

Conference on Prognostic Factors and Staging in Cancer
Management: Contributions of Artificial Neural Networks
and Other Statistical Methods

Supplement to Cancer

Artificial Neural Networks for Prostate Carcinoma Risk Assessment

An Overview

James E. Montie, M.D.
John T. Wei, M.D.

Section of Urology, The University of Michigan, Ann Arbor, Michigan.

Reprinted from *Cancer* 2000;88:2655–60.

Presented at the Conference on Prognostic Factors and Staging in Cancer Management: Contributions of Artificial Neural Networks and Other Statistical Methods, Arlington, Virginia, September 27–28, 1999.

Supported by grant funding from the National Cancer Institute of the National Institutes of Health.

Dr. Montie is an investigator for Zeneca, Inc., and Centeon, Inc.

Address for reprints: James E. Montie, M.D., Section of Urology, The University of Michigan, 2916 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0330.

Received December 30, 1999; accepted March 15, 2000.

This conference on artificial neural networks (ANNs) provided an opportunity to explore current and future strategies for the prediction of outcomes for patients with prostate carcinoma. For example, it is necessary to improve the selection process for prostate biopsy for a patient with clinical factors that raise the suspicion of prostate carcinoma and provide greater precision about possible outcomes after treatment. This paper provides an overview of the role that ANNs may play in the larger picture of prostate carcinoma prognostication. It also presents personal insights garnered from over a decade of clinical experience with staging questions for patients with prostate carcinoma.¹

Historic Perspective

Care for a cancer patient must address three fundamental issues: 1) the diagnosis, 2) the extent of disease, and 3) the appropriate treatment. If a cancer poses a threat to the patient, and if treatment is effective in some but not all patients, then the assessment of the extent of the disease is a prerequisite for the logical pursuit of treatment decisions. In general, of all prognostic factors, a description of the extent of a disease is the most powerful predictor of patient survival.² The extent of a disease has been described historically by stages, with increasing tumor burden or spread designated by a higher stage and associated with a diminished survival. The tumor, lymph node, metastasis (TNM) classification was developed in the 1950s, and it has been refined periodically.^{3–5} Within each reference of T, N, or M, there are categories based on tumor size or extent, number of regional lymph nodes containing metastases, and distant metastases, referred to as “variables.”⁶ The assignment of a T, N, or M variable is dependent on the quality of data obtained from staging studies, such as a digital rectal examination (DRE) or a radionuclide bone scan used for prostate carcinoma. A patient can then be described by using a combination of all three variables. This places the patient into an individual “bin” or category that is mutually exclusive and exhaustive. For example, in prostate carcinoma, there are 16 bins based on the possible permutations of the TNM variables ($4 \times 2 \times 2 = 16$ variables). For ease of use, these bins often are grouped together into “stage groups” based on similar risk for a survival (Fig. 1).

An advantage of the bin model is that it allows for a relatively

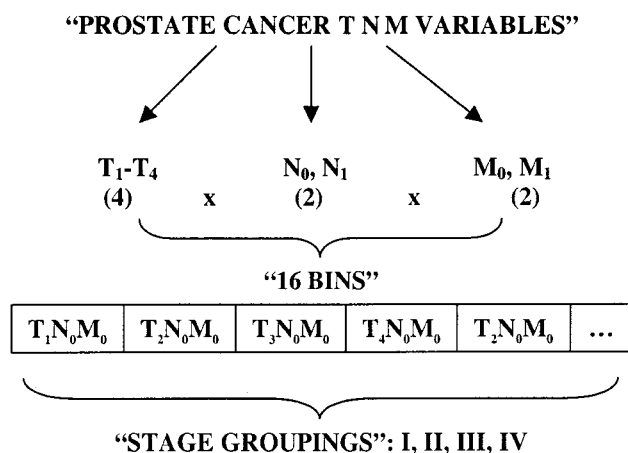


FIGURE 1. TNM bin model for prostate carcinoma.

precise description of the extent of a patient’s cancer, but a major disadvantage of the bin model arises from the exponential increase in the number of bins with the inclusion of additional variables. For patients with prostate carcinoma, adding just one variable, such as the histologic tumor grade (well, moderate, or poorly differentiated), results in the creation of 48 bins ($4 \times 2 \times 2 \times 3 = 48$ variables). One proposed strategy to clarify this complexity is the further expansion of stage grouping. However, the situation then becomes so complicated that it becomes impractical. In addition, the current stage grouping using only the TNM anatomic extent of disease yields four stages, but most of the recently identified prostate carcinomas fall into clinically localized disease or Stage II. Dr. Alan Partin has presented data that 97% of Dr. Patrick Walsh’s patients at Johns Hopkins Hospital are categorized as Stage II.

One disadvantage of any staging classification is that an individual’s stage may change if new data become available. For example, improvements in imaging technology, more precise histologic evaluation, and better identification of micrometastases may change an individual’s stage classification. This artificial “stage migration” of patients, for example, from a T2N0M0 to a T3N0M0 or a T2N1M0, may well change the survival prediction on which staging rules were employed and, thus, undermine the reliability of the staging system.^{2,6} Stage migration certainly is present in prostate carcinoma using transrectal ultrasound (TRUS) needle biopsy of seminal vesicles, step-sectioning reconstruction of the entire prostate to identify microscopic capsular penetration, or identification of micrometastases in the lymph nodes.

Such limitations in the staging classification do not imply a lack of value in the description of the anatomic extent of disease. Additional research may

contribute to survival prediction or treatment success. Practitioners caring for prostate carcinoma patients are fortunate to have two other strong prognostic factors: prostate specific antigen (PSA) and Gleason score. In most studies evaluating prognostic factors, pretreatment PSA level and Gleason score have been shown to be strongly associated with disease outcome. Recent studies focus on how to incorporate additional information from these factors into a workable system for clinical use. An important observation from the 1980s is pertinent when one devises a staging or prognostication system; the National Cancer Institute recognized the impetus to evolve to a TNM staging system from the classic Whitmore system by establishing a “blue-ribbon” panel in 1986 to incorporate the best of both systems. Unfortunately, the classification was complex, and deliberations did not incorporate the responsible agencies worldwide for staging classifications.⁷ Without the support of vital organizations, such as the International Union Against Cancer (UICC) and the American Joint Committee on Cancer (AJCC), the system was not successful. In 1992, a broadly supported classification became available that modified the TNM system to more closely reflect the general categories outlined by Catalona and Whitmore.⁸ This classification defined Stage A as clinically latent prostate carcinoma (T1); Stage B as clinically manifest, early prostate carcinoma (T2); Stage C as clinically manifest, locally advanced prostate carcinoma (T3); and Stage D as clinically manifest, advanced prostate carcinoma with evidence of distant metastases (N+, M+).^{9,10} The TNM system, as defined in 1992 and modified in 1997, is now used broadly worldwide for both clinical care and research studies.¹¹ To be accepted universally by the medical community, a staging or prognostic system must be both practical and supported by existing governing organizations.

Recent Investigations

Patients, clinicians, and researchers need methods with which to evaluate how well an intervention works. Modeling is a process by which statisticians attempt to describe and summarize the clinical milieu by using mathematical equations. The TNM staging classification, in fact, is a simple prognostic model. More complicated prognostic models, such as the Partin, D’Amico, or Kattan nomograms described below, are an improvement on the clinical staging model.¹²⁻¹⁴ In general, there are two purposes for constructing models: to identify factors that are associated with an outcome of interest and to predict a specific outcome, such as survival.

No less than 13 prognostic classic statistical (linear or log-linear) models providing probabilities for a

specific prostate carcinoma treatment outcome have been described as of 1998.¹⁵ The Partin nomograms, which were published in 1993 and were expanded with a larger data set in 1997, apply to patients with clinically localized disease who are undergoing radical prostatectomy and include clinical T stage, PSA level, and Gleason score as variables.¹² Nomograms were developed on the basis of multivariate regression analyses, with the endpoint of pathologic disease stage at the time of radical prostatectomy. In 1987, Pasanski et al. developed a prognostic index for prostate carcinoma patients prior to external radiation therapy.¹⁶ D'Amico and colleagues have been prolific in refining models: They developed a risk score that calculated cancer volume in 1998 and a nomogram in 1999 for use in patients undergoing either radical prostatectomy or external radiation therapy.^{13,17-19} A nomogram developed by Kattan and colleagues in 1998 utilizes a scoring system to provide a probability figure for PSA recurrence after radical prostatectomy.¹⁴ More recently, Vollmer and Montana evaluated models for patients undergoing external radiation therapy using clinical stage, PSA level, and Gleason score as variables.²⁰ The natural log (PSA) for PSA values, T1 and T2 versus T3 and T4, and Gleason score < 7 versus Gleason score ≥ 7 were optimal cut points for assessing the best performance in the models. Although the sensitivity, specificity, and accuracy appeared quite good, it was noteworthy that only 18% of the variance was explained by these variables. The authors concluded that additional prognostic factors are needed for better prediction. Thus, recent nomograms are helpful, and many are available on the Internet; however, these nomograms are less than ideal, because PSA and Gleason score are categorized at different cut points rather than as continuous variables, and the addition of other factors would require complete restructuring of the nomogram.

ANNs are another means with which to model outcome data.²¹ Briefly, ANNs are a class of nonlinear mathematical models that are characterized by a complex structure of interconnected computational elements, the neurons. These computational elements then aggregate a series of inputs (prognostic factors) using a summation operation and produce an output, such as the probability of having a 10-year survival. Inputs to each neuron are multiplied by a weight factor that reflects the excitatory or inhibitory strength of the connection from the input source to the neuron. The sum of the weighted inputs plus a bias term then goes through an activation function that behaves like a "switch" to determine whether or not the neuron will "fire" and, thus, will send out an output signal. The bias term may be viewed as the threshold that the

weighted sum of inputs must exceed before the neuron sends out an output signal. The information-processing capacity of an ANN is a function of the type and quantity of neurons in a given network and by the arrangement of interconnections between neurons. The "knowledge" or the "processing capability" of an ANN is defined by the actual values of the connection weights between neurons in the network. Such "knowledge" of an ANN is acquired through a "learning phase," during which examples of data to be processed repeatedly are fed through the ANN model, and the connection weights in the ANN are adjusted adaptively for the ANN as a whole to satisfy some predetermined performance goals.

Reports on "staging" ANNs have been published, and several more have been presented at conferences.^{22,23} Whereas classic statistical models are capable of identifying significant independent prognostic factors and also can predict outcomes fairly well, ANNs have been limited primarily to predicting outcomes. Which is the preferred method? The classic statistical methods have an advantage, in that they are "transparent," whereas ANNs essentially are a "black box." Nevertheless, one can argue that how these models achieve their goal is less important than how well they do it. After all, what is important to the clinician and the patient is which model best predicts the outcome of interest. A receiver operating characteristics curve is an appropriate test with which to compare how well the various models perform as risk-assessment tools. The data presented at this conference suggest that there is a modest advantage with some ANNs.

For all statistical modeling processes, having more data available during the development phase will strengthen the precision and reliability of its outcome estimate. Fortunately, in urologic oncology, there is an excellent example of the value of a large data set. For years, there has been controversy about the prediction of disease recurrence after chemotherapy for patients with metastatic testis carcinoma to define low, intermediate, and high risk patients. The analysis of an international data set incorporating more than 5000 patients from leading institutions essentially resolved the arguments and presented a system that is accepted worldwide.²⁴ This process demonstrates the tremendous utility of a large, diverse data set from many key investigators with the support of international governing bodies, thus ensuring the acceptance of the final outcomes.

Problems Unique to Prostate Carcinoma

More specific problems with prostate carcinoma become apparent as we evaluate clinical situations. Sev-

eral general problems with prognostic factors discussed elsewhere in the conference relate to 1) heterogeneity and sampling issues, 2) interobserver variability, 3) sample size and subgroup analyses, 4) specimen handling, and 5) endpoints and surrogate biomarkers. For example, information from core biopsies of the prostate for the diagnosis of patients with prostate carcinoma provides potential prognostic information. Variability in the data includes the number of cores sampled (1 vs. 6 vs. 14), whether the biopsy is directed at an "abnormal" area on DRE or TRUS, the use of systematic biopsies, the number of positive cores, the amount of the cancer present measured by the percentage of tissue involved or amount (in millimeters) involved, the use of inked margins, and cost issues relating to processing each core separately.²⁵⁻³² Such concerns are by no means trivial issues and can influence widespread acceptance of a system that incorporates the data provided by biopsies. Another recognized problem in prostate carcinoma is interobserver variability in the grading. Expert pathologists tend to grade higher than community-based physicians with less experience.³³

Tissue sampling for special studies is an additional problem. Most molecular markers require frozen tissue harvested primarily from larger tumors from biopsies or radical prostatectomy specimens, thus introducing bias in a study. Microarrays for gene analysis also are subject to such bias.

Additional problems include the use of subgroups in prostate carcinoma. Endorectal magnetic resonance imaging may prove to be useful primarily only in patients who have larger tumors and, thus, are at greater risk for extracapsular extension.³⁴ In the University of Michigan data base, bladder neck-sparing radical prostatectomy was associated with a greater risk for a positive margin in patients with intermediate-sized tumors but not in patients with small or large tumors.³⁵

Finally, the radical prostatectomy specimen poses several problems. Variability in the time between removal of the prostate and retrieval of the tumor for molecular biology studies may lead to variability in results. The surgical technique may influence margin status, which is recognized as an important prognostic factor.³⁶ The processing in pathology of the prostate varies from routine sampling to a more extensive sampling directed at an abnormal area to a step-sectioning reconstruction to a wholemount reconstruction. Each of these methods provides progressively more precise information on the size and extent of the tumor. In addition, quantification of the amount of extraprostatic extension or a positive margin is not standardized.³⁷ The widespread use of neoadjuvant

androgen-deprivation therapy is of unproven long term benefit, but it drastically alters the histologic findings in prostatectomy specimens and essentially eliminates the possibility of further studies with most molecular biomarkers.³⁸

Approximately 30 prognostic factors have been proposed for prostate carcinoma, a figure probably less than half of those identified for breast carcinoma, reflecting more intense research and greater national funding for breast carcinoma. Expansion of prostate carcinoma research may intensify the similar dilemma currently noted in breast carcinoma, i.e., how to use the new data.

What Is Needed Presently?

A risk-assessment tool is needed that produces a probability estimate for a specific outcome tailored for a particular patient and therapy. The prognostic factors included in such a tool need to explain most of the variability in outcome.³⁹ This information can then be discussed with the patient in the context of patient preference to accept or avoid specific side effects of treatment. All current therapies for patients with prostate carcinoma have side effects, and the patient must always occupy the center of the discussion about their relative importance. Randomized clinical trials in prostate carcinoma patients have proven remarkably difficult to complete based on a variety of factors, including preconceived biases by physicians, difficulty in patient randomization with widely disparate treatments, and inadequate educational tools about the studies for both the patient and the physician. However, the absence of optimal, randomized trials should not preclude the use of the large amount of data available on patients who have been treated around the country, reflecting the real world. Matching of such patients by prognostic factors in large data bases may be tremendously valuable. The UICC and the AJCC always have been interested in predicting survival. This is the primary purpose for clinical and pathologic staging. In an era of rapidly growing applications for biomarkers, new factors to predict prostate carcinoma outcomes frequently are identified; simple clinical or pathologic staging will no longer suffice, nor will simple models involving two or three variables. Both classical statistical models and ANNs are amenable to computer technology. Both forms of models can be developed so that entering certain variables will allow the computer to generate a probability for a specific outcome. In our opinion, the probability figure should be a continuous variable, but the UICC and the AJCC can develop categories based on increasing risk for failure (or success). Such a probability figure, with or without "risk" grouping, may be more helpful

to the clinician who often is confronted by an increasing number of so-called prognostic factors, not the least of which is staging. This probability or grouping can then be applied to the various outcomes, including survival, complications, disease recurrence, etc.

The UICC and the AJCC must change their role from the standardization of staging to the standardization of risk assessment. Soon, a plethora of risk-assessment packages for various cancers will be developed through industry and will become available on the Internet. The UICC and the AJCC should lead the way by facilitating the development of large, international data sets, the promulgation of standards, and the dissemination of information.

Microarrays for gene expression and tissue microarrays for biomarkers will provide an enormous amount of raw data for an individual cancer and present a tremendous challenge to our ability to process the information. A true "prediction" with absolute certainty about an outcome currently is beyond our grasp; nevertheless, improved statistical models, such as ANNs, may allow for greater flexibility and precision and may bring us closer to the ultimate goal.

REFERENCES

1. Montie JE. 1992 staging system for prostate cancer. *Semin Urol* 1993;11:10-3.
2. Burke HB, Henson DE. The American Joint Committee on Cancer. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer* 1993;72:3131-5.
3. Sobin LH, Hermanek P, Hutter RV. TNM classification of malignant tumors. A comparison between the new (1987) and the old editions. *Cancer* 1988;61:2310-4.
4. Sobin LH. Cancer staging: future directions for the TNM classification. *Semin Surg Oncol* 1992;8:107-10.
5. Gospodarowicz M, Benedet L, Hutter RV, Fleming I, Henson DE, Sobin LH. History and international developments in cancer staging. *Cancer Prevent Control* 1998;2:262-8.
6. Fielding LP, Henson DE. Multiple prognostic factors and outcome analysis in patients with cancer. Communication from the American Joint Committee on Cancer. *Cancer* 1993;71:2426-9.
7. Carter HB, Partin AW. Diagnosis and staging of prostate cancer. In: Walsh PC, Retik AB, Vaughan ED, Wein AJ, editors. *Campbell's urology*, 7th ed. Philadelphia: WB Saunders Co., 1998:2526.
8. Catalona WJ, Whitmore WF Jr. New staging systems for prostate cancer. *J Urol* 1989;142:1302-4.
9. Montie JE. 1992 staging system for prostate cancer. *Semin Urol* 1993;11:10-3.
10. Beahrs OH, Henson DE, Hutter RVP, Kennedy BJ. Manual for staging of cancer, 4th ed. Philadelphia: JB Lippincott, 1992.
11. Fleming ID, Cooper JS, Henson DE, Hutter RVP, Kennedy BJ, Murphy GP, et al. AJCC cancer staging handbook, 5th ed. Philadelphia: Lippincott-Raven Publishers, 1997.
12. Partin AW, Kattan MW, Subong EN, Walsh PC, Wojno KJ, Oesterling JE, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA* 1997;277:1445-51.
13. D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen MH, Tomaszewski JE, 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.
14. 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.
15. Vollmer RT, Keetch DW, Humphrey PA. Predicting the pathology results of radical prostatectomy from preoperative information: a validation study. *Cancer* 1998;83:1567-80.
16. Pisansky TM, Kahn MJ, Rasp GM, Cha SS, Haddock MG, Bostwick DG. A multiple prognostic index predictive of disease outcome after irradiation for clinically localized prostate carcinoma. *Cancer* 1997;79:337-44.
17. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Kaplan I, Beard CJ, et al. Calculated prostate cancer volume greater than 4.0 cm³ identifies patients with localized prostate cancer who have a poor prognosis following radical prostatectomy or external-beam radiation therapy. *J Clin Oncol* 1998;16:3094-100.
18. D'Amico AV, Desjardin A, Chen MH, Paik S, Schultz D, Renshaw AA, et al. Analyzing outcome-based staging for clinically localized adenocarcinoma of the prostate. *Cancer* 1998;83:2172-80.
19. D'Amico AV, Desjardin A, Chung A, Chen MH, Schultz D, Whittington R, et al. Assessment of outcome prediction models for patients with localized prostate carcinoma managed with radical prostatectomy or external beam radiation therapy. *Cancer* 1998;82:1887-96.
20. Vollmer RT, Montana GS. Predicting tumor failure in prostate carcinoma after definitive radiation therapy: limitations of models based on prostate-specific antigen, clinical stage, and Gleason score. *Clin Cancer Res* 1999;5:2476-84.
21. Wei JT, Zhang Z, Barnhill SD, Madyastha KR, Zhang H, Oesterling JE. Understanding artificial neural networks and exploring their potential applications for the practicing urologist. *Urology* 1998;52:161-72.
22. Tewari A, Narayan P. Novel staging tool for localized prostate cancer: a pilot study using genetic adaptive neural networks. *J Urol* 1998;160:430-6.
23. Babaian RJ, Fritsche HA, Zhang Z, Zhang KH, Madyastha KR, Barnhill SD. Evaluation of prostaSure index in the detection of prostate cancer: a preliminary report. *Urology* 1998;51:132-6.
24. International Germ Cell Cancer Collaborative Group. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. *J Clin Oncol* 1997;15:594-603.
25. Huland H, Hubner D, Henke RP. Systematic biopsies and digital rectal examination to identify the nerve-sparing side for radical prostatectomy without risk of positive margin in patients with clinical stage T2, N0 prostatic carcinoma. *Urology* 1994;44:211-4.
26. Wills ML, Sauvageot J, Partin AW, Gurganus R, Epstein JI. Ability of sextant biopsies to predict radical prostatectomy stage. *Urology* 1998;51:759-64.
27. Borirakchanyavat S, Bhargava V, Shinohara K, Toke A, Carroll PR, Presti JC, et al. Systematic sextant biopsies in the prediction of extracapsular extension at radical prostatectomy. *Urology* 1997;50:373-8.

28. Eskew LA, Bare RL, McCullough DL. Systematic 5 region prostate biopsy is superior to sextant method for diagnosing carcinoma of the prostate. *J Urol* 1997;157:199–202.
29. Egevad L, Norberg M, Mattson S, Norlen BJ, Busch C. Estimation of prostate cancer volume by multiple core biopsies before radical prostatectomy. *Urology* 1998;52:653–8.
30. Dietrick DD, McNeal JE, Stamey TA. Core cancer length in ultrasound-guided systematic sextant biopsies: a preoperative evaluation of prostate cancer volume. *Urology* 1995;45:987–92.
31. Conrad S, Graefen M, Pichlmeier U, Henke RP, Hammerer PG, Huland H. Systematic sextant biopsies improve preoperative prediction of pelvic lymph node metastases in patients with clinically localized prostatic carcinoma. *J Urol* 1998;159:2023–9.
32. Vashi AR, Wojno KJ, Gillespie B, Oesterling JE. A model for the number of cores per prostate biopsy based on patient age and prostate gland volume. *J Urol* 1998;159:920–4.
33. Weinstein MH, Greenspan DL, Epstein JI. Diagnoses rendered on prostate needle biopsy in community hospitals. *Prostate* 1998;35:50–5.
34. D'Amico AV, Schnall M, Whittington R, Malkowicz SB, Schultz D, Tomaszewski JE, et al. Endorectal coil magnetic resonance imaging identifies locally advanced prostate cancer in select patients with clinically localized disease. *Urology* 1998;51:449–54.
35. Marcovich R, Wojno KJ, Wei JT, Oesterling JE, Montie JE, Sanda MG. Bladder neck-sparing technique adversely affects anti-tumor efficacy of radical prostatectomy in pathologic T3a prostate cancer. *J Urol* 1999;161:386.
36. Blute ML, Bostwick DG, Seay TM, Martin SK, Slezak JM, Bergstralh EJ, et al. Pathologic classification of prostate carcinoma: the impact of margin status. *Cancer* 1998;82:902–8.
37. Sakr WA, Wheeler TM, Blute M, Bodo M, Calle-Rodrigue R, Henson DE, et al. Staging and reporting of prostate cancer—sampling of the radical prostatectomy specimen. *Cancer* 1996;78:366–8.
38. Reuter VE. Pathological changes in benign and malignant prostatic tissue following androgen deprivation therapy. *Urology* 1997;49:16–22.
39. Krongrad A, Lai H, Lai S. Variation in prostate cancer survival explained by significant prognostic factors. *J Urol* 1997;158:1487–90.