

Comparison of Nomograms With Other Methods for Predicting Outcomes in Prostate Cancer: A Critical Analysis of the Literature

Shahrokh F. Shariat,¹ Pierre I. Karakiewicz,² Nazareno Suardi,² and Michael W. Kattan³

Abstract Purpose: Accurate estimates of risk are essential for physicians if they are to recommend a specific management to patients with prostate cancer. Accurate risk estimates are also required for clinical trial design, to ensure homogeneous patient groups. Because there is more than one model available for prediction of most outcomes, model comparisons are necessary for selection of the best model. We describe the criteria based on which to judge predictive tools, describe the limitations of current predictive tools, and compare the different predictive methodologies that have been used in the prostate cancer literature.

Experimental Design: Using MEDLINE, a literature search was done on prostate cancer decision aids from January 1966 to July 2007.

Results: The decision aids consist of nomograms, risk groupings, artificial neural networks, probability tables, and classification and regression tree analyses. The following considerations need to be applied when the qualities of predictive models are assessed: predictive accuracy (internal or ideally external validation), calibration (i.e., performance according to risk level or in specific patient subgroups), generalizability (reproducibility and transportability), and level of complexity relative to established models, to assess whether the new model offers advantages relative to available alternatives. Studies comparing decision aids have shown that nomograms outperform the other methodologies.

Conclusions: Nomograms provide superior individualized disease-related risk estimations that facilitate management-related decisions. Of currently available prediction tools, the nomograms have the highest accuracy and the best discriminating characteristics for predicting outcomes in prostate cancer patients.

Approximately 680,000 men are diagnosed with prostate cancer worldwide each year (1). In the United States, this cancer is the most common solid malignancy and the second leading cause of cancer death in men (2). Accurate estimates of the likelihood of treatment success, complications, and long-term morbidity are essential for patient counseling and informed decision making. Properly informing the patient of the likelihood of success and morbidity will improve patient satisfaction after treatment (3), particularly when complications arise (4). Accurate risk estimates are also required for clinical trial design, to ensure homogeneous patient groups for whom new cancer therapeutics will be investigated.

Traditionally, physician judgment has formed the basis for risk estimation, patient counseling, and decision making. However, humans have difficulty with predicting outcomes due to the biases that exist at all stages of the prediction process

(5–7). Another mode for risk estimation commonly used by clinicians and patients consists of the use of averages for risk or patient categories. With this method, all patients that are included within one category are given the same risk level. Although, the risk may vary within a given risk stratum, this approach offers no possibility for individualization.

To obviate this problem and to obtain more accurate predictions, researchers have developed predictive (probability of an outcome without considering the effect of time) and prognostic (probability of an outcome over time) tools that are based on statistical models. In general, these models have been shown to perform as well as or better than clinical judgment when predicting probabilities of outcome (8). Within the last 5 years, the number of these predictive tools has increased dramatically. Because there is more than one model available for prediction of most outcomes, model comparisons are necessary to identify the most suitable model for a specific application. In this article, we describe the criteria based on which to judge predictive tools. We then describe the limitations of current predictive tools and compare the different predictive methodologies that have been used in the prostate cancer literature (Table 1).

Evaluating Predictive Tools

Decision aids consist of the nomograms (9, 10), risk groupings (11–14), artificial neural networks (ANN; ref. 15),

Authors' Affiliations: ¹Department of Urology, University of Texas Southwestern Medical Centre, Dallas, Texas; ²University of Montreal, Cancer Prognostics and Health Outcomes Unit, Montreal, Quebec, Canada; and ³Department of Quantitative Health Sciences, The Cleveland Clinic, Cleveland, Ohio
Received 10/23/07; revised 2/17/08; accepted 2/19/08.

Requests for reprints: Shahrokh F. Shariat, Department of Urology, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9110. Phone: 469-363-8500; Fax: 214-648-8786; E-mail: Shahrokh.Shariat@UTSouthwestern.edu.

©2008 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-07-4713

Table 1. Techniques for the development of a clinical prediction model or rule

Technique	Advantages/strengths	Disadvantages/limitations
Univariate analysis	Simple statistical methods Easy clinical application	Reduced accuracy
Multivariable analysis (e.g., logistic regression)	Improved accuracy	More involved statistical methods
Neural network	Relatively ease of clinical application Improved accuracy Incorporation of complex variable relationships	May miss complex variable relationships Difficult clinical application and dissemination Less intuitive Unknown effect of any single variable
Nomogram	Improved accuracy Ease of clinical application	Advanced statistics
CART analysis	Improved accuracy Ease of clinical application Intuitive partitioning	Advanced statistics

probability tables such as the most widely known and applied "Partin staging tables" (16, 17), and classification and regression tree (CART) analyses (18, 19). Despite the apparent differences of these prediction tools, their characteristics can be compared using a common approach. The points of comparisons are based on four characteristics: predictive accuracy, performance characteristics according to risk level, generalizability, and level of complexity.

Predictive accuracy. Accuracy quantifies the ability of the model to discriminate between patients with and without the outcome of interest. The accuracy of the model represents the most important consideration for comparison of different models. Determination of the accuracy of the model requires the application of the model under novel testing conditions different from the development cohort. In the absence of an external cohort, models can be subjected to internal validation. Bootstrapping represents the ideal internal validation format, where the development data set is used to simulate model testing under novel conditions (20–24). Split-sample and crossvalidation (leave-one-out validation) represent alternatives (23).

Different metrics can be used to quantify the accuracy of the model. The receiver operating characteristic area under the curve quantifies the accuracy of binary models that do not rely on censored observations. Conversely, the concordance index applies to models that rely on censored data. The concordance index quantifies the probability that, given two randomly drawn patients, the patient who relapses first had a higher probability of the event of interest.

Calibration. The accuracy of the model indicates the overall ability to predict the outcome of interest. However, overall accuracy does not indicate the ability of the model to predict the outcome of interest in specific patient groups or according to risk level. For example, a model that is 80% accurate may predict virtually perfectly well in high-risk patients but may show dismal performance in low-risk ones. The relationship between predicted risk and observed rate of the outcome of interest should be provided for each new model, along with its overall accuracy. Calibration plots provide this type of information and can be obtained for internal as well as external data (20, 22–25).

Generalizability. Due to differences in the patterns of early detection and in the extent of screening, the characteristics of newly diagnosed prostate cancers might not be the same across

populations (26). These population differences may undermine the accuracy of predictive and prognostic models. Moreover, models may perform better in patients who share a specific characteristic but may show significantly worse performance characteristics in other patients. Therefore, it is imperative that the clinician knows whether a specific model is indeed generalizable to the population they intend to apply it to (20, 22–25).

Level of complexity. The level of complexity of a predictive or prognostic model represents an important practical consideration. Excessively complex models, ones that rely on multiple variables, are clearly impractical in busy clinical practice. Similarly, models that rely on variables that are not routinely available are impractical.

Head-to-head comparison. When judging a new tool, one should examine its predictive accuracy, validity, and performance characteristics relative to established models, with the intent of determining whether the new model offers advantages relative to available alternatives (21, 23, 24, 27–30). Head-to-head comparisons represent the most direct and unbiased comparison of objective attributes (accuracy and performance characteristics) of various models. Subsequently, complexity, generalizability, and other considerations can be compared. With this approach, the alternatives are compared directly, without having to judge the concordance index in isolation or against a possibly arbitrary threshold.

The main steps required in a head-to-head comparison consist of the application of the original model to a common external data set that will serve for testing of all models that will be compared with one another. The regression coefficients, taken from the original model, are then applied to each individual observation to derive probabilities (if logistic regression was originally used) or to define the linear predictor (if Cox regression was used). Either metric is then tested against observed rates of the outcome of interest. This results in accuracy that is defined as either the area under the curve or the c-index, respectively, for binary outcomes or for censored data. These steps are repeated for each of the tested models.

A common mistake consists of refitting a new model that relies on the same variables as the original model and calling it the original model. A second common mistake is to perform an internal validation (for example the development data set is used for bootstrapping) and to interpret it as an external validation.

Limitations of Predictive and Prognostic Tools

Besides obvious limitations related to accuracy, performance characteristics, generalizability, and the level of complexity, the most common potential additional limitations of currently available predictive and prognostic tools may be classified in one or several of the following categories:

Study selection criteria. Specific model criteria, such as inclusion and exclusion criteria, do not allow the use of models for patients with different characteristics or who have been exposed to different treatment modalities. For example, if a model development cohort excluded patients treated with neoadjuvant hormonal therapy, then predictions cannot be

made for such patients. Similarly, models developed in patients treated with external beam radiotherapy cannot be applied to patients treated with brachytherapy, intensity-modulated radiotherapy, or any other treatment modality no matter how similar they might seem.

Change over time of the predictive value of model ingredients. Stage and grade migration represent important phenomena that affect cancer control rates. In general, more contemporary prostate cancer patients are diagnosed with more favorable stage and grade. In consequence, tools require periodic reappraisals in contemporary cohorts to ensure temporal validity.

Adjustment for competing risks. Because of the protracted course of prostate cancer, competing causes of mortality are

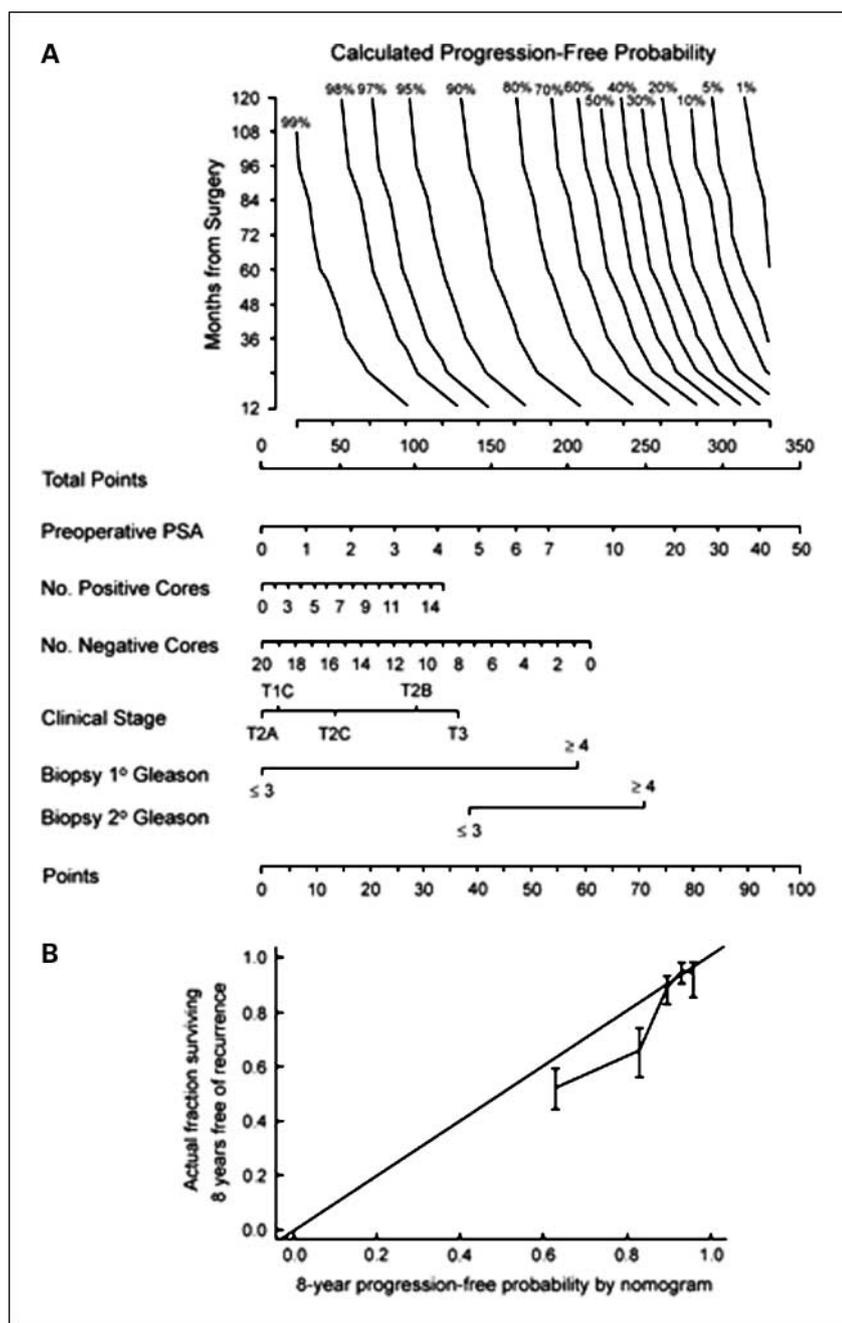


Fig. 1. A, preoperative nomogram estimating the 1- to 10-year biochemical recurrence-free probability after radical prostatectomy alone. B, calibration plot of the nomogram in external validation. The 45° line represents an ideal model in which estimates of recurrence are perfectly calibrated with outcome. Vertical bars, 95% confidence intervals for quintiles in the validation set. Reprinted with permission of the Oxford University Press from Stephenson et al. (35).

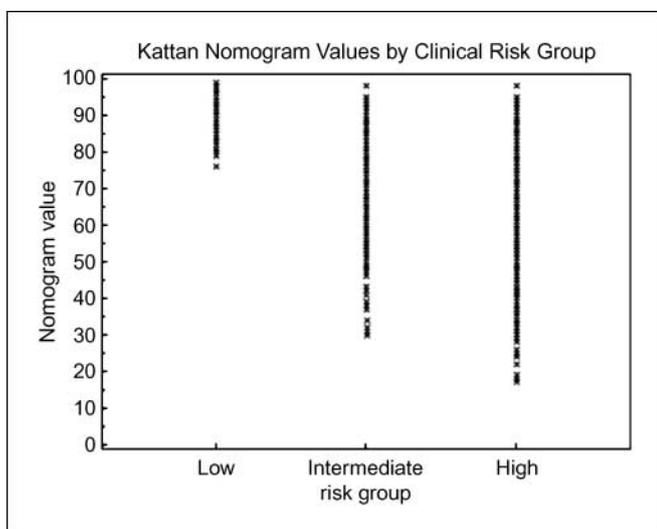


Fig. 2. Preoperative nomogram predicting the 5-year recurrence-free probability after radical prostatectomy (60) for patients classified as low-, intermediate-, or high-risk by D'Amico et al. (41). Reprinted with permission from Mitchell et al. (48). Copyright Elsevier 2005.

extremely important in this patient population (31). Therefore, there is a need for competing-risk modeling to better situate the risk of prostate cancer in the framework of other cause mortality. Such predictions are important to clinicians as well as to patients, especially when overtreatment or suboptimal treatment considerations are addressed. To date, the only modeling tool that allows adjustment for competing risks is the nomogram (32, 33).

Conditional probabilities. Because the risk of disease progression improves with increasing disease-free interval, absence of adjustment for disease-free interval presents the clinician with an excessively somber estimate of cancer control over time. To date, the only modeling tool that allows adjustment for competing risks is the nomogram (Fig. 1; refs. 34, 35).

Nomograms

The statistical definition of a nomogram is a graphical representation of a mathematical formula or algorithm that incorporates several predictors modeled as continuous variables to predict a particular end point based on traditional statistical methods such as multivariable logistic regression or Cox proportional hazards analysis (Fig. 1; refs. 10, 35, 36). By using continuous scales, nomograms calculate the continuous probability of a particular outcome. This obviates the effect of spectrum bias that might be operational, when predictors are stratified. Spectrum bias consists of a forced central effect that is applied to the entire range of observations that decrease within the limits of a given category.

Cubic splines represent one of the strengths of nomogram modeling. Like in neural networks, where all shapes of variables are used, cubic splines allow for nonlinear effects of predictor variables (37). Moreover, nomograms are perfect examples of a predictive or prognostic application that allows graphical representation of variable interactions and depiction of their combined effects.

Comparison of Nomograms with Other Prediction Tools

As more than one model is available for prediction of most outcomes, model comparison is necessary for selection of the best model. Below, we compare the different predictive methodologies that have been used in the prostate cancer literature. Because there is no study comparing the different prediction methods based on prospective data, comparisons are based on retrospective data.

Nomogram versus risk grouping. Physicians often use risk groups to determine the risk of an event. This approach consists of grouping patients with similar characteristics to discriminate between those at low-risk versus those at high-risk for a specific event. Although risk grouping is a logical approach, grouping patients is an inefficient use of the data and tends to reduce the predictive accuracy of a prognostic model (spectrum bias). The misconception related to this approach is that it assumes that all patients within a risk group are equal. However, risk group comprise a heterogeneous group of patients. For example, some patients with clinical stage T_{1c} may have a very favorable prognosis [low prostate-specific antigen (PSA) and biopsy Gleason score of 4-6], whereas others may show less favorable characteristics (elevated PSA and Gleason score of 7-10; ref. 13).

A commonly used risk grouping tool is that developed by D'Amico et al. (11) for pretreatment prediction of biochemical recurrence in patients treated with radical prostatectomy, external-beam radiotherapy, or brachytherapy by placing patients into mutually exclusive risk groups based on clinical stage, biopsy Gleason sum, and pretreatment PSA level (12–14, 38–42). When predicting the outcome for a subset of patients, the relative importance of prognostic variables in another patient group is ignored. In addition, risk grouping requires the conversion of continuous to categorical variables, which limits information about the actual value.

Various studies have documented the superior performance of nomograms compared with risk grouping (21, 27, 28, 30, 43–45). This might stem from the fact that risk groups consist of patients with similar (albeit not identical) characteristics, resulting in heterogeneity within a risk group that reduces the predictive accuracy (28, 45–47). In contrast to risk groups, a nomogram provides an individualized estimate of the predicted probability of the event of interest, which is entirely based on the individual's disease characteristics, without averaging or combining within a category. The heterogeneity inherent in risk groups is illustrated in Fig. 2 (10, 41, 48), where the 5-year recurrence-free probability after radical prostatectomy was calculated using a continuous, multivariable preoperative nomogram among patients classified as low-, medium-, and high-risk using the criteria of D'Amico et al. (41). Although low-risk patients uniformly had a high likelihood of being free of biochemical recurrence based on the probability calculated using the nomogram, a substantial proportion of intermediate- and even high-risk patients had a calculated 5-year recurrence-free probability of $\geq 90\%$. Moreover, a considerable overlap in the risk grouping predictions was evident among intermediate- and high-risk patients.

A risk group is composed of a mixture of patients and is only useful for gauging the prognosis for that group of patients. An individual patient might not be concerned about the outcomes

of his (heterogeneous) group peers. Instead, patients are interested in their own prognosis. Moreover, by incorporating all relevant and informative predictors, nomograms provide more accurate predictions than models based on risk grouping (Kattan, 2001 #4388; Meehl, 1986 #3577; refs. 21, 27, 28, 30, 43–45). Although nomograms are more complex than risk groups, this added complexity results in a better predictive accuracy for both patients and physicians. Moreover, the complexity can be offset, when the electronic versions of nomograms are used.

Finally, the method of counting risk factors/variables should also be avoided because this assumes that each variable exerts an equal prognostic weight on the outcome, which is unlikely to represent the true relationship between variables and prognosis (49–51).

Nomogram versus look-up table. The superior predictive accuracy of multivariable nomograms that rely on variables in their unaltered formats (continuous or categorical) versus look-up tables is illustrated by comparing nomograms with the “Partin tables” (16, 17, 52) predict pathologic features. The Partin tables combined serum PSA level (four categories), clinical stage (seven categories), and biopsy Gleason sum (five categories) to predict the pathologic stage of prostate cancer that is assigned as one of four mutually exclusive

groups, i.e., organ-confined, established extracapsular extension only, seminal vesicle invasion, and lymph node involvement. These tables underestimate the probability of established extracapsular extension as a substantial proportion of patients with lymph node metastases and seminal vesicle invasion will also have established extracapsular extension. Therefore, several studies found that nomograms incorporating PSA level, clinical stage, and Gleason sum modeled as continuous variables had a superior predictive accuracy compared with the Partin tables for predicting organ-confined disease, seminal vesicle invasion, and lymph node invasion (21, 27, 53–57).

Another example of the superiority of nomograms over look-up tables has been shown by Chun et al. (30, 58). They showed that a logistic regression-based nomogram that included preoperative PSA, clinical stage, primary, and secondary biopsy Gleason grades had an accuracy of 80.4% for prediction of the probability of Gleason sum upgrading between biopsy and radical prostatectomy. In contrast, a previously published look-up table (59) based on preoperative PSA, clinical stage, and prostate gland volume had an accuracy of only 52.3% ($P < 0.001$). In addition, the nomogram had a virtually ideal performance, whereas the look-up table had important departures from ideal prediction. Taken together, these findings

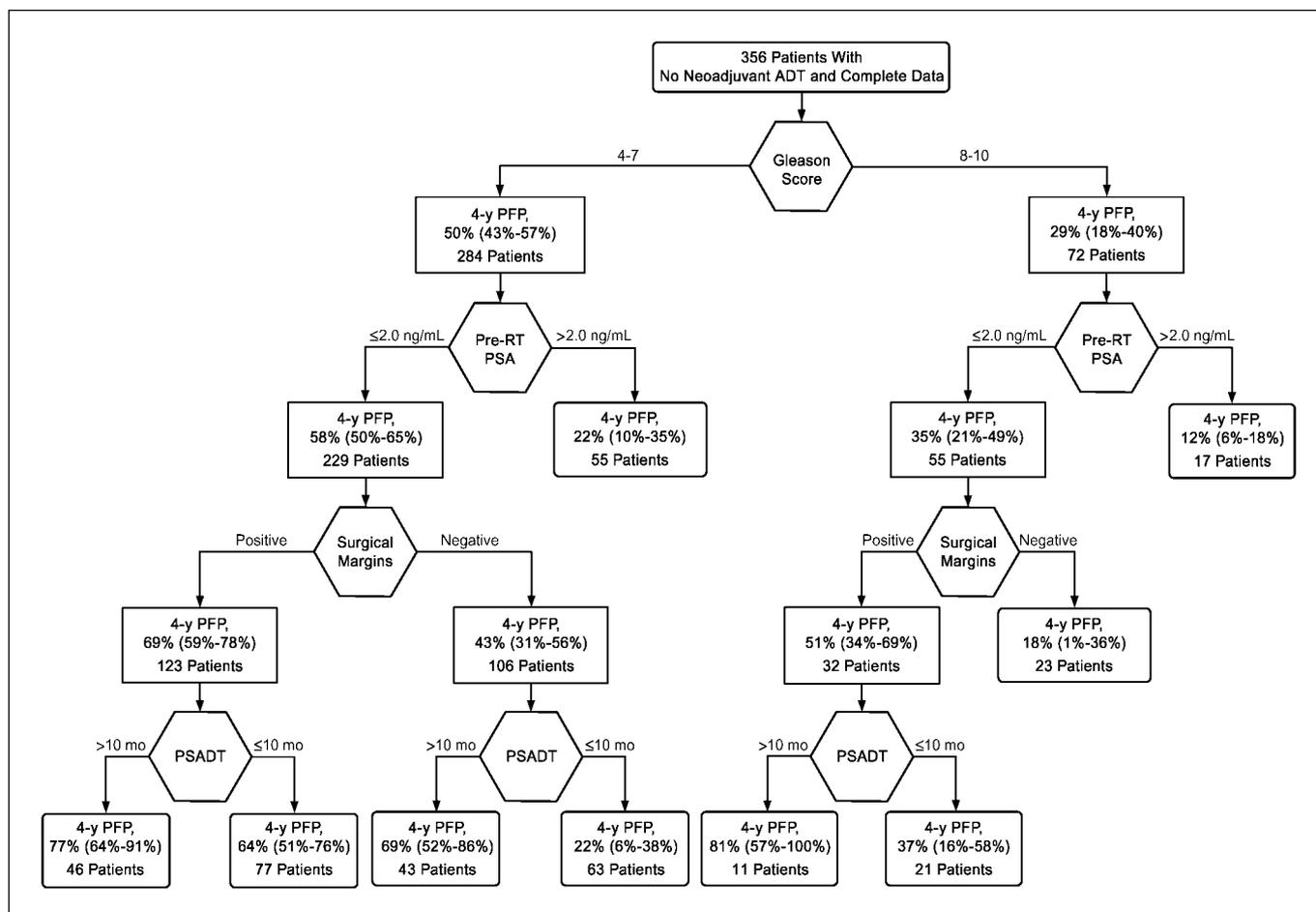


Fig. 3. Four-year actuarial progression-free probability after salvage radiotherapy. Progression-free probability (PFP) stratified by Gleason sum, preradiotherapy PSA level, surgical margins, and PSA doubling time (PSADT). Patients receiving neoadjuvant androgen deprivation therapy (ADT) were excluded from this analysis. All values in parentheses are 95% confidence intervals. RT, radiotherapy. Reprinted with permission from Stephenson et al. (60), JAMA 2004;291:1325–32. Copyright©2004. American Medical Association. All rights reserved.

support that nomograms are more accurate than look-up tables and perform better throughout the range of predicted probabilities.

Nomogram versus tree analysis. CART analysis is another type of predictive model that uses nonparametric techniques to evaluate data. It has the capacity to account for complex relationships, and presents the results in a clinically useful form. In this type of analysis, there is progressive splitting of the population into subgroups that are based on the predictive independent variables. The variables that are chosen, the discriminatory values of the variable, and the order in which the splitting occurs are all produced by the underlying mathematical algorithm to maximize predictive accuracy. A simplified example of a CART-recursive partitioning based on Cox proportional hazards regression analyses is shown in Fig. 3 (60). In this analysis, the clinician simply follows the paths of the tree that best describe the characteristics of the patient being evaluated and arrives at the prediction of the outcome of interest for that particular patient.

Tree analysis is relatively easy to use for the clinician. First, in contrast to many logistic regression models, there are no complicated equations. The structure of the tree is one that is appealing intuitively and congruent with methods of decision-making that a physician already uses on many occasions. For example, in trying to understand the best diagnostic test or treatment for a given patient, clinicians will use specific patient characteristics to determine progressively which modalities are most appropriate or which outcomes are most likely. The CART not only uses this type of logic but also provides a formal structure and quantitative outcome assessment that can optimize the actual clinical decision.

Thus, CART offer greater model-fitting flexibility than traditional statistical methods (61), and theoretically might lead to enhanced predictive accuracy if data sets contain highly predictive nonlinear or interactive effects. However, there are several issues to consider when deciding whether to use CART analysis. It is often important to estimate the overall effect of a single independent variable on the outcome of interest. This is especially true in studies with specific hypotheses about the effects of an independent variable or group of variables on the outcome. Because CART analysis is intended to identify distinct population subgroups, its hierarchical nature does not allow the estimation of net effects of a single variable (62). Regression techniques, however, are largely used to estimate the "average" effect of an independent variable on the probability of having a dependent variable while accounting for others factors. Thus, CART analysis cannot be used as a substitute for proven regression techniques in this type of situation. Moreover, CART analysis can become very complex and difficult to interpret. Trees can grow into multiple levels and thereby result in splits that are not particularly important.

Several studies have shown that traditional statistical methods perform better than CART analysis. For example, using three real world data sets, Kattan (21) found that Cox proportional hazards regression model provided superior predictive accuracy than four tree-based methods. Similarly, Chun et al. (29, 50, 58) compared a CART analysis (19) with a nomogram (63, 64) for prediction of the side of extracapsular extension (30). The nomogram yielded a predictive accuracy of 84% versus 70% for the CART model. Moreover, the nomogram calibration plot was virtually ideal, whereas the

CART calibration plot had appreciable divergence from ideal prediction. Thus, the nomogram was statistically significantly more accurate than the CART model and did better throughout the range of predicted probabilities.

Nomogram versus ANNs. In the last 10 years, a new class of techniques known as ANN has been proposed as a supplement or alternative to standard statistical techniques. For the purpose of predicting medical outcomes, an ANN can be considered a computer intensive classification method. It is a computational method that uses multifactorial analysis. It contains layers of richly interconnected computing nodes, for which weights are adjusted when data are presented to the network during a "training" process. Successful training can result in ANNs that predict output values or recognize patterns in multifactorial data (65). Theoretically, an ANN should have considerable advantages over standard statistical approaches. Neural networks automatically allow arbitrary nonlinear relations between the independent and dependent variables and all possible interactions between the dependent variables. Standard statistical approaches (e.g., logistic or Cox regression) require additional modeling to allow this flexibility. In addition, ANNs do not require explicit distributional assumptions (such as normality). These and other proposed advantages have generated considerable interest in the use of neural network techniques for the classification of medical outcomes.

However, ANNs are not without drawbacks. The primary disadvantage of an ANN is its black box quality, that is, without extra effort, it is difficult if not impossible to gain insight into a problem based on an ANN model. Regression techniques, for example, allow the user to sequentially eliminate possible explanatory variables that do not contribute to the fit of the model. Similarly, based on the underlying statistical theory, regression techniques allow hypothesis testing regarding both the univariate and multivariate association between each explanatory variable and the outcome of interest. Furthermore, it yields other insights into the prediction model, such as hazard ratios and tests of significance for the predictors. These features are not available for ANN. Moreover, regression analysis offer the added advantage of reproducibility and interpretability through the generation of hazard ratios and tests of significance for the predictors (21). The same result is achieved each time it is run on a particular data set, which is not necessarily true for machine learning techniques because they use random processes for sampling and/or coefficient estimation. In addition, regression analyses are common in many statistical software packages and are relatively fast to perform.

Based on a review of 28 studies comparing ANN and regression modeling, Sargent (66) concluded that ANN should not replace standard statistical approaches as the method of choice for the classification of medical data. In the eight largest studies (sample size, >5,000), regression and ANN tied in seven cases, with regression winning in the remaining case. In the more moderate-size data sets, ANN tended to be equivalent or outperform regression, although it is unclear whether this is an artifact due to publication bias. The author pointed out that the regression methods are clearly superior to the ANN with respect to inferences based on the output. Inference and interpretation are frequently key desired outcomes of a modeling exercise. In addition to insight into the disease process, regression models provide explicit information regarding the relative importance

of each independent variable. This information can be valuable in planning subsequent interventions, in eliminating possibly unnecessary tests and procedures (such as blood or tissue studies that are shown not to relate to the outcome of interest), and in determining which are the most critical data to store in a database. Similarly, in a review of the literature, Schwarzer et al. (67) concluded that machine learning methods often have failed to perform better than traditional statistical methods outlining numerous design flaws in the studies that show the superiority of neural networks. For example, on numerous occasions, the neural network was provided with additional data not available to the statistical method. In addition, many different neural networks were compared with a single statistical model, possibly contributing to a chance finding. In some cases, the wrong statistical model was used as the benchmark.

Terrin et al. (68) did a simulation study that compared the external validity of logistic regression analyses, CART, and ANN on data simulated from a specified population and on data from perturbed forms of the population not representative of the original distribution. They found that logistic regression models had the best performance followed by ANNs, and then CARTs. Similarly, using three real-world data sets, Kattan (21) found that Cox proportional hazards regression model provided comparable or superior predictive accuracy than two neural networks. Likewise, Chun et al. (29) compared an ANN (69) with a nomogram (70) for predicting initial biopsy outcome in a cohort of 3,980 patients subjected to at least an 8-core initial biopsy. The nomogram (70.6%) was 3.6% ($P < 0.001$) more accurate than the ANN (67.0%). The nomogram calibration plot gave virtually ideal predictions. Conversely, the ANN had important departures from ideal predictions, which were manifested by underestimation throughout the range of predicted probabilities. These examples of direct comparison between nomograms and ANNs on the same data set support that nomograms are statistically significantly more accurate and better calibrated than ANNs.

References

- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
- Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007;57:43–66.
- Miles BJ, Giesler B, Kattan MW. Recall and attitudes in patients with prostate cancer. *Urology* 1999;53:169–74.
- Clark JA, Wray NP, Ashton CM. Living with treatment decisions: regrets and quality of life among men treated for metastatic prostate cancer. *J Clin Oncol* 2001;19:72–80.
- Elstein AS. Heuristics and biases: selected errors in clinical reasoning. *Acad Med* 1999;74:791–4.
- Vlaev I, Chater N. Game relativity: how context influences strategic decision making. *J Exp Psychol Learn Mem Cogn* 2006;32:131–49.
- Hogarth RM, Karelaia N. Heuristic and linear models of judgment: matching rules and environments. *Psychol Rev* 2007;114:733–58.
- Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002;20:82–8.
- Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol* 1999;17:1499–507.
- Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90:766–71.
- D'Amico AV, Whittington R, Malkowicz SB, et al. The combination of preoperative prostate specific antigen and postoperative pathological findings to predict prostate specific antigen outcome in clinically localized prostate cancer. *J Urol* 1998;160:2096–101.
- D'Amico AV, Keshaviah A, Manola J, et al. Clinical utility of the percentage of positive prostate biopsies in predicting prostate cancer-specific and overall survival after radiotherapy for patients with localized prostate cancer. *Int J Radiat Oncol Biol Phys* 2002;53:581–7.
- D'Amico AV, Whittington R, Malkowicz SB, et al. Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. *J Clin Oncol* 1999;17:168–72.
- D'Amico AV, Whittington R, Malkowicz SB, et al. Combination of the preoperative PSA level, biopsy gleason score, percentage of positive biopsies, and MRI T-stage to predict early PSA failure in men with clinically localized prostate cancer. *Urology* 2000;55:572–7.
- Snow PB, Smith DS, Catalona WJ. Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study. *J Urol* 1994;152:1923–6.
- Partin AW, Kattan MW, Subong EN, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update [see comments] [published erratum appears in *JAMA* 1997 Jul 9;278(2):118]. *JAMA* 1997;277:1445–51.
- Makarov DV, Trock BJ, Humphreys EB, et al. Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005. *Urology* 2007;69:1095–101.
- Conrad S, Graefen M, Pichlmeier U, et al. Prospective validation of an algorithm with systematic sextant biopsy to predict pelvic lymph node metastasis in patients with clinically localized prostatic carcinoma. *J Urol* 2002;167:521–5.
- Graefen M, Haese A, Pichlmeier U, et al. A validated

Conclusions

The above discussion is meant to provide guidelines in the process of decision aid selection. Continuous, multivariable models such as nomograms are a highly appealing means of calculating accurate predictions with or without the use of a computer. Many nomograms have been constructed for patients with prostate cancer. Nomograms currently represent the most accurate and discriminating tools for predicting outcomes in patients with prostate cancer. When faced with the difficult decision of choosing among the treatment options for each clinical stage of prostate cancer, the nomograms provide patients with accurate estimates of outcomes. Equipped with this information, the patient is more likely to be confident in his treatment decision and less likely to experience regret in the future. However, it should be emphasized that nomogram predictions must be interpreted as such; they do not make treatment recommendations or act as a surrogate for physician-patient interactions, nor do they provide definitive information on symptomatic disease progression or complications associated with treatments.

Many more nomograms, as well as improvements to existing nomograms, are needed. For example, none of the nomograms predicts with perfect accuracy. Novel biomarkers, larger data sets, better data collection methods, and more sophisticated modeling procedures are needed to improve predictive accuracy. In addition, better accuracy might be accomplished by modeling physician and/or hospital-specific data for patients being treated by that physician or at that hospital. Finally, nomograms that predict the likelihood of metastatic progression, cancer-specific mortality, and long-term urinary and sexual function are likely to have great utility for the patient and physician when exploring treatment alternatives.

Disclosure of Potential Conflicts of Interest

M. Kattan is a consultant to Sanofi-Aventis.

- strategy for side specific prediction of organ confined prostate cancer: a tool to select for nerve sparing radical prostatectomy. *J Urol* 2001;165:857–63.
20. Bradley E, Tibshirani, R.J. *Monographs on statistics and applied probability: an introduction to the bootstrap*: Chapman and Hall/CRC; 1993.
21. Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. *J Urol* 2003;170:S6–9; discussion S10.
22. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003;56:441–7.
23. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
24. Steyerberg EW, Harrell FE, Jr., Goodman PH. Neural networks, logistic regression, and calibration. *Med Decis Making* 1998;18:349–50.
25. Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schroder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol* 2007;177:107–12; discussion 12.
26. Briganti A, Shariat SF, Chun FK, et al. Differences in the rate of lymph node invasion in men with clinically localized prostate cancer might be related to the continent of origin. *BJU Int* 2007;100:528–32; Epub 2007 Jun 15.
27. Kattan MW. Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. *Curr Opin Urol* 2003;13:111–6.
28. Kattan MW, Shariat SF, Andrews B, et al. The addition of interleukin-6 soluble receptor and transforming growth factor β 1 improves a preoperative nomogram for predicting biochemical progression in patients with clinically localized prostate cancer. *J Clin Oncol* 2003; 21:3573–9.
29. Chun FK, Graefen M, Briganti A, et al. Initial biopsy outcome prediction-head-to-head comparison of a logistic regression-based nomogram versus artificial neural network. *Eur Urol* 2007;51:1236–40; discussion 41–3.
30. Chun FK, Karakiewicz PI, Briganti A, et al. A critical appraisal of logistic regression-based nomograms, artificial neural networks, classification and regression-tree models, look-up tables and risk-group stratification models for prostate cancer. *BJU Int* 2007;99:794–800.
31. Albertsen PC, Hanley JA, Gleason DF, Barry MJ. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. *JAMA* 1998;280:975–80.
32. Kattan MW, Giri D, Panageas KS, et al. A tool for predicting breast carcinoma mortality in women who do not receive adjuvant therapy. *Cancer* 2004;101: 2509–15.
33. Kattan MW, Heller G, Brennan MF. A competing-risks nomogram for sarcoma-specific death following local recurrence. *Stat Med* 2003;22:3515–25.
34. Stephenson AJ, Scardino PT, Eastham JA, et al. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2005;23:7005–12.
35. Stephenson AJ, Scardino PT, Eastham JA, et al. Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006;98:715–7.
36. Kattan MW. Nomograms. Introduction. *Semin Urol Oncol* 2002;20:79–81.
37. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
38. D'Amico AV, Cote K, Loffredo M, Renshaw AA, Schultz D. Determinants of prostate cancer-specific survival after radiation therapy for patients with clinically localized prostate cancer. *J Clin Oncol* 2002;20: 4567–73.
39. D'Amico AV, Cote K, Loffredo M, Renshaw AA, Schultz D. Determinants of prostate cancer specific survival following radiation therapy during the prostate specific antigen era. *J Urol* 2003;170:S42–6; discussion S6–7.
40. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy or external beam radiation therapy for patients with clinically localized prostate carcinoma in the prostate specific antigen era. *Cancer* 2002;95:281–6.
41. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;280:969–74.
42. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. *J Clin Oncol* 2003;21:2163–72.
43. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. *J Clin Oncol* 2003;21:3647–50.
44. Kattan MW, Leung DH, Brennan MF. Postoperative nomogram for 12-year sarcoma-specific death. *J Clin Oncol* 2002;20:791–6.
45. Shariat SF, Karakiewicz PI, Palapattu GS, et al. Nomograms provide improved accuracy for predicting survival after radical cystectomy. *Clin Cancer Res* 2006;12:6663–76.
46. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for renal cell carcinoma. *J Urol* 2001;166:63–7.
47. Kattan MW, Zelefsky MJ, Kupelian PA, et al. Pretreatment nomogram that predicts 5-year probability of metastasis following three-dimensional conformal radiation therapy for localized prostate cancer. *J Clin Oncol* 2003;21:4568–71.
48. Mitchell JA, Cooperberg MR, Elkin EP, et al. Ability of 2 pretreatment risk assessment methods to predict prostate cancer recurrence after radical prostatectomy: data from CaPSURE. *J Urol* 2005;173:1126–31.
49. Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST-elevation MI: A method for prognostication and therapeutic decision making. *JAMA* 2000;284:835–42.
50. Cooperberg MR, Freedland SJ, Pasta DJ, et al. Multi-institutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. *Cancer* 2006;107:2384–91.
51. Cooperberg MR, Pasta DJ, Elkin EP, et al. 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. *J Urol* 2005;173: 1938–42.
52. Partin AW, Yoo J, Carter HB, et al. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer [see comments]. *J Urol* 1993;150: 110–4.
53. Cagiannos I, Karakiewicz P, Eastham JA, et al. A preoperative nomogram identifying decreased risk of positive pelvic lymph nodes in patients with prostate cancer. *J Urol* 2003;170:1798–803.
54. Kattan MW, Eastham JA, Wheeler TM, et al. Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. *J Urol* 2003;170: 1792–7.
55. Koh H, Kattan MW, Scardino PT, et al. A nomogram to predict seminal vesicle invasion by the extent and location of cancer in systematic biopsy results. *J Urol* 2003;170:1203–8.
56. Briganti A, Chun FK, Salonia A, et al. Validation of a nomogram predicting the probability of lymph node invasion based on the extent of pelvic lymphadenectomy in patients with clinically localized prostate cancer. *BJU Int* 2006;98:788–93.
57. Gallina A, Chun FK, Briganti A, et al. Development and split-sample validation of a nomogram predicting the probability of seminal vesicle invasion at radical prostatectomy. *Eur Urol* 2007;52:98–105.
58. Chun FK, Steuber T, Erbersdobler A, et al. Development and internal validation of a nomogram predicting the probability of prostate cancer Gleason sum upgrading between biopsy and radical prostatectomy pathology. *Eur Urol* 2006;49:820–6.
59. D'Amico AV, Renshaw AA, Arsenault L, Schultz D, Richie JP. Clinical predictors of upgrading to Gleason grade 4 or 5 disease at radical prostatectomy: potential implications for patient selection for radiation and androgen suppression therapy. *Int J Radiat Oncol Biol Phys* 1999;45:841–6.
60. Stephenson AJ, Shariat SF, Zelefsky MJ, et al. Salvage radiotherapy for recurrent prostate cancer after radical prostatectomy. *JAMA* 2004;291: 1325–32.
61. Kattan MW, Cooper RB. A simulation of factors affecting machine learning techniques: an examination of partitioning and class proportions. *Omega Int J Mgmt Sci* 2000;28:501.
62. Marshall DJ. The use of classification and regression trees in clinical epidemiology. *J Clin Epidemiol* 2001;54:603–9.
63. Ohori M, Kattan MW, Koh H, et al. Predicting the presence and side of extracapsular extension: a nomogram for staging prostate cancer. *J Urol* 2004;171: 1844–9; discussion 9.
64. Steuber T, Graefen M, Haese A, et al. Validation of a nomogram for prediction of side specific extracapsular extension at radical prostatectomy. *J Urol* 2006;175: 939–44; discussion 44.
65. Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. *Cancer* 2001;91:1615–35.
66. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001;91:1636–42.
67. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. *Semin Urol Oncol* 2002;20:89–95.
68. Terrin N, Schmid CH, Griffith JL, D'Agostino RB, Selker HP. External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. *J Clin Epidemiol* 2003; 56:721–9.
69. Stephan C, Cammann H, Semjonow A, et al. Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies. *Clin Chem* 2002;48: 1279–87.
70. Karakiewicz PI, Benayoun S, Kattan MW, et al. Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen. *J Urol* 2005;173:1930–4.