Evaluating a Marker's Contribution to a Nomogram: The GEMCaP Example

Michael W. Kattan

A postoperative nomogram developed in 1999 by Kattan et al. has been externally validated, but needs improvement. This tool predicts well but not perfectly. The Genomic Evaluators of Metastatic Prostate Cancer (GEMCaP) biomarkers hold promise for improving this tool; however, a larger data set that permits more involved analyses is needed. Clin Cancer Res; 16(1): 1–3. ©2010 AACR.

In the current issue of Clinical Cancer Research, Paris et al. (1) have worked to extend the predictive performance of a postoperative nomogram, a tool which predicted the probability that a man's prostate cancer would progress following surgery. Paris et al. extended the performance of the tool by adding genomic markers as additional prognostic factors. A particularly attractive aspect of the Paris et al. study is the recognition of the complexity involved in demonstrating new marker value. Most new marker evaluation studies fail to appreciate how a new marker should be evaluated. Clearly, this evaluation must occur before the translational implication of the marker work can be assessed.

In 1999, Kattan et al. first published on the postoperative nomogram tool (2). The tool used postoperative information and predicted a 7-year end point (see Fig. 1). By Cox proportional hazards regression analysis, they modeled the clinical and pathologic data and disease follow-up for 996 men with clinical stage T1a to T3c N0M0 prostate cancer that were treated with radical prostatectomy by a single surgeon at our institution. The predictors included pretreatment serum prostate-specific antigen level, specimen Gleason sum, prostatic capsular invasion, surgical margin status, seminal vesicle invasion, and lymph node status. The outcome predicted was defined as either clinical evidence of disease recurrence, a rising serum prostate-specific antigen level (two measurements of 0.4 ng/mL or greater and rising), or initiation of adjuvant therapy.

The original tool has been tested in a variety of settings with generally good performance characteristics. For example, Graefen et al. used clinical and pathologic data of 2,908 patients from four international institutions for validation, with 2,465 complete records (3). The nomogram-predicted probabilities of 7-year freedom from recurrence were compared with actual follow-up. First, the concordance index was calculated for all patients and stratified by the time period of surgery. Second, calibration of the nomogram was measured by comparing the predicted freedom from recurrence with that of an ideal nomogram. The overall concordance index was 0.80 when applied to the validation data set, with individual institution concordance indices ranging from 0.77 to 0.82. The predictive accuracy of the nomogram seemed to be higher in patients who were operated on between 1997 and 2000 [area under the receiver operating characteristic curve (AUC), 0.83] compared with those treated between 1987 and 1996 [AUC, 0.78].

A tool like this has a few different uses. First, it serves a role in patient counseling when the patient wants to know his prognosis. Nomograms are typically more accurate than other approaches which might be easier to use, such as risk group assignment or tables of probabilities. Second, the tool can be used for identifying patients who are in need of an adjuvant therapy. Although the nomogram does not predict the success of the adjuvant therapy, it at least predicts how well the absence of adjuvant therapy is likely to perform. Third, patients at high risk could be identified for possible eligibility in clinical trials. Clinical trials should generally be restricted to the patients at highest risk for reasons of ethics and clinical trial efficiency, and nomograms, through accurate risk assessment, are attractive for this purpose.

However, the postoperative nomogram does not predict perfectly and needs improvement. Stephenson et al. enhanced the tool in several ways (4). They extended the predictions to 10 years after radical prostatectomy and have enabled the nomogram predictions to be adjusted for the disease-free interval that a patient has maintained after radical prostatectomy. Cox regression analysis was used to model 1,881 patients that underwent radical prostatectomy for clinically localized prostate cancer by two high-volume surgeons. The model was externally validated on two independent cohorts of 1,782 patients and 1,357 patients, respectively. The end point predicted was a rising prostate-specific antigen level, clinical progression, radiotherapy more than 12 months postoperatively, or
initiation of systemic therapy. The concordance index of the nomogram, when applied to the independent validation sets, was 0.81 and 0.79. However, it is not clear whether the enhanced tool offers improvement in predictive accuracy. Kattan et al. examined whether adding surgeon experience (i.e., the number of cases performed by the surgeon prior to the current case) improved the predictive accuracy of the postoperative nomogram (5). As measured by the concordance index, they found that inclusion of surgeon experience had little effect on the predictive accuracy of the postoperative nomogram. It has been argued that a change in the concordance index is not very sensitive to an improvement in prediction accuracy (6), but a mere examination of the effect on predictions reveals that accommodating surgeon experience in the predicted probability has negligible results. The predictions with and without surgeon experience were very similar.

It is very attractive to evaluate a new marker by comparing predictions and predictive accuracy from models that lack and contain a novel marker. Incremental predictive accuracy is a more direct evaluation of the benefit of a new marker, more interpretable than a P value or hazard ratio. There are several reasons for this belief (7). First, an individual patient’s optimal prediction, generally, will come from a multivariable model. Almost never would a single marker, absent any modeling, be ideal for prediction. If a model of markers provides the most accurate prediction, models of markers should be evaluated. Second, the P value tests whether the association with the marker is 0, which is not testing the question of direct interest:

---

**Fig. 1.** Postoperative nomogram for predicting prostate-specific antigen recurrence after radical prostatectomy. For prostatic capsular invasion, "None" refers to L0-1, "Inv. Capsule" refers to L2, "Focal" refers to L3F, and "Established" refers to L3E. Reprinted with permission. ©2008 American Society of Clinical Oncology. All rights reserved. Kattan et al. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. J Clin Oncol, vol. 17 (5), 1999: 1499-1507.

---

**Postoperative Nomogram for Prostate Cancer Recurrence**

<table>
<thead>
<tr>
<th>Points</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preop PSA</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Gleason Sum</td>
<td>53</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pros. Cap. Inv.</td>
<td>None</td>
<td>Pos</td>
<td>Focal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical Margins</td>
<td>Neg</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seminal Ves. Invasion</td>
<td>No</td>
<td>Pos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph Nodes</td>
<td>Neg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total Points** 0 10 20 30 40 50 60 70 80 90 100

| 84-Month Recurrence Free Prob. | 0.99 | 0.98 | 0.95 | 0.9 | 0.8 | 0.7 | 0.5 | 0.3 | 0.1 | 0.01 |

**Instructions for Physician:** Locate the patient’s PSA on the PSA axis. Draw a line straight upwards to the Points axis to determine how many points towards recurrence the patient receives for his PSA. Repeat this process for the other axes, each time drawing straight upward to the Points axis. Sum the points achieved for each predictor and locate this sum on the Total Points axis. Draw a line straight down to find the patient’s probability of remaining recurrence free for 84 months assuming he does not die of another cause first.

**Instruction to Patient:** "Mr. X, if we had 100 men exactly like you, we would expect between-<predicted percentage from nomogram = 10%> and <predicted percentage = 10%> to remain free of their disease at 7 years following radical prostatectomy, and recurrence after 7 years is very rare."

1998 Michael W. Kattan and Peter T. Scardino
MSKCC Division of Urology
whether a new marker improves predictive ability. These are different questions (8). Third, when examining the $P$ value for a novel marker, this value may depend on the other variables in the multivariable model. For example, the use of cutoffs or transforms for the established marker could affect the $P$ value of the new marker. A comparison of the best models with and without a marker of interest provides a more objective alternative because the emphasis is shifted to the predictive accuracies of the models; the modeling should be used that provides the most accurate predictions, an objective goal.

This concept is to compare the predictive accuracy of the best model one can make which contains the new marker, and compare that with the accuracy of the best model one can make which lacks the new marker. The accuracy comparison should be reflective of what is expected when each model is applied to prospective patients (i.e., using a validation data set or otherwise corrected for optimism).

In the current issue of Clinical Cancer Research (1), Paris et al. examined three versions of a set of biomarkers they call GEMCaP (Genomic Evaluators of Metastatic Prostate Cancer). They found that two of the three versions seemed to improve the concordance index. However, due to sample size concerns, Paris et al. were not able to examine a continuous GEMCaP value added to a continuous nomogram prediction in a fully optimism-corrected manner (e.g., bootstrapped). If future research is able to show this, and that the alteration in predicted probabilities from these types of models is clinically significant, GEMCaP might be a very valuable addition to our ability to predict progression probabilities.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

Received 11/25/09; accepted 11/29/09; published OnlineFirst 12/22/09.

**References**

Evaluating a Marker's Contribution to a Nomogram: The GEMCaP Example

Michael W. Kattan

Clin Cancer Res  Published OnlineFirst December 22, 2009.