biopsies with adequate core tissue (≥ 50 mm of total tissue of a possible 90 mm). Linear regression showed a linear relationship
between Gleason grades in biopsy and radical prostatectomy, with an R² of 0.57. If the biopsy specimen contained at least 20% Gleason grade 4/5 cancer, the prostate contained at least 20% Gleason grade 4/5 in 90% of the men. If the biopsy result showed no Gleason grade 4/5 cancer, the prostate had less than 20% Gleason grade 4/5 cancer 90% of the time. Previous investigators have not shown good correlations between Gleason score in biopsy specimens and radical specimens, but this is to be expected when a single score of 7 can vary in its % Gleason grade 4/5 from 5% to 95%.

Estimating cancer volume is more difficult. The 4 volume ranges of 0.5 to 2.0, 2.0 to 6.0, 6.0 to 12.0 and more than 12.0 cm³ are associated with failure rates of 1.5%, 2.0 to 6.0, 6.0 to 12.0, and more than 12.0, respectively. In these 379 peripheral zone cancers, these ranges are sufficiently broad that it should be possible with further research to estimate them. The area of cancer in each biopsy core specimen, number of positive findings in core specimens of 6 to 10 systematic biopsies of the prostate, serum PSA level (half of which is related to cancer volume), and clinical stage from digital rectal examination are all related to cancer volume and should be part of any future algorithm for estimating cancer volume preoperatively. Although the analysis in this article of morphologic variables that determine failure is based on treatment only by radical prostatectomy, cancer volume and the % Gleason grade 4/5 cancer in peripheral zone cancers are so powerful that it is highly likely the results of any other therapeutic modality, such as radiation or chemotherapy, will be determined by the same morphologic variables described in this research. Katcher et al. using radiation therapy and a proportional hazards model, found that only Gleason grade and preradiation serum PSA level were independently predictive of biochemical failure in men followed up for at least 5 years (cancer volume was not available and clinical stage was rejected by their Cox model, despite the fact that there is volume information in at least T1c to T2c stages).

Once we know the activated genes involved in progression of prostate cancer, molecular markers may offer additional hallmarks of cancer progression and our failure to cure this all-too-common cancer. So far, clinical applicability of markers like p53, E-cadherin, and bcl-2 has not been shown. Our data in the Figure suggest that it will take a powerful molecular marker to replace the simple estimation of the % Gleason grade 4/5 in prostate cancer, in which every 10% increase is accompanied by an increase in failure to eradicate this disease in the 85% of men who have peripheral zone cancers.

Finally, the 72 men (19%) from the 379 who had no Gleason grade 4/5 cancer in their prostate had a biochemical failure rate of only 5.6%. This failure rate is so low that it raises the question of whether they needed the surgery, a question that cannot be answered by randomized trials. On the other hand, we hope that this demonstration of increased failure rates associated with each 10% increase in % Gleason grade 4/5 cancer will contribute to ending the debate about which patients with prostate cancer should be treated.

Funding/Support: This study was supported in part by a grant from the Lucas Foundation, Menlo Park, Calif.

REFERENCES