THE UNIVERSITY OF CALIFORNIA, SAN FRANCISCO CANCER OF THE PROSTATE RISK ASSESSMENT SCORE: A STRAIGHTFORWARD AND RELIABLE PREOPERATIVE PREDICTOR OF DISEASE RECURRENCE AFTER RADICAL PROSTATECTOMY

MATTHEW R. COOPERBERG, DAVID J. PASTA, ERIC P. ELKIN, MARK S. LITWIN, DAVID M. LATINI, JANEEN DU CHANE AND PETER R. CARROLL*

From the University of California, San Francisco, San Francisco, and University of California, Los Angeles, Los Angeles, California (MSL), and TAP Pharmaceutical Products, Inc., Lake Forest, Illinois (JD)

ABSTRACT

Purpose: Multivariate prognostic instruments aim to predict risk of recurrence among patients with localized prostate cancer. We devised a novel risk assessment tool which would be a strong predictor of outcome across various levels of risk, and which could be easily applied and intuitively understood.

Materials and Methods: We studied 1,439 men diagnosed between 1992 and 2001 who had undergone radical prostatectomy and were followed in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) database, a longitudinal, community based disease registry of patients with prostate cancer. Disease recurrence was defined as prostate specific antigen (PSA) 0.2 ng/ml or greater on 2 consecutive occasions following prostatectomy or a second cancer treatment more than 6 months after surgery. The University of California, San Francisco-Cancer of the Prostate Risk Assessment (UCSF-CAPRA) score was developed using preoperative PSA, Gleason score, clinical T stage, biopsy results and age. The index was developed and validated using Cox proportional hazards and life table analyses.

Results: A total of 210 patients (15%) had recurrence, 145 by PSA criteria and 65 by second treatment. Based on the results of the Cox analysis, points were assigned based on PSA (0 to 4 points), Gleason score (0 to 3), T stage (0 to 1), age (0 to 1) and percent of biopsy positive cores (0 to 1). The UCSF-CAPRA score range is 0 to 10, with roughly double the risk of recurrence for each 2-point increase in score. Recurrence-free survival at 5 years ranged from 85% for a UCSF-CAPRA score of 0 to 1 (95% CI 73%–92%) to 8% for a score of 7 to 10 (95% CI 0%–28%). The concordance index for the UCSF-CAPRA score was 0.66.

Conclusions: The UCSF-CAPRA score is a straightforward yet powerful preoperative risk assessment tool. It must be externally validated in future studies.

KEY WORDS: prostatic neoplasms, risk factors, prognosis, prostate-specific antigen

An estimated 230,110 new cases of prostate cancer were predicted for 2004 and 29,900 men were expected to die of the disease.1 While this mortality figure is second only to lung cancer, the high diagnosis-to-mortality ratio underscores the well documented quandary presented by the fact that many men diagnosed with prostate cancer do not die of the disease.2 Definitive local therapy yields excellent long-term survival rates, and has been shown to reduce prostate cancer metastases and cause specific mortality.3 However, all available active treatments may exert a significant impact on patient health related quality of life.4 Therefore, clinicians must attempt to determine at diagnosis who among their patients might do well with active surveillance, who should receive immediate local treatment, who requires aggressive multimodal therapy and who should be treated presump-tively for advanced disease.

Submitted for publication September 13, 2004.

Supported by TAP and the National Institutes of Health/National Cancer Institute University of California, San Francisco SPORE Grant p50 e28572.

CaPSURE is a research collaboration between UCSF and TAP Pharmaceutical Products, Inc. (Lake Forest, Illinois).

Correspondence: UCSF/Mt. Zion Cancer Center, 1800 Divisadero St., 3d Floor, San Francisco, California 94115-1711 (telephone: 415-353-7098; FAX: 415-353-7093; e-mail: pcarroll@urol.ucsf.edu).

See Editorial on page 1848.
**Variables definitions.** The prostate specific antigen (PSA) value used was the highest PSA recorded in the 9 months before diagnosis. The 2002 clinical TNM stage was the highest reported from 1 month before to 3 months after diagnosis. Highest total and highest primary Gleason scores were recorded from the diagnostic biopsy site. Percent positive biopsies (PPB) was calculated from detailed reported biopsy data. Disease recurrence after radical prostatectomy (RP) was defined as 2 consecutive PSA values of 0.2 ng/ml or greater at any time postoperatively, or any additional treatment more than 6 months after RP. The date of recurrence was defined as the earlier of the second PSA 0.2 ng/ml or greater, or the date additional treatment was initiated. If there was no disease recurrence patient followup was censored at the date of the last recorded PSA.

**Patient cohort.** As of July 2003, 10,018 patients were enrolled in CaPSURE. Of these patients, 4,128 elected RP as the primary treatment for prostate cancer. We included 2,154 patients diagnosed between 1992 and 2001 with clinically localized disease (clinical stage T1c-3a, N0/x, M0/x) who did not receive neoadjuvant or adjuvant radiation or hormonal therapy. We excluded from study patients with unknown PSA, Gleason score, clinical T stage or PPB. We also limited analysis to patients with at least a sextant biopsy, PSA of 2 ng/ml or greater at diagnosis, and at least 2 followup PSA or evidence of additional treatment more than 6 months after RP. A total of 1,439 patients meeting these criteria constituted our analytic dataset.

**Development of the University of California, San Francisco-Cancer of the Prostate Risk Assessment (UCSF-CAPRA).** Our goal in developing this predictive index was to maximize the ability of the score to predict disease-free survival while maintaining the simplicity and clinical applicability of the tool. The variables initially considered for inclusion in the index included PSA, Gleason score, T stage, PPB, age at diagnosis and ethnicity. We began by including all of these variables in a Cox proportional hazards model with detailed diagnosis and ethnicity. We began by including all of these variables in a Cox proportional hazards model with detailed categories (PSA as 2 to 4, 4.1 to 6, 6.1 to 8, 8.1 to 10, 10.1 to 20, 20.1 to 30, greater than 30; Gleason as 1–2/1–2, 1–2/3, 3/1–2, 3/3, 3/4–5, 4–5/1–3, 4–5/4–5; T stage as T1c, T2a, T2b, T2c, T3a; PPB as less than 15%, 15% to 25%, 26% to 33%, 34% to 50%, 51% to 66%, 67% to 79%, 80% or greater; age as younger than 50 years, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 or older; and ethnicity as black, white or other).

The results of the initial model were reviewed to determine whether any variables could be eliminated and which category levels could be collapsed. Ethnicity was not a significant predictor of disease-free survival (hazard ratio [HR] for black ethnicity 1.16, p = 0.53, HR for other/unknown ethnicity 0.31, p = 0.10). Further refinement of the model was accomplished primarily through an examination of the model’s parameter estimates (PEs). Category levels within each variable with similar PEs could be combined to reduce the model. This process was repeated iteratively until the final model was reached.

We also built models incorporating various mathematical transformations of PSA including linear, logarithmic, truncated, sigmoidal, cubic spline and piecewise linear. In fact there was a slight improvement in the performance of the model using the piecewise linear PSA function rather than categorized PSA levels. However, the relatively small gain in accuracy would not be worth the large loss in simplicity and ease of use and, therefore, categorized PSA was retained in the final model. Finally, in addition to the various categorizations of PPB, we examined the percent of biopsy cores positive from the more involved side of the prostate only, as well as the absolute number, rather than percentage, of cores positive. However, these variables did not provide any additional predictive information.

Once the final model was specified, the PEs were used to assign points for each level of the variables in the model. We decided that each 2-point increase in the final index should represent approximately a doubling of risk for the outcome of disease recurrence. Proportional hazards model PEs are calculated on the log scale, thus a PE of 0.7 would result in a doubling of risk and each 0.35 increase in PE would be worth 1 point. Points were thus assigned to each level of the final variables in the index. The final UCSF-CAPRA score for each patient was calculated by summing the points for each variable in the model.

**Predictive performance of the UCSF-CAPRA.** UCSF-CAPRA scores were calculated for the men in the analytic population and included in Cox proportional hazards regression models with HR calculated for each UCSF-CAPRA score. Life table and Kaplan-Meier analyses were used to determine the probability of DFS at 3 and 5 years for each UCSF-CAPRA score level.

**Validation tests.** We calculated Kattan nomogram scores for each man in the dataset, and classified each according to a modification of the D’Amico et al risk groupings. A patient with PSA less than 10 ng/ml, no Gleason pattern 4 or 5 disease and a clinical stage of T1 or T2a was low risk. Intermediate risk patients were those with PSA 10.1 to 20 ng/ml, Gleason 7 or less with or 4 or 5 as the secondary pattern, or clinical stage T2b-c. High risk patients had PSA greater than 20 ng/ml, Gleason greater than 7 or 4 or 5 as the primary pattern, or clinical stage T3a.

Relationships among the UCSF-CAPRA, Kattan and D’Amico scores were assessed with Pearson correlation coefficients (r), frequency analysis of D’Amico categories by UCSF-CAPRA level and mean Kattan nomogram score per UCSF-CAPRA level. We also calculated the concordance index (c-index) for each algorithm. The c-index in survival analysis is the proportion of randomly paired patients for whom the patient with the higher probability of recurrence (higher UCSF-CAPRA score, higher D’Amico risk category, lower Kattan nomogram score) also had the earlier observed disease recurrence. The concordance index ranges from 0 to 1, with 1 indicating perfect concordance and 0.5 indicating no concordance. All analyses were performed using SAS® version 8.2, except for the c-index, which was calculated using S-PLUS® version 6.0.

**RESULTS**

Of the 1,439 patients, 210 (15%) had recurrence at a median of 21 months. Among these patients treatment failed in 145 (69%) by PSA criteria and in 65 (31%) by second treatment. Those in whom treatment failed by the second treatment criterion did so at a median of 45 months after prostatectomy with a median PSA of 0.3 ng/ml at second treatment. The remaining 1,229 patients were censored at a median of 24 months. Mean patient age was 62 years. There were 122 black women (8.4%), 52 (3.6%) were of other or unknown ethnicity and the remainder were white.

The final UCSF-CAPRA scoring system is presented in table 1, which also illustrates the distribution of patients and crude recurrence rates across the levels of these variables. Points for each variable are totaled to yield a final UCSF-CAPRA score of 0 to 10. The parameter estimates and HRs for each are given in table 2.

Table 3 presents the distribution of patients across the possible UCSF-CAPRA scores as well as the crude recurrence rates for each score. Of the patients 88% had scores in the range of 1 to 4, only 2% had scores greater than 6 and none had a score of 10. Crude recurrence rates were 10.9% for patients with scores 0 to 3 and 78.6% for those with scores of 7 or greater. In the Cox regression analysis, the PE for each incremental point in the index was 0.42 (p < 0.0001), with a HR for recurrence of 1.5 (95% CI 1.4–1.6). This result indicates roughly a doubling of risk of recurrence on average with every 2-point increase in UCSF-CAPRA score. The PEs and
HR values with 95% CI for each individual score are presented in Table 4. This table also includes the results of the Kaplan-Meier analysis. Patients with UCSF-CAPRA scores of 0 or 1 had 3 and 5-year recurrence-free survival (RFS) rates of 91% and 85%, respectively. For those with UCSF-CAPRA scores of 7 or more these rates were 24% and 8%, respectively. The figure presents the survival curves.

Correlation among the UCSF-CAPRA and the D'Amico and Kattan instruments is presented as $r = 0.77$ between UCSF-CAPRA and Kattan, $r = 0.74$ between UCSF-CAPRA and D'Amico, and $r = 0.71$ between Kattan and D'Amico (all $p < 0.0001$). Table 5 presents the median Kaplan score and distribution of D'Amico risk groups at each UCSF-CAPRA score. The c-indexes for the 3 systems were 0.63 (range 0.55 to 0.72) for D'Amico, 0.65 (range 0.56 to 0.75) for Kattan and 0.66 (range 0.57 to 0.75) for UCSF-CAPRA.

**DISCUSSION**

As of April 2000 Ross et al had counted 42 nomograms published for risk assessment at the various stages of prostate cancer while others have been published more recently. Another offering to this already rich literature certainly must be justified. The D'Amico classification performs well in identifying patients at low risk but there is significant overlap among patients in the intermediate and high risk categories. This is not surprising since the classification does not account for multiple adverse risk factors, such that a patient with PSA 19 ng/ml, Gleason 4, stage T2b disease, for example, would be considered at lower risk than one with PSA 5 ng/ml, Gleason 4, stage T1c disease.

Other instruments in current use such as the Kattan nomogram or the tables published by Partin et al, better integrate multiple risk factors, but are difficult to apply without paper copies of the instruments at hand. A handheld computer version of the Kattan nomogram is available but with this approach the derivation of the predicted outcome score, and the relative weights given each variable, become completely opaque to clinician and patient. We have also recently shown that the Kattan nomogram may overestimate the likelihood of RFS in the community setting, especially among patients with relatively low risk tumors.
The UCSF-CAPRA score performed well in this community based cohort with a concordance index comparable to that of the Kattan nomogram. As should be expected from an RP cohort, the distribution of scores was shifted toward low scores, representing lower risk. Consistent with other nomograms, PSA and Gleason score remain the greatest predictors of risk. In the absence of active treatment there is natural fluctuation in PSA levels. The decision to use the highest PSA before treatment was arbitrary, but the highest and last PSA levels were different among only 4.6% of patients. Moreover, only 1.6% would have a different UCSF-CAPRA score if the last rather than the highest PSA were used. We chose to exclude from study 75 patients with PSA less than 2 ng/ml who constitute a small fraction of the patients in CaPSURE and do not represent typical patients with prostate cancer. They had unusually indolent tumors, and to include them would have precluded the development of a tool applicable to the majority of contemporary newly diagnosed patients.

Similar reasoning led to the exclusion of patients from study with fewer than 6 biopsy cores taken. Again, we designed an instrument germane to contemporary patients newly diagnosed with prostate cancer in 2004, and to include those with fewer cores would have significantly distorted the analysis of PPB. The importance of systematic biopsy is increasingly well recognized among CaPSURE urologists. By 2002 only 2.8% of patients in our cohort had fewer than 6 biopsy cores taken. Therefore, we excluded these patients from analysis at the cost of excluding more patients diagnosed earlier in the study period and, thus, shortening our mean followup time.

Clinical T stage as assessed by digital rectal examination was not a significant predictor of outcome in our model except in the case of palpable extracapsular extension (stage cT3a) which raises the score by 1. Age at diagnosis is not a significant variable in most extant nomograms. Although there was not a statistically significant difference in our dataset, the parameter estimate for men younger than 50 years compared with those older than 50 years at diagnosis was large enough to merit inclusion in the UCSF-CAPRA score. Of the

### Table 4. Results of Cox model regression and Kaplan-Meier analysis

<table>
<thead>
<tr>
<th>UCSF-CAPRA Score</th>
<th>Parameter Estimate</th>
<th>p Value</th>
<th>HR</th>
<th>HR 95% CI</th>
<th>3-Yr %RFS (95% CI)</th>
<th>5-Yr %RFS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1 Reference</td>
<td>0.25</td>
<td>0.31</td>
<td>1.28</td>
<td>0.79–2.08</td>
<td>91 (85–95)</td>
<td>85 (73–92)</td>
</tr>
<tr>
<td>2</td>
<td>0.86</td>
<td>&lt;0.001</td>
<td>2.36</td>
<td>1.49–3.72</td>
<td>89 (83–94)</td>
<td>81 (69–89)</td>
</tr>
<tr>
<td>3</td>
<td>0.87</td>
<td>0.001</td>
<td>2.38</td>
<td>1.40–4.03</td>
<td>81 (73–87)</td>
<td>66 (54–76)</td>
</tr>
<tr>
<td>4</td>
<td>1.20</td>
<td>&lt;0.001</td>
<td>3.32</td>
<td>1.89–5.80</td>
<td>69 (51–82)</td>
<td>59 (40–74)</td>
</tr>
<tr>
<td>5</td>
<td>1.96</td>
<td>&lt;0.001</td>
<td>7.11</td>
<td>3.84–13.15</td>
<td>54 (27–75)</td>
<td>34 (12–57)</td>
</tr>
<tr>
<td>6 or Greater</td>
<td>2.56</td>
<td>&lt;0.001</td>
<td>17.38</td>
<td>9.92–30.46</td>
<td>24 (9–43)</td>
<td>8 (0–28)</td>
</tr>
</tbody>
</table>

### Table 5. Comparison of UCSF-CAPRA scores with Kattan scores and D’Amico classification

<table>
<thead>
<tr>
<th>UCSF-CAPRA Score</th>
<th>Kattan Median</th>
<th>Kattan Range</th>
<th>D’Amico Group % Low</th>
<th>% Intermediate</th>
<th>% High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1 Reference</td>
<td>91</td>
<td>79–96</td>
<td>76</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td>65–96</td>
<td>58</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>47–96</td>
<td>16</td>
<td>78</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
<td>39–91</td>
<td>0</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>69</td>
<td>27–84</td>
<td>0</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td>6 or Greater</td>
<td>60</td>
<td>21–83</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Survival curves for 5-year recurrence-free survival among patients with each UCSF-CAPRA score

### Table 5. Comparison of UCSF-CAPRA scores with Kattan scores and D’Amico classification

The UCSF-CAPRA score performed well in this community based cohort with a concordance index comparable to that of the Kattan nomogram. As should be expected from an RP cohort, the distribution of scores was shifted toward low scores, representing lower risk. Consistent with other nomograms, PSA and Gleason score remain the greatest predictors of risk. In the absence of active treatment there is natural fluctuation in PSA levels. The decision to use the highest PSA before treatment was arbitrary, but the highest and last PSA levels were different among only 4.6% of patients. Moreover, only 1.6% would have a different UCSF-CAPRA score if the last rather than the highest PSA were used. We chose to exclude from study 75 patients with PSA less than 2 ng/ml who constitute a small fraction of the patients in CaPSURE and do not represent typical patients with prostate cancer. They had unusually indolent tumors, and to include them would have precluded the development of a tool applicable to the majority of contemporary newly diagnosed patients.

Similar reasoning led to the exclusion of patients from study with fewer than 6 biopsy cores taken. Again, we designed an instrument germane to contemporary patients newly diagnosed with prostate cancer in 2004, and to include those with fewer cores would have significantly distorted the analysis of PPB. The importance of systematic biopsy is increasingly well recognized among CaPSURE urologists. By 2002 only 2.8% of patients in our cohort had fewer than 6 biopsy cores taken. Therefore, we excluded these patients from analysis at the cost of excluding more patients diagnosed earlier in the study period and, thus, shortening our mean followup time.

Clinical T stage as assessed by digital rectal examination was not a significant predictor of outcome in our model except in the case of palpable extracapsular extension (stage cT3a) which raises the score by 1. Age at diagnosis is not a significant variable in most extant nomograms. Although there was not a statistically significant difference in our dataset, the parameter estimate for men younger than 50 years compared with those older than 50 years at diagnosis was large enough to merit inclusion in the UCSF-CAPRA score. Of the
patients in our cohort 4% were younger than 50 years and were half as likely to have recurrence as men older than 50 years. This finding is consonant with recent reports from academic series which report better pathological and biochemical outcomes among men younger than 50 years. Evidence from numerous studies during the last several years indicates that information derived from the results of the diagnostic biopsy adds significantly to accurate risk assessment among patients with newly diagnosed localized disease. In the UCSF-CAPRA model PPB only contributed 1 point despite analysis of this variable in multiple forms. It may be the case that variability in the community setting in urologists’ biopsy techniques and/or pathologists’ interpretations of specimens, or some other unexplained factor, reduces the predictive power of this variable relative to that found in academic series. Of note, a recent attempt at improving the Kattan nomogram by including PPB and other biopsy data produced only minor improvements in its predictive power. PPB may also not be the best measure of tumor burden on biopsy. For example, Freedland et al demonstrated that the total percentage of biopsy tissue involved was a better predictor of pathological and biochemical outcomes than the percentage of cores positive. However, this variable is not yet collected in CaPSURE and, therefore, could not be considered for inclusion in UCSF-CAPRA.

CONCLUSIONS

Within this large cohort of primarily community based RP cases, the UCSF-CAPRA score proved to have predictive accuracy for biochemical recurrence comparable to the Kattan nomogram, and is significantly easier to calculate and apply both for clinical and research purposes. This novel instrument clearly must be validated in other cohorts of RP cases, and we also plan to test the instrument in cohorts of patients who underwent radiotherapy. We recognize that biochemical recurrence is not an ideal proxy for disease specific or overall mortality, and we will investigate the UCSF-CAPRA score’s ability to predict these outcomes as more of the patients in CaPSURE reach them. No nomogram can replace clinical decision making for men with localized prostate cancer. Patients and physicians must weigh factors such as patient preferences for various quality of life outcomes, as well as risk of disease recurrence or progression. Nonetheless, the UCSF-CAPRA offers practical guidance for these decisions and will also serve as a useful risk stratification tool for future prostate cancer studies.

REFERENCES

17. Freedland, S. J., Casathy, G. S., Dorey, F. and Aronson, W. J.: Percent prostate needle biopsy tissue with cancer is more predictive of biochemical failure or adverse pathologic after radical prostatectomy than prostate specific antigen or Gleason score. J Urol, 167: 516, 2002