The Use of Artificial Intelligence Technology to Predict Lymph Node Spread in Men with Clinically Localized Prostate Carcinoma

E. David Crawford, M.D.[1]
Joseph T. Batuello, M.D., J.D.[2]
Peter Snow, Ph.D.[3]
Eduard J. Gamito, B.S.[2]
David G. McLeod, M.D., J.D.[4]
Alan W. Partin, M.D., Ph.D.[5]
Nelson Stone, M.D.[6]
James Montie, M.D.[7]
Richard Stock, M.D.[8]
John Lynch, M.D.[9]
Jeff Brandt, Ph.D.[3]

[1] Section of Urologic Oncology, University of Colorado Health Sciences Center, Denver, Colorado.

[2] ANNs in CaP Project, Denver, Colorado.

[3] Xaim Incorporated, Colorado Springs, Colorado.

[4] Walter Reed Army Medical Center, Urology Clinic, Washington, DC.

[5] Johns Hopkins University, Baltimore, Maryland.

[6] Urology and Radiation Oncology, Mt. Sinai School of Medicine, New York, New York.

[7] University of Michigan Medical Center/Urology-Ann Arbor, Michigan.

[8] Department of Radiation Oncology, Mt. Sinai Medical Center, New York, New York.

[9] Department of Surgery, Georgetown University Medical Center, Washington, DC.

Supported by Institute for Clinical Research, Inc. and a Department of Defense (DOD) grant to D.G.M. and a Specialized Program of Research Excellence (SPORE) Grant NCI-CA-58236 to A.W.P.

Address for reprints: E. David Crawford, M.D., Professor of Surgery/Radiation Oncology, Section of Urologic Oncology, University of Colorado Health Sciences Center, 4200 East Ninth Ave., Box C-324, Denver, CO 80262.

Received August 3, 1999; revision received December 7, 1999; accepted December 7, 1999

BACKGROUND. The current study assesses artificial intelligence methods to identify prostate carcinoma patients at low risk for lymph node spread. If patients can be assigned accurately to a low risk group, unnecessary lymph node dissections can be avoided, thereby reducing morbidity and costs.

METHODS. A rule-derivation technology for simple decision-tree analysis was trained and validated using patient data from a large database (4133 patients) to derive low risk cutoff values for Gleason sum and prostate specific antigen (PSA) level. An empiric analysis was used to derive a low risk cutoff value for clinical TNM stage. These cutoff values then were applied to 2 additional, smaller databases (227 and 330 patients, respectively) from separate institutions.

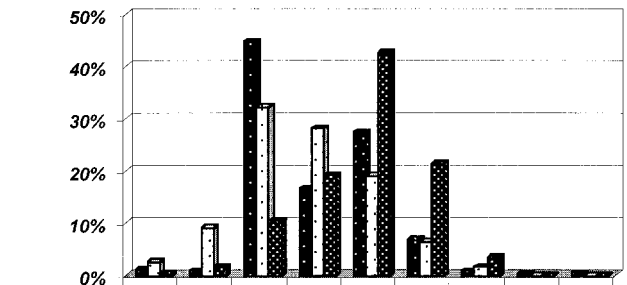
RESULTS. The decision-tree protocol derived cutoff values of ≤ 6 for Gleason sum and ≤ 10.6 ng/mL for PSA. The empiric analysis yielded a clinical TNM stage low risk cutoff value of $\leq$ T2a. When these cutoff values were applied to the larger database, 44% of patients were classified as being at low risk for lymph node metastases (0.8% false-negative rate). When the same cutoff values were applied to the smaller databases, between 11 and 43% of patients were classified as low risk with a false-negative rate of between 0.0 and 0.7%.

CONCLUSIONS. The results of the current study indicate that a population of prostate carcinoma patients at low risk for lymph node metastases can be identified accurately using a simple decision algorithm that considers preoperative PSA, Gleason sum, and clinical TNM stage. The risk of lymph node metastases in these patients is $\leq 1\%$; therefore, pelvic lymph node dissection may be avoided safely. The implications of these findings in surgical and nonsurgical treatment are significant. ***Cancer*** **2000;88:2105–9.** *© 2000 American Cancer Society.*

KEYWORDS: prostate carcinoma, artificial intelligence, decision tree, lymphadenectomy, metastases.

The trend in clinical practice has been toward limiting the invasiveness of diagnostic and therapeutic interventions. Limiting invasive procedures without sacrificing diagnostic accuracy or therapeutic effectiveness is expected to reduce the morbidity and mortality of those procedures, as well as the associated costs.

It is generally reported that the incidence of lymph node metastases found by pelvic lymph node dissection, either as part of a radical prostatectomy or as part of pretreatment staging, is 10% or less.[1–5] Therefore, approximately 90% of these procedures and their associated morbidity and costs may be unnecessary. The goal of this study is to assess the use of artificial intelligence methods to identify those patients with prostate cancer who are at low risk of lymph node spread, by using clinical parameters that currently are obtained in the



	T1a	T1b	T1c	T2a	T2b	T2c	T3a	T3b	T3c
WRAMC	1.2%	0.9%	44.8%	16.7%	27.6%	7.0%	0.9%	0.3%	0.6%
Johns Hopkins	2.6%	9.3%	32.3%	28.2%	19.3%	6.6%	1.7%	0.0%	0.0%
Mt. Sinai	0.4%	1.8%	10.6%	19.4%	42.7%	21.6%	3.5%	0.0%	0.0%

FIGURE 1. Distribution of TNM stage is shown. WRAMC: Walter Reed Army Medical Center.

T1a	Non-palpable, < 5% of tissue with cancer, at TURP low-grade.
T1b	Non-palpable, >5% of tissue with cancer at TURP and/or high-grade.
T1c	Non-palpable, with elevated PSA.
T2a	Palpable, half of one lobe or less.
T2b	Palpable, more than half of one lobe, not both lobes.
T2c	Palpable, involves both lobes.
T3a	Palpable, unilateral capsular penetration.

FIGURE 2. Definition of each clinical stage[9–10] is shown. TURP: transurethral resection of the prostate.

preoperative evaluation. If such a population can be determined with sufficient accuracy, unnecessary lymph node dissections in those patients can be avoided, significantly reducing morbidity and costs.

Other investigators have proposed rules, derived using traditional statistical and empiric methods, for identifying patients in whom lymph node dissections may be safely foregone.[2,6–8] In this study, we applied a rule-derivation technology for simple decision-tree analysis to derive low risk cutoffs for two clinical parameters: preoperative Gleason sum and prostate specific antigen (PSA) level. An empiric analysis was used to determine whether a cutoff for clinical stage would enhance the results obtained by the decision-tree protocol.

METHODS

The data used in this study were extracted from three databases that contained patient information collected from five centers. Informed consent was obtained where appropriate, and the databases used contained only clinical data with no patient identifiers. The first database (the "development database") has been previously described by Partin et al.[9] It contained clinical information on 4133 consecutive patients with clinically localized prostate carcinoma who underwent radical retropubic prostatectomy and staging lymphadenectomy at one of these three centers: Johns Hopkins (3116 patients), Baylor College of Medicine (782 patients), and the University of Michigan School of Medicine (235). Patients were included in this protocol if they had 1) a preoperative prostate specific antigen collected at least 4 weeks after transrectal or digitally guided needle biopsy, transurethral resection of the prostate, or both; 2) evaluation of histologic grade according to the Gleason scoring system; and 3) no preoperative hormonal or radiation therapy. Exclusion criteria were: 1) elevated prostatic acid phosphatase (clinical stage D0); 2) absence of preoperative Gleason sum due to diagnosis by fine-needle aspiration; or 3) inadequately documented preoperative stage. The prevalence of positive lymph nodes in this database was 212 (5.1%). The preoperative PSA range for these patients was 0.1–431.8 with a mean PSA of 9.3 ng/mL. The range for Gleason sum was from 2 to 10 with a mean Gleason sum of 6.1. The TNM range was T1a to T3a. The distribution of TNM stage for patients in all three databases used in this study is shown in Figure 1. Preoperative stage was assigned according to the definitions given in Figure 2 (which are not AJCC/UICC definitions).[9,10]

Two smaller databases were used to further test the derived cutoffs. The first of the two "verification" databases has been described previously by Stone et al.[11] and consisted of 227 consecutive patients with clinically localized prostate carcinoma who underwent staging laparoscopic lymphadenectomy before definitive treatment at Mount Sinai Medical Center. The patients in this database had negative bone scans and computed tomography examinations for pelvic adenopathy. Sextant biopsies were performed and graded by one pathologist using the Gleason system. Clinical stage was determined using the TNM (American Joint Committee on Cancer) staging system. The prevalence of lymph node spread in this database was 22 (9.7%). The PSA range was 1.6–190 ng/mL with a mean PSA of 19.2 ng/mL. The TNM range was T1a to T3a. The Gleason sum range was 2–10 with a mean of 6.

The second verification database consisted of 330 patients from Walter Reed Army Medical Center (WRAMC) with clinically localized prostate carcinoma who underwent radical prostatectomy and open pelvic lymphadenectomy. The prevalence of lymph node spread in this database was 10 (3.0 %). The PSA range was 0.0–172.2 ng/mL with a mean PSA of 12.2 ng/mL. The TNM range was T1a to T3c. The Gleason sum range was from 2 to 10 with a mean of 5.

Fifty percent (2067) of the patients were randomly

selected from the development database for use in determining the cutoff of the decision-tree algorithm. A proprietary rule-derivation procedure based on binary recursive partitioning[12] (Xaim Inc., Colorado Springs, CO) evaluated the preoperative PSA, Gleason sum, TNM stage, and lymph node status to determine the cutoff for each clinical parameter. This methodology maximized the number of patients that could be reasonably classified as low risk, without allowing an unacceptably high false-negative rate. The binary recursive partitioning technology used in this study did not recognize TNM score as having discriminating power for this purpose.

To verify this result, we compared the distribution of TNM scores between those patients classified as low risk who had lymph node metastases (false-negatives) and those who did not (true-negatives). A discrepancy was noted between the two groups, which was determined to be statistically significant using chi-square analysis. This discrepancy suggests that TNM score could be used to further improve the negative predictive value of the classification scheme. A TNM cutoff that produced an acceptable balance between false-negative fraction (arbitrarily selected to be $< 1\%$) and percentage of patients classified as low risk was empirically derived by trial and error. The cutoff derived then was applied to the remainder of the development database as well as to all patients in the verification database.

RESULTS

Of the 4133 patients in the development database, 2353 (57%) were classified as low risk for developing lymph node disease by using only PSA and Gleason sum. Of those patients in the low risk group, 32 had lymph node metastases, giving a false-negative rate of 1.4%.

Of the 32 patients comprising the false-negative population, 14 had clinical stage T2a or below, (i.e., tumors affecting $<$ 50% of a single lobe on digital rectal exam), and 18 had more extensive palpable tumors. These values were compared with the 2321 low risk patients without lymph node spread (the true-negatives) in which 1814 had clinical stages of T2a and below, and 507 had stages T2b and above. The difference in clinical stage between the true-negative and false-negative populations was statistically significant ($P < 0.00001$).

When the TNM score was added to the analysis, 1829 patients or 44% were classified as low risk, with 14 having lymph node disease. This amounts to a false-negative rate of 0.8%.

The same rules using PSA and Gleason sum alone were applied to the first verification database (Mt. Sinai) resulting in 49 of 227 patients (22%) being classified as low risk, with one patient having lymph node positive disease (2% false-negatives). Consideration of TNM score resulted in 25 of 227 (11%) being classified as low risk, with none of those (0%) having lymph node involvement.

When the PSA and Gleason cutoffs were applied to the second verification database (WRAMC), 195 of 330 patients (59%) were classified as low risk for lymph node spread with an associated false-negative rate of 1.0%. Adding the TNM cutoff yielded a low risk group of 141 (43%) and a false-negative rate of 0.7%.

DISCUSSION

Diagnostic methods that enable earlier detection of prostate carcinoma have naturally resulted in a lower prevalence of lymph node metastases in patients with that disease.[3] This has raised the possibility of identifying a population of prostate carcinoma patients, based on clinical and pathologic data, whose risk of lymph node involvement is so low that pelvic lymph node dissection may be reasonably foregone.

This study used a simple decision-tree analysis to define the low risk population. A proprietary artificial intelligence algorithm was used to derive the cutoff for decision-tree analysis. These cutoffs were similar to those published by other investigators using different methods.[2,7,8] Using these cutoffs, we were able to identify a population of patients in whom the prevalence of lymph node disease was approximately 2%. This result was found to be applicable to patients treated at different institutions, even though the clinical and pathologic characteristics of those patients varied markedly between institutions.

There was no cross-validation of subjective variables such as Gleason score or clinical exam between institutions. Nevertheless, the decision rule derived by the artificial intelligence algorithm was valid when applied to patients from different institutions. This suggests that the distinction between palpable tumors as defined in Figure 1 is sufficient to minimize errors from subjective variations between clinicians.

Although the simple decision-tree analysis using only PSA and Gleason sum was able to classify 57% of patients in the development database as low risk, the same classification applied to only 22% of patients from the Mt. Sinai database and 59% of the WRAMC database. Nevertheless, the classification was equally valid between the three groups, with a false-negative rate of $\leq 2\%$ in each population. Because the results that were obtained from databases derived at different institutions were valid, it is difficult to interpret the significance of any variations that may exist between digtal rectal exam (DRE) and Gleason sum at different

institutions. The subjective components of DRE and Gleason sum interpretation do not appear to influence the validity of the rules derived here.

The artificial intelligence tool was unable to discern a significant relationship between TNM score and lymph node metastases. This result may have been an artifact of the database used for training the rule-derivation program, or may have been the result of the ordinal nature of the TNM data; i.e., the intervals between any two adjacent TNM values is somewhat arbitrary with no exact correlation to the severity of disease. This result may also follow from complex interactions between the clinical variables and lymph node status. For example, it has been observed that poorly differentiated tumors may lose the ability to produce PSA; therefore, the PSA may be positively correlated at moderate values and negatively correlated at higher values.

When we examined the distribution of TNM scores in the patients classified as low risk, we noted a discrepancy between those with lymph node disease and those without (i.e., between the false-negatives and true-negatives). This implied that consideration of TNM score could further improve the negative prediction value of the classification scheme, as other investigators have observed.[7] Using the empirically derived cutoff of stage T2a, we were able to reduce the false-negative rate to 0.8% in the development database. The TNM cutoff reduced the false-negative rates in the Mt. Sinai and WRAMC databases to 0.0% and 0.7%, respectively. The trade-off for this improvement in negative predictive value was that the percentage of patients classified as low risk declined to 44% in the development database, 11% in the Mt. Sinai database, and 43% in the WRAMC database.

The classification of patients as low risk or not low risk conforms to the decision that the clinician and patient must make in determining the extent of surgery. In this study, we were able to identify patients at low risk of lymph node metastases by using clinical and pathologic data typically collected before surgical intervention. The rule derived is very simple: If the preoperative PSA is $\leq$ 10.6 ng/mL, the biopsy Gleason sum is $\leq$ 6, and the tumor does not palpably involve more than 50% of a single lobe, there is a very low chance of lymph node metastases.

Although lymph node dissection is a relatively safe procedure, it is associated with some clinical risk as well as financial cost. If the prevalence of lymph node metastases were sufficiently low, the clinical benefit derived from lymph node dissection would justify neither the cost nor the risk. In the Medicare population, there were an estimated 69,000 radical prostatectomies in 1992–1993. An estimate of the costs attributable to lymph node dissection is difficult to make because this procedure affects not only operating room time and pathology services but may affect the duration of postprocedure hospitalization and impact the type of surgery selected for radical prostatectomy.[6] Furthermore, an accurate prediction of lymph node status may affect the decision to undertake extensive imaging studies as part of the presurgical evaluation and influence the staging procedures undertaken for purposes of radiation therapy, cryosurgery, and perineal prostatectomy. Taken together, it is apparent that foregoing lymph node dissection and imaging studies in patients at low risk of lymph node disease would result in significant cost savings.

In this study, patient clinical characteristics were successfully analyzed using an automated, computer-based algorithm. The classification process used a simple decision-tree approach that made no assumptions about interactions between variables. More sophisticated computer analysis techniques, such as artificial neural networks, may enhance the accuracy of the classifications presented here by accounting for these interactions. The existence of this computer analysis technique has been suggested by other investigators.[13] This should be a fertile field for further research.

It is also noted that clinical stage as determined by the digital rectal exam has a beneficial effect on the negative predictive value of the rules presented here but may adversely affect the number of patients that can be spared lymph node dissection. This is because the difference in classifying a patient as low risk or not low risk depends on whether a tumor is classified as Stage T2a or T2b, which in practical terms is whether or not the palpable tumor affects less than or more than 50% of a lobe. Presumably, the value of TNM staging in predicting lymph node metastases arises from the correlation of tumor size with both age and aggressiveness of the tumor. More accurate estimates of tumor size by using transrectal ultrasound or other imaging modalities may enhance the predictive value of the rules presented in this article and may be an area worthy of further research.

This investigation demonstrates that patients can be classified as low risk by using a defined number of clinical parameters. The classification is independent of any subjective clinical judgement not contained in the input variables. This study highlights the strengths and weaknesses of using artificial intelligence methods for analyzing patient data from large databases and multiple institutions. As databases become larger, more complex, and heterogeneous, it is expected that the artificial intelligence methods for analyzing them also will develop yielding clinical information beyond

that currently available. The finding that artificial intelligence methods yield valid, although not unique results, indicates that they are valid methods of data analysis.

CONCLUSION

A population of prostate carcinoma patients at low risk of lymph node metastases can be identified using a simple decision-tree algorithm that considers preoperative PSA, Gleason sum, and clinical stage. The methodology of this article allows identification of a cohort of prostate carcinoma patients with a defined level of risk ($< 1\%$). It is up to the clinician and patient to decide how such a classification and accompanying risk of lymph node disease influences a particular decision.

REFERENCES

1. Alagiri M, Colton MD, Seidmon EJ, Greenberg RE, Hanno PM. The staging pelvic lymphadenectomy: implications as an adjunctive procedure for clinically localized prostate cancer. *J Urol* 1997;80:243–6.
2. Bishoff JT, Reyes A, Thompson IM, Harris MJ, St. Clair SR, Gomella L, et al. Pelvic lymphadenectomy can be omitted in selected patients with carcinoma of the prostate: development of a system of patient selection. *Urology* 1995;45:270–4.
3. Danella JF, de Kernion JB, Smith RB, Steckel J. The contemporary incidence of lymph node metastases in prostate cancer: implications for laparoscopic lymph node dissection. *J Urol* 1993;149:1488–91.
4. Petro JA, Catalona WJ. Lower incidence of unsuspected lymph node metastases in 521 consecutive patients with clinically localized prostate cancer. *J Urol* 1992;147:1574–5.
5. Rogers E, Gürpinar T, Dillioglugil Ö, Kattan MW, Goad JR, Scardino PT, et al. The role of digital rectal examination, biopsy Gleason sum and prostate specific antigen in selecting patients who require pelvic lymphnode dissections for prostate cancer. *Br J Urol* 1996;78:419–25.
6. Bluestein DL, Bostwick DG, Bergstralh EJ, Oesterling JE. Eliminating the need for bilateral pelvic lymphadenectomy in select patients with prostate cancer. *J Urol* 1994;151:1315–20.
7. Narayan P, Fournier G, Gajendran V, Leidich R, Lo R, Wolf JS, et al. Utility of preoperative serum prostate specific-antigen concentration and biopsy Gleason sum in predicting risk of pelvic lymph node metastases in prostate cancer. *Urology* 1994;44:519–24.
8. Sands ME, Zagar GK, Pollack M, von Eschenbach AC. Serum prostate specific antigen, clinical stage, pathologic grade and the incidence of nodal metastases in prostate cancer. *Urology* 1994;44:215–20.
9. Partin AW, Kattan MW, Subing ENP, Walsh PC, Wojno KJ, Oesterling JE, et al. Combination of prostate-specific antigen, clinical stage and Gleason sum to predict pathological stage of localized prostate cancer: a multi-institutional update. *JAMA* 1997:277:1445–51.
10. Schroder FF, Hermanek P, Denis L, Fair WR, Gospodarowicz M. TNM classification of prostate cancer. *Prostate* 1992;4:129.
11. Stone NN. Stock RG. Parikh D. Yeghiayan P. Unger P. Perineural invasion and seminal vesicle involvement predict pelvic lymph node metastasis in men with localized carcinoma of the prostate. *J Urol* 1998;160:1722–6.
12. Kim H, Koehler GJ. Theory and practice of decision tree induction. *Omega: Int J Manage Sci* 1995;23:637–52.
13. Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, et al. The use of prostate specific antigen, clinical stage and Gleason score to predict pathologic stage in men with localized prostate cancer. *J Urol* 1993;150:110–4.