# *Prostate cancer nomograms are superior to neural networks*

Pierre I. Karakiewicz, MD,[1,2] Felix K.-H. Chun, MD,[2,3] Alberto Briganti, MD,[2] Paul Perrotte, MD,[1] Michael McCormack, MD,[1] François Bénard, MD,[1] Luc Valiquette, MD,[1] Markus Graefen, MD,[3] Fred Saad, MD[1]

[1]Department of Urology, University of Montreal, Montreal, Quebec, Canada
[2]Cancer Prognostics and Health Outcomes Unit, University of Montreal, Montreal, Quebec, Canada
[3]Department of Urology, University of Hamburg, Germany
[4]Department of Urology, University Vita Salute, Milan, Italy

KARAKIEWICZ PI, CHUN FKH, BRIGANTI A, PERROTTE P, MCCORMACK M, BENARD F, VALIQUETTE L, GRAEFEN M, SAAD F. Principles of prostate cancer nomograms and how do they measure up to neural networks. The Canadian Journal of Urology. 2006;13(Supplement 2):18-25.

***Introduction:*** *Several nomograms have been developed to predict PCa related outcomes. Neural networks represent an alternative.*
***Methods:*** *We provide a descriptive and an analytic comparison of nomograms and neural networks, with focus on PCa detection.*
***Results:*** *Our results indicate that nomograms have several advantages that distinguish them from neural networks. These are both quantitative and qualitative.*
***Conclusion:*** *In the field of PCa detection, nomograms appear to outweigh the benefits of neural networks. However, the neural network methodology represents a valid alternative, which should not be underestimated.*

**Key Words:** nomogram, artificial neural network, prostate cancer, prediction models

## Introduction

The field of prognostics has exploded in the last decade and clinicians have been provided with numerous tools to assist with medical decision-making in the most evidence-based fashion. Most of these tools consist of nomograms, look-up tables and neural network models.[1-16] They address numerous prostate cancer (PCa) outcomes, which range from prediction of biopsy outcome[1] in men considered at risk of PCa to prediction of death from hormone-refractory PCa.[2]

Address correspondence to Dr. Pierre I. Karakiewicz, Cancer Prognostics and Health Outcomes Unit, University of Montreal Health Center (CHUM), 1058, rue St-Denis, Montréal, Québec, H2X 3J4 Canada

## Choice of decision aids

The presence of several decision aids requires a careful selection of tools that should be used for prediction of the outcomes of interest. The following criteria provide an objective and systematic approach in that complex process:

(1) Level of complexity represents an important consideration. Excessively complex models are difficult to integrate in busy clinical practice. For example, lengthy logistic regression equations require the use multiple functions and access to scientific calculators. These are clearly impractical in a busy clinical practice. Neural networks can accurately predict several outcomes of interest.[3-11] Despite high accuracy,

the use of these models is restricted to centers with adequate computer infrastructure, as predictions require access to and expertise with specific software. Look-up tables, such as the Partin Tables[12] or nomograms represent user-friendly alternatives. In either paper based format or within palm digital assistants, they are ideally suited for busy clinical practice.[12-16]

(2) Accuracy represents the second consideration. Current statistical methods offer the possibility of assessing model's predictive accuracy. Usually, it is quantified using receiver operating characteristics derived area under the curve and is expressed as a percentage. Values range from 0.5 to 1.0, where 0.5 is equivalent to a flip of a coin and 1.0 represents perfect prediction. No model is perfect and acceptable accuracy ranges from 70% to 80%.[1-16] Accuracy should be confirmed in either an external cohort or internally, using statistical methods such as bootstrapping.[17,18]

(3) Performance characteristics represent another important consideration. Accuracy indicates the overall ability of the model to predict the outcome of interest. However, the overall predictive accuracy does not inform the user on how good or how bad the predictions may be in specific patient subgroups. Some models may be ideally suited to predict in high-risk patients, but may predict poorly in low risk patients. Other models may predict well throughout the range of predictions.

(4) Model generalizability is important, as patient characteristics can vary. For example, PCa characteristics may not be the same in Europe as in the United States.[14] Prior to using a tool, the clinician should ensure that it was validated in patients with similar disease characteristics.

(5) Finally, when judging a new tool,[19,20] one should examine its accuracy, validity and performance characteristics relative to established models, with the intent of determining whether the new model offers advantages relative to available alternatives.

Availability of several high quality predictive models should encourage the clinician to adopt these tools into everyday clinical practice. Arguments favoring such behavior include standardization of care and of decision-making. Moreover, nomograms predict more accurately than clinicians.[15] For example, Specht and colleagues[15] addressed the ability to predict presence of axillary nodal metastases in women with invasive breast cancer. Nomogram predictions were compared to 17 breast cancer specialists from the Memorial Sloan-Kettering Cancer Center. The nomogram predictions were 18% (p=0.01) more accurate than those of the expert clinicians. This implies that if predictions were made for 100 consecutive women, 18 would have been staged incorrectly if expert clinician predictions were used instead of nomogram predictions. Thus, it appears that nomograms have better ability to predict the outcomes of interest than even expert clinicians. It is conceivable that the advantage related to the use of nomogram predictions may be even more important if clinical ratings were obtained from less expert clinicians.

In addition to methodological and practical considerations, patient perspective also deserves a mention when the use of most unbiased decision-tools is considered. Patients are becoming increasingly aware of the existence of predictive tools. This trend is likely to increase in future years. Patients are also increasingly demanding to actively participate in decision-making, which may in part be explained by the following observations:

(1) Advances in therapeutics have offered numerous treatment options and men no longer accept a paternalistic physician-centered treatment decision-making. Instead, they demand to know the efficacy and detailed side effect profiles of treatment alternatives.

(2) The patient is increasingly recognized as a pivotal player in medical decision-making. Decisions can no longer be made by the physician alone. For example, the American Urological Association suggests a detailed informed consent prior to PSA testing.

(3) Health care 'consumerism' is a growing phenomenon in North America and Europe. Patients select what option of health care to purchase, rather than passively receiving a given treatment modality.

(4) Attention to bioethical considerations has greatly increased over the past decade and has promoted autonomous decision-making.

Thus, it may be postulated that increasingly greater emphasis will be placed on standardized predictions, which will further promote the development of new tools and/or the improvement of existing predictive tools. These considerations may motivate clinicians to adopt the use of decision-tools. Their motivation may also stem from the wealth of clinical data that are used for the development and validation of each model. Most decision tools are based on thousands of observations and it is virtually impossible to achieve that level of clinical exposure and expertise on an individual level. Moreover, most clinicians do not have

the capacity to systematically record or remember the risk characteristics of several thousands of patients. Additionally, unlike computers, clinicians are incapable of systematically and cumulatively processing the recorded risk characteristics and outcomes of historic cases, to derive an estimated probability of outcome for a new case at hand. Thus, it may be expected that the majority of physician-derived estimates are not as accurate as computer-derived decision models.[15] Despite this advantage, these tools are not meant to replace clinical judgment. Their input needs to be weighed against the pros and cons of several other considerations, such as comorbidity, case-mix, cost or social, religious or emotional considerations.

## Prostate cancer detection nomograms

Several authors developed nomograms for prediction of prostate cancer on needle biopsy. One is limited to men with serum PSA values less than 4 ng/ml.[21] This restriction precludes inclusion of many men with PSA values in excess of 4, in whom a biopsy may not always be indicated due to age and/or comorbidity. However, in these men it might be desirable to quantify the probability of finding cancer. Another nomogram relies on ultrasound findings to determine the probability of finding PCa.[22] This requirement also undermines the practical application of this tool. Clinicians decide whether to perform a biopsy well ahead of the ultrasonic assessment of the gland. These findings emphasize the importance of inclusion of readily and routinely available predictor variables, which represent a *sine qua non* of any predictive tool developed for broad use.

To circumvent the limitations of previously developed models, we recently developed a nomogram predicting the probability of prostate cancer on needle biopsy in men undergoing an initial biopsy.[1] This tool only requires the input of variables that are routinely available at the time of a prostatic evaluation, namely age, DRE findings, serum PSA and percent-free PSA. The combined predictive accuracy of this model is 78% in the development cohort and 77% in the external validation cohort, Figure 1. The

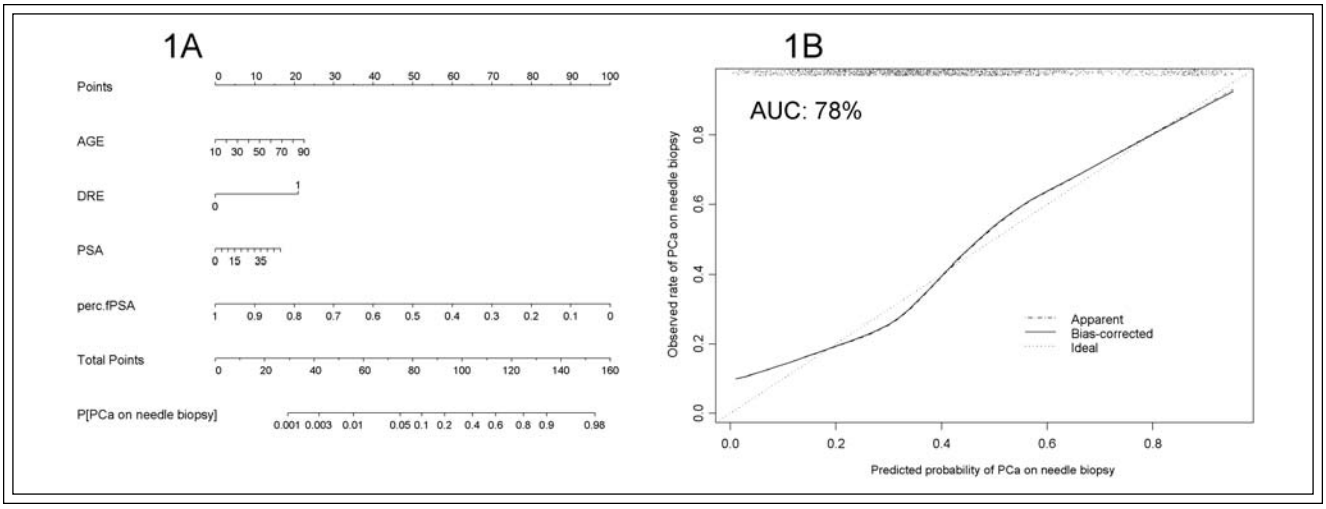


**Figure 1a.** Initial biopsy nomogram based on four variables (age, DRE, PSA and %fPSA).
Instructions for physicians: to obtain nomogram predicted probability of biopsy outcome, locate patient values at each axis. Draw a vertical line to the "Point" axis to determine how many points are attributed for each variable value. Sum the points for all variables. Locate the sum on the "Total Points" line. Draw a vertical line towards the "P(PCa on needle biopsy)" - axis to determine the patient's probability of presence of prostate cancer on initial prostate biopsy.
**Figure 1b.** Calibration plot of the initial biopsy nomogram.
Instructions for readers: perfect prediction would correspond to the 45-degree line. Points estimated below the 45-degree line correspond to nomogram over prediction, whereas points situated above 45-degree line correspond to nomogram under prediction.
DRE: digital rectal examination (1=suspicious, 0=normal)
PSA: prostate specific antigen
perc.fPSA: percent free prostate specific antigen

benefit related to the use of this tool resides in its ability to consider the simultaneous contribution of four variables. Its multivariate performance is appreciably higher than that of PSA (64%), age (52%), DRE (62.9%) or %fPSA (73%) alone. These substantially lower predictive accuracy estimates clearly demonstrate the benefit related to consideration of all four variables.

Besides the combined contribution of the four variables, the nomogram users can ascertain the relative importance of each predictor variable, as all risk factors are graphically depicted in the form of 'risk axes'. For example, assessment of the nomogram axes indicates that the effect of a suspicious DRE, as well as the effect of serum PSA in excess of 50 ng/ml has a limited effect on the probability of diagnosing prostate cancer on needle biopsy. Suspicious DRE contributes to 20 risk points. Similarly, PSA of 50 ng/ml contributes to approximately the same number of risk points. Conversely, %fPSA can contribute to as many as 100 risk points. Thus, the effect of %fPSA is five-fold stronger than that of the other predictors. Such information cannot be derived from look-up tables or from neural networks.

Moreover, assessment of nomogram axes can situate the user with regard to the magnitude of the effect associated with each of predictor. For example, the PSA risk axis indicates a limited magnitude of the effect of PSA. Men with PSA values between 10 ng/ml and 15 ng/ml are given 10 risk points. Suspicious DRE and age of 75 years both contribute approximately 20 risk points. These contributions are modest at best, in the light of %fPSA, where a value of 10% contributes 90 risk points. The above example illustrates how useful the graphical display of nomogram axes can be with regard to familiarizing the clinician with differential contribution of key risk factors.

## Neural networks

The graphical display of risk factors, which allows a clear and user-friendly depiction of the risk variables, distinguishes nomograms from neural networks, where graphical display cannot be provided in paper format. Although, the structure of neural networks can be presented in schematic form, Figure 2, the actual effect of the input variables on the output cannot. This is due to the numerous interactions that are allowed, when data are processed from the input units towards hidden neuron layers and then eventually to one or several output units. Neural networks predicting the outcome of needle biopsy

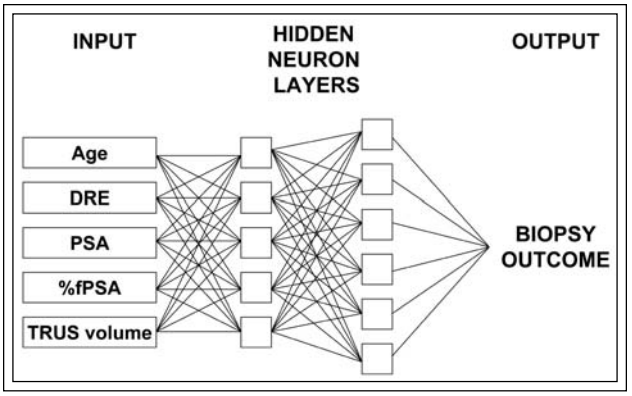


**Figure 2.** Schematic architecture of an artificial neural network to predict prostate cancer on initial biopsy.

have been generally limited to one layer of hidden neurons. Two investigators relied on 'several' layers of hidden neurons.[3,4]

Use of several layers of hidden neurons renders the computational data manipulations highly complex and lacks transparence. Multiple interactions are allowed between input variables at each level. These are weighed to promote the most accurate prediction of the outcome of interest, for example of presence of cancer on needle biopsy. At each hidden neuron, binary outputs are transmitted to the next level of hidden neurons. These resemble multiple outputs within a logistic regression model. Interactions between these outputs, which again can be weighed to further promote accuracy, increase the complexity of the model. The process contributes to highly accurate prediction of the outcome of interest, which in several reported neural network models closely approximates the 85%-95% range. Although accuracy is of key importance, models that underlie predictions need to be tested before their discriminant ability can be taken at face value. Unfortunately, lack of familiarity with biostatistical considerations frequently severely undermines the validity of reported predicted accuracy estimates, which are exaggerated and reported in a biased and methodologically incorrect way. Thus, despite good intentions many investigators report spuriously high ability to predict the outcome of interest.

## Head-to-head comparison of a neural network and a nomogram

To substantiate the claim that neural network predictions are less accurate when they are subjected to strict external validity tests, we compared the ability to predict presence of cancer on biopsy between our

nomogram and a neural network model that was made available by investigators at the Charité Hospital in Berlin, Germany.[5] The nomogram is based on four input variables, namely age, digital rectal examination findings, serum PSA and %fPSA and its maximum predictive accuracy was estimated at 78%.[1] The neural network additionally includes prostate volume as a risk variable and its predictive accuracy has been estimated at 84%. Prostate volume represents an important predictor of PCa risk on needle biopsy in several contemporary analyses.[23-26] Thus, its inclusion should bias the ability of the network to predict more accurately than the nomogram, where this variable is not considered. Moreover, unlike the neural network, the nomogram variables are not allowed to interact with one another, which should further undermine the predictive ability of the nomogram.

Both models were tested on a cohort of 4093 patients subjected to at least 8-core initial biopsy. Despite these *à priori* disadvantages, our results have indicated that the nomogram (70.6%) was 3.6% more accurate than the neural network (67.0%). Both models predicted less accurately than in the original studies, where they were described.[1,5] The decrease in predictive accuracy relative to original data was related to development of both tools on populations subjected to virtually exclusive sextant biopsies, while their head-to-head comparison was performed on a cohort exposed to extended biopsy schemes.

Besides overall model accuracy, we explored the performance characteristics of the nomogram and then of the neural network, as these are instrumental when the decision to adopt one tool versus another. As shown in Figure 3A, the performance characteristics of the nomogram virtually paralleled the ideal 45-degree prediction line. Conversely, the neural network demonstrated important departures from ideal predictions, Figure 3B, which were manifested by severe under estimation throughout the range of predicted probabilities. The most important departures were recorded for predicted probabilities between 20% and 80%.

Taken together, our comparison demonstrated that neural networks do not exceed the ability of logistic regression models to predict the outcome of interest. Moreover, we have shown that the performance characteristics of the nomogram, which consist of a comparison between predicted and observed rate of PCa on needle biopsy were far superior to the neural network. This example of a head-to-head comparison between a nomogram and a neural network shows

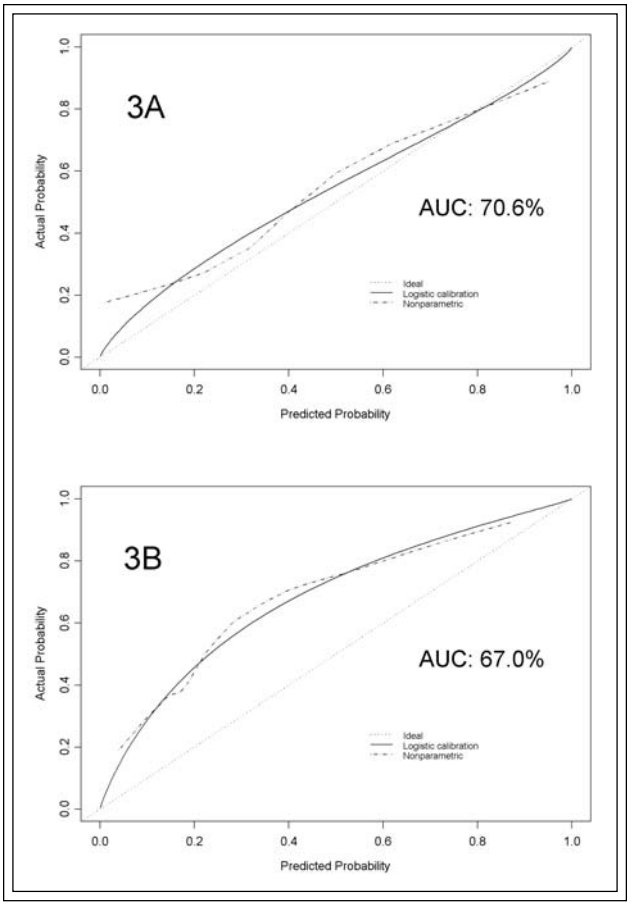


**Figure 3.** Local regression nonparametric smoothing plots which demonstrate performance of external validations of a previously published initial biopsy nomogram[1] (A) and of a previously published artificial neural network[5] (B) to predict initial biopsy outcome.

**Figure 3a.** External validation (n=4093) of the previously published four variables (age, DRE, PSA, %fPSA) sextant nomogram[1] for prediction of prostate cancer in men exposed to initial biopsy, where X-axis represents predicted probability and Y-axis represents observed fraction with evidence of prostate cancer.

**Figure 3b.** External validation (n=4093) of the previously published artificial neural network[5] (age, DRE, PSA, %fPSA, prostate volume) for prediction of prostate cancer in men exposed to initial biopsy where X-axis represents predicted probability and Y-axis represents observed fraction with evidence of prostate cancer.

Perfect predictions correspond to the 45-degree line. Points estimated below the 45-degree line correspond to nomogram over prediction, whereas points situated above 45-degree line correspond to nomogram under prediction.

that nomograms appear to be more accurate and appear to be associated with better performance characteristics. This directly contradicts several urological publications, where the accuracy of neural networks was substantially higher than that of nomograms. Moreover, this example illustrates some of the concerns that experts in prognostics have voiced about the true predictive ability of neural networks.

Nonetheless, our findings and their interpretation are not meant to suggest that the neural network methodology should be abandoned. Instead, they indicate the need for methodologically sound application and critical appraisal of this approach.

## Concerns with current neural network applications

Several important problems have been identified with the methodology of contributions addressing neural network models. The potential for the emergence of these problems has been signaled as early as in 1977, when statistical packages, such as SPSS, SAS and others became widely available for non-specialists.[27] Despite these early warnings, a recent review of existing neural networks for prediction and diagnostic classification in oncology found numerous crucial methodological mistakes in 43 identified articles.[28] These were summarized as (1) biased and/or inefficient estimation, (2) overfitting and fitting of implausible functions, (3) incorrect or missing description of the complexity of the network, (4) use of inadequate statistical competitors or insufficient statistical comparisons, and (5) naive and inappropriate application to survival data.

(1) Mistakes in the estimation of predictive accuracy represent without doubt the most dangerous flaw of many neural networks in oncology. These relate to inappropriate use of data sets to estimate the predictive ability of these models. For example, most reports divide the data sets into learning and validation sets. Such methodology is appropriate, when the accuracy of a regression models is tested. Neural networks behave differently. Therefore, validation sets demonstrate excessively optimistic predicted accuracy, relative to regression models. The degree of optimism has been estimated at between 9% and 13%.[28] Thus, a model may be reported to predict accurately 90% of the time, while in reality only 80% of predictions are correct.

Appropriate assessment of predictive accuracy requires the use of test sets. These can be derived from the original cohort. However, such approach results in fewer observations that can be used for learning and validation. Alternatively, cross-validation techniques can be used, where the test set is generated from a randomly drawn proportion of the population. Another test set can then be randomly identified and the process may be repeated several times. Each time the predictive accuracy of the neural network is determined. Once all repetitions have been completed, an average predictive accuracy is determined. This method is more sophisticated than simple splitting of the dataset between learning, validation and test sets. It allows testing of the unbiasedness of the model on substantially larger test sets, relative to when the cohort is split into three subsets. The most efficient validation may be provided by computer-intensive resampling technique called bootstrapping.[18] This methodology replicates the process of test set generation from an underlying validation set by drawing sample sizes with replacement from the original validation dataset. Each resample is of the same size as the original validation set. Use of resampling maximizes the efficiency of predictive accuracy testing. Thus, instead of dividing the population between three subsets, only the learning and validation sets are required. Use of cross validation or bootstrapping techniques may allow fewer instances of overfitting, where neural networks learning sets rely on few dozens observations and numerous input nodes. It is of note, that regression models do not require a test set. Instead, their validation may be achieved either using the split sample or cross-validation methodologies. Finally, resampling with replacement represents a frequently used and efficient validation approach of regression models.

(2) Overfitting may undermine the validity of neural networks, which have the ability to closely reflect the underlying data. For example neural networks, which are based on few observations but numerous hidden units, have a tendency to result in implausible functions to describe the relation between the input nodes and the output node. Such models may be associated with spuriously high accuracy, which may be difficult to confirm in a test set. Despite great ability to replicate the relations between input and output nodes, the neural networks do require between 5 and 10 observations for each parameter to be estimated. Thus, readers are cautioned about taking at face value the predictive ability of neural networks that bypass that key consideration.

(3) Incorrect or incomplete description of the neural network represents a common limitation in the ability of the reader to independently assess the properties of the network at hand as well as these

of the learning, validation and testing steps. For example, it is not uncommon that multiple hidden layers are mentioned, without specifying how many.

(4) Excessively optimistic performance of neural networks may be due to comparisons with inappropriate, insufficient or inadequate statistical competitors. For example, neural networks are frequently compared to logistic regression models. Although we have demonstrated that logistic regression models can favorably out-compete neural networks, this is not invariably the case. The advantage of neural networks resides in their complexity relative to straightforward regression models, which rely on linear relations between predictors and the outcome of interest. In order to provide comparable conditions, regression models should be fitted with multiple interaction terms and with cubic as well as quadratic predictor terms. Such methodology would results in comparable ability of the predictors to interact with one another in a non-linear fashion, as in neural networks.

(5) The above methodological problems are compounded by inappropriate applications of neural networks. For example, the statistical assumptions governing neural networks generally do not allow the use of censored data. Thus, neural networks are not amenable to modeling of survival data. Many investigators have attempted to circumvent this methodological limitation by either ignoring censored cases, or omitting censored cases, or imputing censored cases, or finally by using time to event data as an additional input. All these approaches are methodologically flawed and are known to results in biased estimate of the outcome of interest.

Finally, neural networks have been popularized in the medical literature by over inflated praises, such as 'ability tom learn...makes them formidable tools in the fight against cancer,[29] and 'neural computation may be as beneficial to medicine and urology in the twenty-first century as molecular biology has been in the twentieth'.[30] Despite these praises, neural networks have made little difference in the diagnosis or management of localized PCa, despite their introduction in the early 1990s.[31] Besides severely limited availability, practical considerations related to the misuses of neural network methodology have without doubt contributed to the observed marginal use of these tools in clinical practice.

## Conclusion

Prediction of several PCa related outcomes can be achieved with nomograms, look-up tables or neural networks. While look-up tables represent a simplification of logistic regression, nomograms and neural networks represent two distinct methodological approaches towards prediction of clinical outcomes. Nomograms offer several advantages. They allow users to understand the underlying effect of risk factors on the outcome of interest. Their predictive accuracy and performance characteristics can be easily tested and graphically displayed. Finally, their accuracy and performance characteristics are at least as good as these of neural networks. These properties have resulted in the use of nomograms in clinical practice across several continents.[32,33] Conversely, neural networks are less popular. Despite numerous methodological flaws in existing neural network models, this approach represents a valid alternative to nomograms, as long as its methodology is used with equal scrutiny to that employed in nomogram applications.[34] □

References


1. Karakiewicz PI, Benayoun S, Kattan MW, Perrotte P, Valiquette L, Scardino PT, Cagiannos I, Heinzer H, Tanguay S, Aprikian AG, Huland H, Graefen M. 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(6):1930-1934.
2. Svatek R, Karakiewicz PI, Shulman M, Karam J, Perrotte P, Benaim E. Pre-Treatment Nomogram for Disease-Specific Survival of Patients with Chemotherapy-Naive Androgen Independent Prostate Cancer. *Eur Urol* 2006;Epub ahead of print.
3. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. *J Clin Oncol* 2002;20(4):921-929.
4. Remzi M, Anagnostou T, Ravery V, Zlotta A, Stephan C, Marberger M, Djavan B. An artificial neural network to predict the outcome of repeat prostate biopsies. *Urology* 2003;62(3):456-460.
5. Stephan C, Cammann H, Semjonow A, Diamandis EP, Wymenga LF, Lein M, Sinha P, Loening SA, Jung K. Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies. *Clin Chem* 2002;48(8):1279-1287.
6. Matsui Y, Utsunomiya N, Ichioka K, Ueda N, Yoshimura K, Terai A, Arai Y. The use of artificial neural network analysis to improve the predictive accuracy of prostate biopsy in the Japanese population. *Jpn J Clin Oncol* 2004;34(10):602-607.
7. Babaian RJ, Fritsche H, Ayala A, Bhadkamkar V, Johnston DA, Naccarato W, Zhang Z. Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology* 2000;56(6):1000-1006.

8. Porter CR, Gamito EJ, Crawford ED, Bartsch G, Presti JC Jr, Tewari A, O'Donnell C. Model to predict prostate biopsy outcome in large screening population with independent validation in referral setting. *Urology* 2005;65(5):937-941.
9. Poulakis V, Witzsch U, de Vries R, Emmerlich V, Meves M, Altmannsberger HM, Becht E. Preoperative neural network using combined magnetic resonance imaging variables, prostate specific antigen, and Gleason score to predict prostate cancer recurrence after radical prostatectomy. *Eur Urol* 2004;46(5):571-578.
10. Poulakis V, Witzsch U, De Vries R, Emmerlich V, Meves M, Altmannsberger HM, Becht E. Preoperative neural network using combined magnetic resonance imaging variables, prostate specific antigen and Gleason score to predict prostate cancer stage. *J Urol* 2004;172(4 Pt 1):1306-1310.
11. Han M, Snow PB, Epstein JI, Chan TY, Jones KA, Walsh PC, Partin AW. A neural network predicts progression for men with gleason score 3+4 versus 4+3 tumors after radical prostatectomy. *Urology* 2000;56(6):994-999.
12. Partin AW, Kattan MW, Subong EN, Walsh PC, Wojno KJ, Oesterling JE, Scardino PT, Pearson JD. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA* 1997;277(18):1445-1451.
13. Finne P, Auvinen A, Aro J, Juusela H, Maattanen L, Rannikko S, Hakama M, Tammela TL, Stenman UH. Estimation of prostate cancer risk on the basis of total and free prostate-specific antigen, prostate volume and digital rectal examination. *Eur Urol* 2002;41(6):619-626.
14. Steuber T, Graefen M, Haese A, Erbersdobler A, Chun FK, Schlomm T, Perrotte P, Huland H, Karakiewicz PI. Validation of a nomogram for prediction of side specific extracapsular extension at radical prostatectomy. *J Urol* 2006;175(3 Pt 1):939-944.
15. Specht MC, Kattan MW, Gonen M, Fey J, Van Zee KJ. Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram. *Ann Surg Oncol* 2005;12(8):654-659.
16. Chun FK, Steuber T, Erbersdobler A, Currlin E, Walz J, Schlomm T, Haese A, Heinzer H, McCormack M, Huland H, Graefen M, Karakiewicz PI. Development and Internal Validation of a Nomogram Predicting the Probability of Prostate Cancer Gleason Sum Upgrading Between Biopsy and Radical Prostatectomy Pathology. *Eur Urol* 2005;Epub ahead of print.
17. Bradley E, Tibshirani RJ. Monographs on statistics and applied probability: An introduction to the bootstrap. Champman and Hall/CRC, p.275, 1993.
18. 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(8):774-781.
19. Katz EM, Kattan MW. How to judge a tumor marker. *Nat Clin Pract Oncol* 2005;2:482-483.
20. Kattan MW. Statistical prediction models, artificial neural networks, and the sophism "I am a patient, not a statistic". *J Clin Oncol* 2002;20(4):885-887.
21. Eastham JA, May R, Robertson JL, Sartor O, Kattan MW. Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology* 1999;54:709.
22. Garzotto M, Hudson RG, Peters L, Hsieh YC, Barrera E, Mori M et al. Predictive modeling for the presence of prostate carcinoma using clinical, laboratory, and ultrasound parameters in patients with prostate specific antigen levels £ 10 ng/mL. *Cancer* 2003;98:1417.
23. Karakiewicz PI, Bazinet M, Aprikian AG, Trudel C, Aronson S, Nachabe M, Peloquint F, Dessureault J, Goyal MS, Begin LR, Elhilali MM. Outcome of sextant biopsy according to gland volume. *Urology* 1997;49(1):55-59.
24. Uzzo RG, Wei JT, Waldbaum RS, Perlmutter AP, Byrne JC, Vaughan ED Jr. The influence of prostate size on cancer detection. *Urology* 1995;46(6):831-836.
25. Eskicorapci SY, Guliyev F, Akdogan B, Dogan HS, Ergen A, Ozen H. Individualization of the biopsy protocol according to the prostate gland volume for prostate cancer detection. *J Urol* 2005;173(5):1536-1540.
26. Freedland SJ, Isaacs WB, Platz EA, Terris MK, Aronson WJ, Amling CL, Presti JC Jr, Kane CJ. Prostate size and risk of high-grade, advanced prostate cancer and biochemical progression after radical prostatectomy: a search database study. *J Clin Oncol* 2005;23(30):7546-7554.
27. Lachenbruch PA. Some misuses of discriminant analysis. *Methods Inf Med* 1977;16(4):255-258.
28. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* 2000;19(4):541-561.
29. Burke HB. Artificial neural networks for cancer research: outcome prediction. *Semin Surg Oncol* 1994;10(1):73-79.
30. Niederberger CS. This month in Investigative Urology. Commentary on the use of neural networks in clinical urology. *J Urol* 1995;153(5):1362.
31. Snow PB, Smith DS, Catalona WJ. Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study. *J Urol* 1994;152(5 Pt 2):1923-1926.
32. Graefen M, Karakiewicz PI, Cagiannos I, Quinn DI, Henshall SM, Grygiel JJ, Sutherland RL, Stricker PD, Klein E, Kupelian P, Skinner DG, Lieskovsky G, Bochner B, Huland H, Hammerer PG, Haese A, Erbersdobler A, Eastham JA, de Kernion J, Cangiano T, Schroder FH, Wildhagen MF, van der Kwast TH, Scardino PT, Kattan MW. International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2002;20(15):3206-3212.
33. Graefen M, Karakiewicz PI, Cagiannos I, Klein E, Kupelian PA, Quinn DI, Henshall SM, Grygiel JJ, Sutherland RL, Stricker PD, de Kernion J, Cangiano T, Schroder FH, Wildhagen MF, Scardino PT, Kattan MW. Validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer. *J Clin Oncol* 2002;20(4):951-956.
34. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. *Semin Urol Oncol* 2002;20(2):89-95.