Juyoung A. Jung, MD
Fergus V. Coakley, MD
Daniel B. Vigneron, PhD
Mark G. Swanson, PhD
Aliya Qayyum, MD
Vivian Weinberg, PhD
Kirk D. Jones, MD
Peter R. Carroll, MD
John Kurhanewicz, PhD

Index terms:
Magnetic resonance (MR), spectroscopy, three-dimensional, 844.121411, 844.121415, 844.121416, 844.12145
Prostate neoplasms, MR, 844.121411, 844.121415, 844.121416, 844.12145

Published online
10.1148/radiol.2333030672
Radiology 2004; 233:701–708

Abbreviations:
CC/C = choline-creatine–to-citrate ratio
3D = three-dimensional

[1] From the Departments of Radiology (J.A.J., F.V.C., D.B.V., M.G.S., A.Q., J.K.), Pathology (K.D.J.), and Urology (P.R.C.), University of California San Francisco, San Francisco, Calif; and Biostatistics Core, University of California San Francisco Comprehensive Cancer Center, San Francisco, Calif (V.W.). Received April 28, 2003; revision requested July 10; final revision received March 6, 2004; accepted April 15. Supported by NIH grants R01 CA59897, R01 CA79980, and R33 CA88214. **Address correspondence to** J.K., University of California San Francisco Magnetic Resonance Science Center, Box 1290, AC 109, 1 Irving St, San Francisco CA 94143-1290 (e-mail: *johnk@mrsc.ucsf.edu*).

Authors stated no financial relationship to disclose.

Author contributions:
Guarantor of integrity of entire study, J.K.; study concepts and design, J.K., J.A.J., F.V.C.; literature research, J.A.J.; clinical studies, J.A.J., F.V.C., J.K., P.R.C.; experimental studies, J.K., J.A.J., F.V.C.; data acquisition, J.K., M.G.S., K.D.J., D.B.V.; data analysis/interpretation, J.K., M.G.S., F.V.C., A.Q.; statistical analysis, V.W., J.K., J.A.J., F.V.C.; manuscript preparation, definition of intellectual content, revision/review, and final version approval, J.A.J., F.V.C., J.K.; manuscript editing, J.A.J., F.V.C., D.B.V., M.G.S., A.Q., P.R.C., J.K.

Prostate Depiction at Endorectal MR Spectroscopic Imaging: Investigation of a Standardized Evaluation System[1]

PURPOSE: To investigate the accuracy and interobserver variability of a standardized evaluation system for endorectal three-dimensional (3D) magnetic resonance (MR) spectroscopic imaging of the prostate.

MATERIALS AND METHODS: The human research committee approved the study, and all patients provided written informed consent. Endorectal MR imaging and MR spectroscopic imaging were performed in 37 patients before they underwent radical prostatectomy. For the 22 patients with good or excellent MR spectroscopic imaging data, step-section histopathologic tumor maps were used to identify spectroscopic voxels of unequivocally benign (n = 306) or malignant (n = 81) peripheral zone tissue. Two independent spectroscopists, unaware of all other findings, scored the spectra of the selected voxels by using a scale of 1 (benign) to 5 (malignant) that was based on standardized metabolic criteria. Descriptive statistical, receiver operating characteristics (ROC), and κ statistical analyses of the data obtained by both readers were performed by using two definitions of cancer: one based on a voxel score of 3–5 and the other based on a score of 4 or 5.

RESULTS: The scoring system had good accuracy (74.2%–85.0%) in the differentiation between benign and malignant tissue voxels, with areas under the ROC curve of 0.89 for reader 1 and 0.87 for reader 2. Specificities of 84.6% and 89.3% were achieved when a voxel score of 4 or 5 was used to identify cancer, and sensitivities of 90% and 93% were achieved when a score of 3–5 was used to identify cancer. Readers demonstrated excellent interobserver agreement (κ values, 0.79 and 0.80).

CONCLUSION: The good accuracy and excellent interobserver agreement achieved by using the standardized five-point scale to interpret peripheral zone metabolism demonstrate the potential effectiveness of using metabolic information to identify prostate cancer, and the clinical usefulness of this system warrants testing in prospective clinical trials of MR imaging combined with MR spectroscopic imaging.

Approximately 8% of American men will receive a diagnosis of prostate cancer during their lifetime, and 20% of these men will die of the disease (1). Despite this sizeable mortality rate, many cases of prostate cancer are subclinical, and microscopic foci of incidental prostate cancer can be detected in up to 40% of men at autopsy (2). The management of early-stage prostate cancer is controversial because the methods currently available for the noninvasive evaluation of this malignancy have substantial limitations in the differentiation of patients with indolent and incidental disease from those with progressive and life-threatening disease (3–5). As a result, there have been several studies to investigate the role of magnetic resonance (MR) imaging in the evaluation and staging of prostate cancer, with relatively promising results (6,7).

More recently, the addition of metabolic information obtained at three-dimensional (3D) endorectal MR spectroscopic imaging to the anatomic data obtained at MR imaging has been shown to improve tumor localization, tumor staging, tumor volume estimation, and assessment of tumor aggressiveness (8–10). Specifically, in MR spectra acquired from regions of prostate cancer, citrate and polyamine levels are either substantially reduced or absent, whereas choline levels are elevated relative to creatine (11).

The metabolic criteria used to identify prostate cancer have evolved from empiric observations made at nearly 4000 clinical MR imaging–MR spectroscopic imaging examinations and from data obtained at ex vivo high-resolution magic-angle spinning spectroscopy of biopsy and surgical tissues that were assessed at subsequent full histopathologic analysis (12,13). To our knowledge, however, these metabolic criteria have not yet been systematically standardized or validated for interobserver agreement and accuracy. Therefore, this study was undertaken to investigate the accuracy and interobserver variability of a standardized evaluation system for endorectal 3D MR spectroscopic imaging of the prostate.

MATERIALS AND METHODS

Study Subjects

This was a retrospective single-institution cross-sectional study. We included all patients who underwent radical prostatectomy between January and December 1999 and had undergone MR imaging and 3D MR spectroscopic imaging of the prostate (n = 37). The 37 patients were referred for MR imaging after a diagnosis of prostate cancer had been established with biopsy. Patients were recruited as part of an ongoing National Institutes of Health study to investigate the use of MR imaging in patients with prostate cancer, and they were also included in a separate study to investigate the measurement of prostate cancer tumor volume with MR imaging and MR spectroscopic imaging (10). The University of California San Francisco Committee on Human Research approved all of the study procedures, and written informed consent was obtained from all patients.

The quality of each 3D MR spectroscopic image was rated (by J.K.) as excellent (n = 10), good (n = 12), fair (n = 10), or poor (n = 5) on the basis of signal-to-noise ratio data, magnetic field homogeneity, and the presence or absence of water- and lipid-induced baseline distortions. Specifically, an image was considered to have excellent spectral quality when the signal-to-noise ratios of all metabolites were higher than 10, all metabolic resonances were well resolved, and there were no baseline distortions due to residual water or lipid. An image was considered to have good spectral quality when the signal-to-noise ratios of all metabolites were between 8 and 10, all metabolic resonances were reasonably well resolved, and there were minimal baseline distortions due to residual water or lipid. Images with lower signal-to-noise ratios were considered to have fair spectral quality provided there was no lipid contamination. Images with substantial lipid contamination were considered to have poor spectral quality. Patients with poor or fair spectral data (n = 15) were excluded, and the remaining 22 patients formed the final study group.

The mean age of the final study group was 56 years (age range, 44–69 years). The mean serum prostate-specific antigen level was 5.9 ng/mL (range, 2.2–17.2 ng/mL), with a median value of 5.0 ng/mL. In general, a prostate-specific antigen level of 4 ng/mL or lower is considered normal; however, more exact normal ranges account for the patient's race and age (14). The mean Gleason score was 6.6 (score range, 5–8), with a median score of 7.0. The mean period between the MR imaging and MR spectroscopic imaging examinations and radical prostatectomy was 46 days (range, 2–160 days). None of the patients underwent preoperative hormonal or radiation therapy.

MR Imaging Techniques

MR imaging was performed by using a 1.5-T whole-body MR imaging unit (Signa; GE Medical Systems, Milwaukee, Wis). The patients were imaged while in the supine position by using a body coil for signal excitation and a pelvic phased-array coil (GE Medical Systems) combined with a balloon-covered expandable endorectal coil (Medrad, Pittsburgh, Pa) for signal reception. Transverse spin-echo T1-weighted images were obtained from the aortic bifurcation to the symphysis pubis by using the following parameters: 700/8 (repetition time msec/echo time msec), a 5-mm section thickness, a 1-mm intersection gap, a 24-cm field of view, a 256 × 192 matrix, a transverse frequency direction, and one acquired signal.

Thin-section high-spatial-resolution transverse and coronal T2-weighted fast spin-echo MR images of the prostate and seminal vesicles were obtained by using the following parameters: 6000/96 (effective), an echo train length of 16, a 3-mm section thickness, no intersection gap, a 14-cm field of view, a 256 × 192 matrix, an anteroposterior frequency direction (to prevent obscuration of the prostate by endorectal coil motion artifact), and three acquired signals. All MR images were routinely postprocessed to compensate for the reception profile of the endorectal and pelvic phased-array coils (15).

After two authors (J.K. and M.G.S.) reviewed the transverse T2-weighted MR images, a spectroscopic MR imaging volume was selected to maximize coverage of the prostate while minimizing the inclusion of periprostatic fat and rectal air. Three-dimensional MR spectroscopic imaging data were acquired by using a water- and lipid-suppressed double-spin-echo point-resolved spectroscopic sequence optimized for the quantitative detection of both choline and citrate. Water and lipid suppression was achieved by using the band selective inversion with gradient dephasing technique (16). To eliminate signals from adjacent tissues, especially periprostatic lipids and the rectal wall (17), outer voxel saturation pulses also were used. Data sets were acquired as 16 × 8 × 8 phase-encoded spectral arrays (1024 voxels) by using a nominal spectral resolution of 0.24–0.34 cm^3, 1000/130, and a 17-minute acquisition time. The total examination time, including the time spent placing the coil and positioning the patient, was 1 hour.

MR Spectroscopic Imaging Data Processing and Analysis

Three-dimensional MR spectroscopic imaging data were processed, aligned with the corresponding MR imaging data, displayed, and analyzed by using a combination of in-house software and Interactive Display Language (Research Systems, Boulder, Colo) software tools (18). The raw spectral data were apodized with a 1-Hz Gaussian function and Fourier transformed in the time domain and in three spatial domains. Choline, creatine, and citrate peak parameters (ie, peak area, peak height, peak location, and line width) were estimated by using an iterative procedure that was used to first identify statistically significant peaks—that is, those with a signal-to-

noise ratio higher than 5—in the magnitude spectrum and then estimate a frequency shift that best aligned these peaks with the expected locations of choline, creatine, citrate, and residual water. Subsequently, the spectra were phased by using the phase of the residual water and the metabolite resonances, and baseline values were corrected by using a local nonlinear fit to the nonpeak regions of the spectra.

Areas of peak choline, creatine, and citrate levels were localized by using numeric integration over a frequency range that was determined according to the metabolite peak locations and widths (18). In regions of healthy tissue, the peak polyamine level that resonated between the peak choline and creatine levels could not be sufficiently resolved and was incorporated into the area of the peak choline-plus-creatine (choline-creatine) level. The quantification algorithm also yielded estimates of spectral noise and the metabolite peak area-to-noise ratio (18).

Histopathologic Identification of Benign and Malignant Spectroscopic Voxels

Radical prostatectomy specimens were coated with India ink and fixed in 10% buffered formaldehyde. Transverse step sections were obtained in 3–4-mm intervals in a plane perpendicular to the long axis of the prostate. A pathologist (K.D.J.) with 6 years of experience recorded the location of all peripheral zone tumor nodules on a standardized diagram of the prostate and recorded the maximum transverse diameter of all tumor nodules. Two radiologists (F.V.C. and A.Q., with 8 and 4 years experience, respectively) in consensus labeled the voxels on the spectroscopic grid that were considered to clearly consist of benign or malignant tissue on the basis of MR imaging and histopathologic tumor map findings. Voxels were only labeled as consisting of benign or malignant tissue when there was a clear-cut concordance between the MR imaging (consensus reading by F.V.C. and A.Q.) and histopathologic findings—with no postbiopsy hemorrhage seen on the T1-weighted MR images—and no potential for partial volume effects (ie, overlap of tumorous voxel with surrounding healthy tissue, ejaculatory zone, or central gland tissues). Allowances for differences in registration between the MR imaging and histopathologic findings were made. On the basis of this standard of reference, 387 voxels—out of a total of 1430 peripheral zone voxels in the 22 patients—were labeled as benign (n = 306) or malignant (n = 81).

Interpretation of 3D MR Spectroscopic Imaging Data

Two experienced spectroscopists (J.K. and M.G.S., with 13 and 4 years experience, respectively) independently reviewed the spectra of the 387 voxels that had been labeled as benign or malignant. These readers were aware that the spectra were derived from patients with biopsy-proved prostate cancer and represented either benign or malignant tissue, but they were unaware of which voxels had been labeled benign or malignant and of all other clinical, histopathologic, and MR imaging findings.

The spectra were interpreted and scored on the basis of prior research findings and current understanding of prostate cancer metabolism (12,13). The citrate level is characteristically high in healthy prostatic tissue because the presence of zinc inhibits the first enzyme in the Krebs cycle (19,20). Citrate levels decrease with prostate cancer, but they can also be reduced owing to prostatitis or postbiopsy hemorrhage (21). The level of choline, a cell membrane constituent, increases with prostate cancer owing to increased membrane turnover, changes in cellular density, and phosholipid metabolism (22,23). Polyamine levels have been shown to decrease with prostate cancer (24,25). The polyamine level peak occurs between the creatine and choline level peaks and cannot be entirely resolved from these peaks. However, decreased polyamine levels can be subjectively recognized as a sharper separation of the creatine and choline level peaks (12).

The peak area choline-creatine–to-citrate ratio (CC/C) and choline-to-creatine ratio were calculated. The choline-to-creatine peak area ratio can be estimated only in cancerous regions because in healthy voxels, there is poor in vivo spectral resolution of these metabolites owing to the presence of a large polyamine resonance. Spectroscopic voxels were scored on a standardized five-point scale by using the following criteria:

1. A primary score of 1–5 was assigned on the basis of the mean normal CC/C (26). The mean CC/C was determined to be 0.22 ± 0.013 (standard deviation) on the basis of the results of a previously published study (26) involving the use of the MR spectroscopic imaging data acquisition and processing techniques used in the present study. A score of 1 was assigned to voxels with a CC/C 1 standard deviation from the mean normal value. A score of 2 was assigned to voxels with a CC/C more than 1 but 2 or fewer standard deviations above the mean normal value. A score of 3 was assigned to voxels with a CC/C more than 2 but 3 or fewer standard deviations above the mean normal value. A score of 4 was assigned to voxels with a CC/C more than 3 but 4 or fewer standard deviations above the mean normal value. A score of 5 was assigned to voxels with a CC/C more than 4 standard deviations above the mean normal value.

2. An initial adjustment to the primary voxel score was made to account for elevation of the choline level relative to the creatine level and for reduced polyamine levels. When the choline-to-creatine ratio was greater than or equal to 2, with a primary voxel score of 2 or 3, the overall score was increased to 4. When the choline-to-creatine ratio was less than 2 or there was no reduction in polyamine levels, with a primary score of 4 or 5, the overall score was decreased by 1—that is, to 3 or 4, respectively.

3. A final adjustment to the score was made to account for poor spectral signal-to-noise ratios. Poor signal-to-noise ratio was defined as a peak area-to-noise ratio of less than 8 for voxels with a score of 3–5 and of less than 5 for voxels with a score of 1 or 2. In the presence of a poor signal-to-noise ratio, a score of 1 became 3, a score of 2 or 4 became 3, and a score of 5 became 4. A score of 3 was not changed owing to low signal-to-noise ratio criteria.

This standardized voxel-scoring system yielded a final score of 1–5 and was designed so that the following interpretative scale could be applied: Voxels with a score of 1 were considered to be likely benign; voxels with a score of 2, probably benign; voxels with a score of 3, equivocal; voxels with a score of 4, probably malignant; and voxels with a score of 5, likely malignant. In addition to using this five-point scoring system, readers were allowed to deem spectra unusable if they showed substantial lipid contamination or misalignment of metabolite resonance peaks.

Statistical Analyses

For all labeled voxels that were considered to be usable by both spectroscopists, reader-assigned scores of 1–5 were compared with the histopathologic reference-

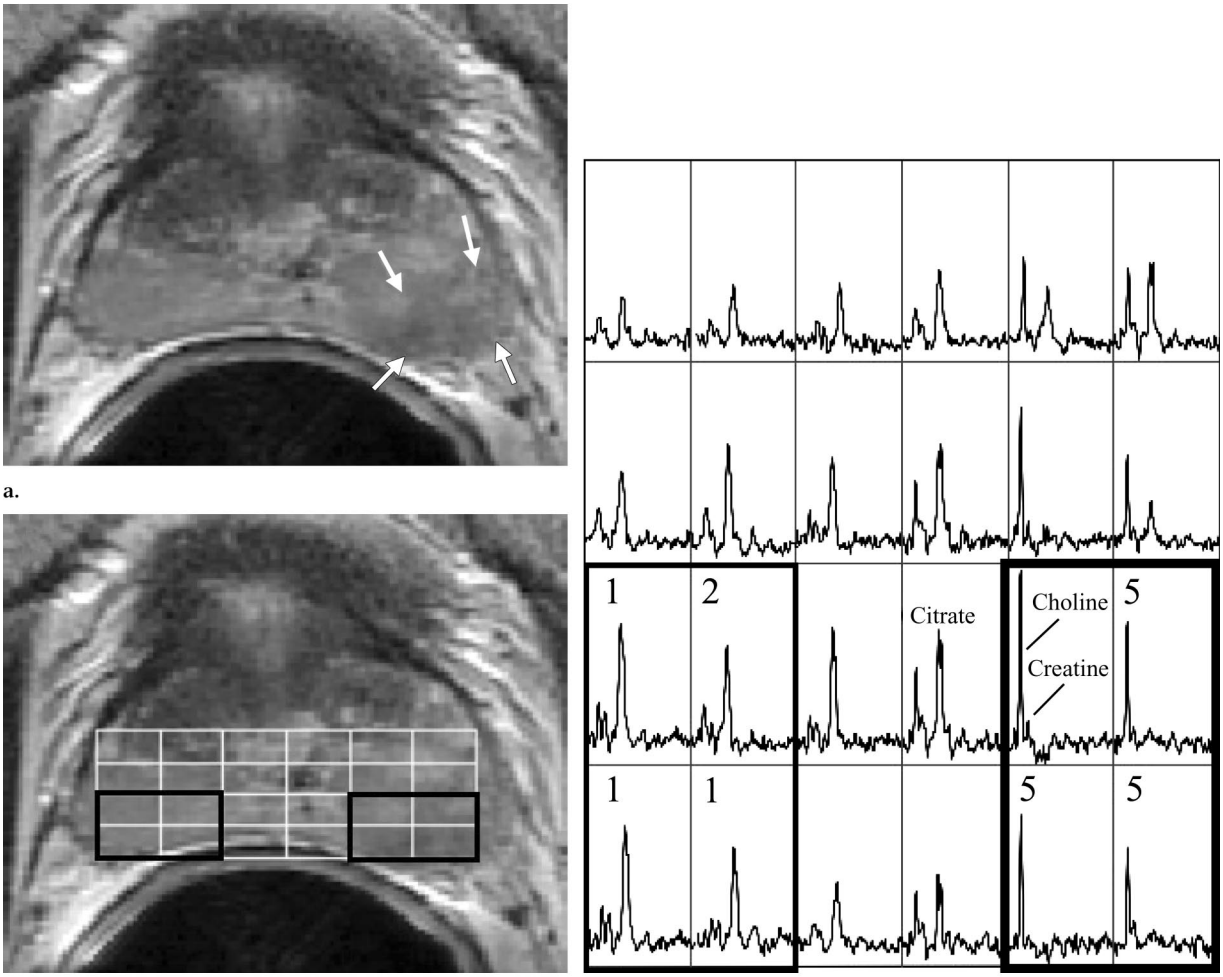

Figure 1. **(a)** Representative reception profile–corrected transverse fast spin-echo T2-weighted MR image (6000/96 [effective]) shows region of low signal intensity (arrows) that correlated with Gleason 4 + 3 cancer in the left middle region of the prostate gland on the corresponding step-section histopathologic diagram (not shown). **(b, c)** The selected volume for spectroscopy (white box) and a portion of the 16 × 8 × 8 spectral phase encode grid from one of eight axial spectroscopic sections are shown overlaid (fine white line) on the transverse T2-weighted MR image (**b**), with the corresponding axial 0.3-cm^3 proton spectral array (**c**). **(b)** The radiologists defined the voxels in the outlined portion of the grid on the right side of the image as a region of prostate cancer and those in the outlined portion of the grid on the left side as a region of healthy peripheral zone tissue. **(c)** The spectroscopists accurately judged the corresponding spectra outlined on the right side of the array to represent "likely cancer" (score of 5) and those outlined on the left side to represent healthy tissue (score of 1 or 2).

standard diagnoses of benign or malignant tissue. Interreader agreement was evaluated by using κ statistics. The degree of observer agreement was graded as follows: a κ value of 0–0.20 indicated slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement (27). Descriptive statistical data (sensitivity, specificity, and positive and negative predictive values) were determined (by V.W.) by using two dichotomized rating systems: one with voxel scores of 1–3 indicating benignancy and a score of 4 or 5 indicating malignancy and the other with a voxel score of 1 or 2 indicating benignancy and a voxel score of 3–5 indicating malignancy.

Because of the lack of independence of voxels in each patient, estimates of sensitivity and specificity were first calculated for each patient and a χ^2 test was used to determine the difference in sensitivity and specificity among the patients, with each patient representing a stratum. The overall statistical estimates were similar with and without patient stratification. Therefore, the final results of this analysis were based on pooled data. Receiver operating characteristic curves of true- versus false-positive proportions of malignant voxels were generated for each reader, and the areas under the curves were compared by using the method of Hanley and McNeil (28).

RESULTS

Figure 1 illustrates the approach that we used in this study to develop and evaluate a standardized system for analyzing data obtained at MR spectroscopic imaging of the prostate. Figure 1a shows a representative reception profile–corrected T2-weighted fast spin-echo transverse MR image obtained at the middle of the prostate gland in a patient with cancer detected in the left lobe at radical prostatectomy. On the

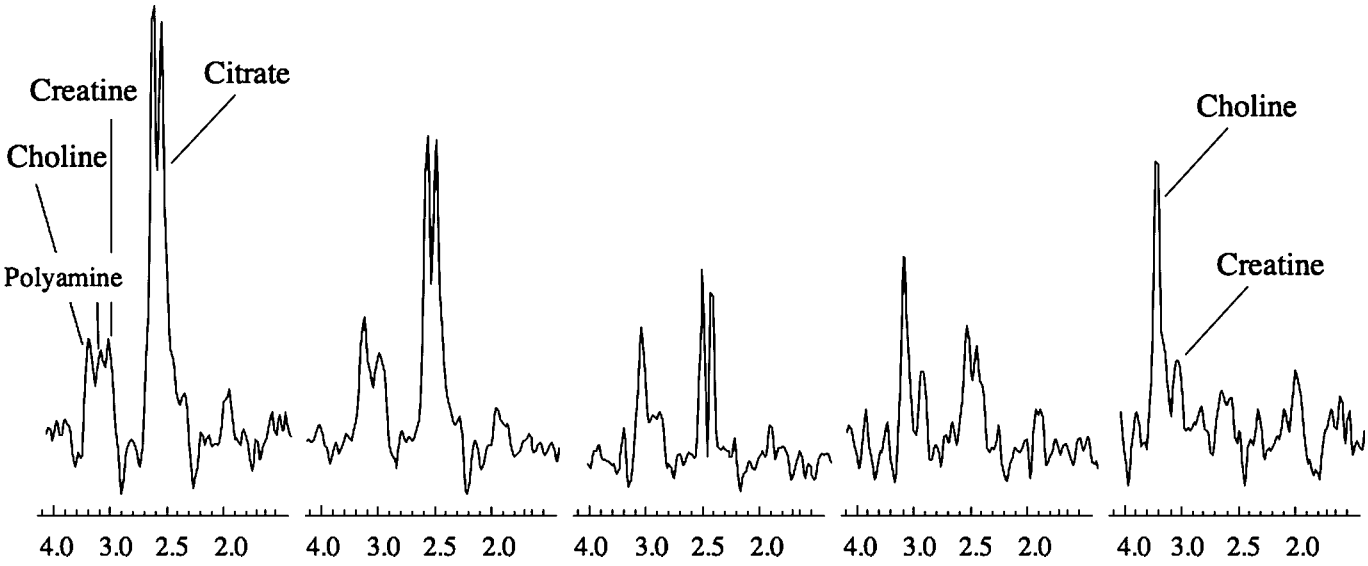

Figure 2. Representative proton spectra acquired from a 3D MR spectroscopic imaging data set for a patient with prostate cancer demonstrate the spectral patterns associated with the standardized five-point scale used to interpret peripheral zone metabolism.

TABLE 1
Three-dimensional MR Spectroscopic Imaging–derived Cancer Scores Assigned by Readers 1 and 2

	Reader 1			Reader 2		
Cancer Score	Total	Benign	Malignant	Total	Benign	Malignant
1	157 (41)	156	1	98 (26)	98	0
2	70 (18)	63	7	115 (30)	109	6
3	65 (17)	48	17	69 (18)	46	23
4	46 (12)	23	23	54 (14)	35	19
5	42 (11)	9	33	44 (12)	11	33
Total	380	299	81	380	299	81

Note.—Data are numbers of voxels. Numbers in parentheses are percentages based on a total of 380 voxels.

basis of knowledge gleaned from the postprostatectomy histopathologic diagram, the two radiologists in consensus selected spectroscopic voxels that were clearly in the region of the tumor (low signal intensity) in the left lobe of the prostate (right side of image) and voxels that were in a high-signal-intensity region of the contralateral normal lobe (Fig 1b).

The quality of the spectral data was rated as excellent, and both spectroscopists independently scored the voxels the same way without knowledge of the MR imaging and histopathologic results (Fig 1c). The voxels with a score of 5 (Fig 1b, right side), as compared with the voxels with a score of 1 or 2 on the contralateral side, demonstrated dramatically elevated choline-to-creatine ratios and an absence of citrate and polyamines. If metabolic changes in regions of prostate cancer were always as clear as those demonstrated in Figure 1, there would be no need for a systematic scoring system. In actual practice, however, the spectral patterns identified in patients with prostate cancer can range from clearly normal to clearly abnormal. Representative spectra that correspond to scores used in the proposed interpretative five-point standardized scoring system are shown in Figure 2. Numbers along the axis at the bottom of Figure 2 represent frequencies, in parts per million. Each metabolite resonates at a distinct frequency.

Both spectroscopists independently determined that 380 of the 387 selected voxels were usable, and both of them judged the same seven voxels to be unusable. In Table 1, the results of using the five-point scoring system for all 380 voxels (299 benign and 81 malignant voxels) that were considered usable by both readers are compared with the histopathologic reference-standard diagnoses of benign or malignant tissue. Both spectroscopists judged nearly the same percentage of voxels to be clearly normal (59% [reader 1] and 56% [reader 2] of voxels assigned score of 1 or 2), equivocal (17% [reader 1] and 18% [reader 2] of voxels assigned score of 3), and clearly abnormal (23% [reader 1] and 26% [reader 2] of voxels assigned score of 4 or 5). The clearly normal voxels demonstrated a mean CC/C of 0.25 ± 0.12, which is not substantially different from a previously reported value of 0.22 ± 0.013 (23). However, the mean CC/C in the healthy peripheral zone will most likely change according to the specific MR spectroscopic imaging acquisition and processing protocol used.

The diagnoses of benign or malignant tissue assigned to voxels at the histopathologic reference-standard examination and by the two readers with use of both score definitions of cancer (voxel score of 3–5 and voxel score of 4 or 5) are shown in Table 2. The readers showed substantial agreement with use of the two dichotomized scales: The κ value was 0.80 when cancer was defined as tissue with a voxel score of 3–5 and 0.79 when it was defined as tissue with a score of 4 or 5 (Table 3).

The descriptive statistical values calculated by using the two dichotomized rating schemes are shown in Table 4. For both spectroscopists, sensitivity was

moderate (69% and 64%) and specificity was high (89.3% and 84.6%) when a voxel score of 4 or 5 was used to define cancer. Sensitivity was high (90% and 93%) and specificity was moderate (73.2% and 69.2%) when a voxel score of 3–5 was used to define cancer. The majority of negative predictive values for the two readers were higher than 90% with use of either cancer definition, but the positive predictive values were only moderate (44.9%–63.6%). The areas under the receiver operating characteristic curve were 0.89 for reader 1 and 0.87 for reader 2 (Fig 3). These values were not significantly different ($P = .26$).

TABLE 2
Diagnoses Assigned to Benign and Malignant Voxels at Histopathologic (Reference-Standard) Analysis and by Readers 1 and 2 with Use of Both Score Definitions of Cancer

	Reader 1		Reader 2	
Final Diagnosis	Malignant, Cancer Score 3–5	Malignant, Cancer Score 4 or 5	Malignant, Cancer Score 3–5	Malignant, Cancer Score 4 or 5
Benign ($n = 299$)	80	32	92	46
Malignant ($n = 81$)	73	56	75	52

Note.—Data are numbers of voxels.

TABLE 3
Distribution of Cancer Scores Assigned by Readers 1 and 2

	Score Assigned by Reader 2				
Score Assigned by Reader 1	1	2	3	4	5
1	85	66	6	0	0
2	13	37	18	2	0
3	0	11	36	16	2
4	0	1	9	26	10
5	0	0	0	10	32

Note.—Data are numbers of voxels ($n = 380$).

TABLE 4
Three-dimensional MR Spectroscopic Imaging Tumor Detection on a Voxel-by-Voxel Basis for Readers 1 and 2

	Cancer, Score 4 or 5		Cancer, Score 3–5	
Parameter	Reader 1	Reader 2	Reader 1	Reader 2
Sensitivity	69 (56/81)	64 (52/81)	90 (73/81)	93 (75/81)
Specificity	89.3 (267/299)	84.6 (253/299)	73.2 (219/299)	69.2 (207/299)
Positive predictive value	63.6 (56/88)	53.1 (52/98)	47.7 (73/153)	44.9 (75/167)
Negative predictive value	91.4 (267/292)	89.7 (253/282)	96.5 (219/227)	97.2 (207/213)
Accuracy	85.0 (323/380)	80.3 (305/380)	76.8 (292/380)	74.2 (282/380)

Note.—Data are percentages. Numbers in parentheses are the numbers of prostatic voxels used to calculate the percentages.

DISCUSSION

We undertook this study to investigate the accuracy and objectivity of a standardized scoring system for endorectal 3D MR spectroscopic imaging evaluation of the prostate, because this modality is increasingly being used to assist in tumor localization (9,10), staging (8), and characterization (12) and in radiation treatment planning (29,30). Our study results show that the described standardized scoring system had good accuracy (74.2%–85.0%; areas under receiver operating characteristic curve, 0.87 and 0.89) and excellent interobserver agreement with use of a dichotomized scoring system in which a voxel score of 4 or 5 and voxel scores of 3–5 indicated the presence of cancer ($\kappa = 0.79$ and $\kappa = 0.80$, respectively). These data illustrate the ability of two independent readers to accurately and similarly label voxels as benign or malignant with 3D MR spectroscopic imaging.

Our study data indicate that specificities of 84.6% and 89.3%, as compared with a previously reported specificity of approximately 75% (9), were achieved when a voxel score of 4 or 5 was used to identify cancer. The higher specificity in our study most likely reflects our incorporation of additional criteria, such as the choline-to-creatine ratio and the presence or absence of polyamines, into the current standardized evaluation system, as well as the fact that only good or excellent spectra were selected for interpretation.

Sensitivity ranged from 64% to 93% (Table 2), depending on the reader and the cancer dichotomization scheme used. As anticipated, the less stringent dichotomization of cancer versus healthy tissue based on a voxel score of 3–5 resulted in higher sensitivity and lower specificity (ie, increased false-positive cases), whereas the dichotomization based on a score of 4 or 5 resulted in higher specificity and lower sensitivity (ie, increased false-negative cases). Because the use of MR imaging alone results in high sensitivity and moderate specificity (9), the dichotomization scheme based on a score of 4 or 5 is more appropriate when MR imaging is used in conjunction with 3D MR spectroscopic imaging.

Our study had a number of limitations. First, the analysis methods used may have reflected a best-case scenario, because only clearly benign and clearly malignant voxels were presented to the spectroscopic readers for analysis, and all of the voxels were selected from images of good or excellent spectral quality. However, as MR imaging technology continues to improve and with the emergence of clinical 3-T imaging units, we expect the quality of MR spectroscopic imaging to surpass that of current standards. Another confounding factor that may have lowered the positive predictive values reported in this study was the relatively small proportion of malignant voxels, which probably reflected the overall small volume of disease in these surgically treated patients.

With a software package for prostate spectroscopic imaging now being commercially available and a multi-institutional study of endorectal prostate MR

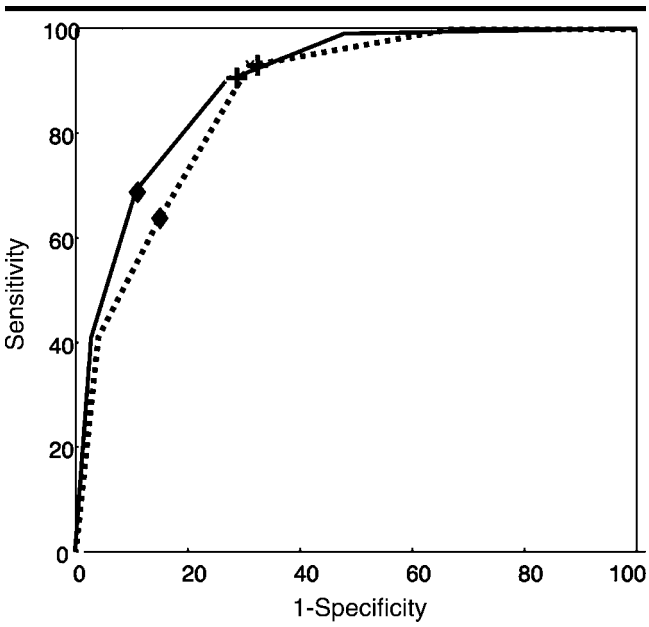

Figure 3. Receiver operating characteristic curves for 3D MR spectroscopic imaging results obtained by reader 1 (continuous line) and reader 2 (dashed line). The areas under the receiver operating characteristic curves for readers 1 (0.89) and 2 (0.87) were not significantly different. + = sensitivity and specificity achieved by readers with use of score of 3–5 to define metabolically abnormal tissue, ♦ = sensitivity and specificity achieved by readers with use of score of 4 or 5 to define metabolically abnormal tissue.

imaging combined with MR spectroscopic imaging having been recently launched by the American College of Radiology Imaging Network, additional research is needed to further refine and validate the current evaluation system for use in spectroscopic analysis. Specifically, future studies need to be conducted to evaluate the described scoring system combined with T2-weighted MR imaging and with all spectra—regardless of their quality—incorporated so that daily clinical practice can be better replicated.

In addition, future studies will need to address how strictly the presented quantitative rules for using this standardized scoring system need to be adhered to and whether a more qualitative assessment based on pattern recognition will yield equal or better accuracy. The future use of in vitro examinations performed with high-resolution magic-angle spectroscopy of ex vivo prostate tissue samples should lead to an increase in the number of metabolic markers used to identify prostate cancer and thereby to improved overall accuracy of cancer diagnosis and characterization.

In conclusion, the good accuracy and interobserver agreement achieved by using the described standardized five-point scale for interpreting peripheral zone metabolism indicate the potential usefulness of this system for metabolically identifying prostate cancer. However, this study was designed for the development and evaluation of a grading system for MR spectroscopic imaging data rather than for the in vivo determination of the accuracy of the system, which has yet to be confirmed by the results of prospective trials, such as the ongoing American College of Radiology Imaging Network Study.

References

1. American Cancer Society. Cancer facts and figures: 1999. Atlanta, Georgia: American Cancer Society.
2. Stamey TA, McNeal JE. Adenocarcinoma of the prostate. In: Walsh PC, Retik AB, Stamey TA, Vaughan ED, eds. Campbell's urology. 6th ed. Vol 2. Philadelphia, Pa: Saunders, 1992; 1159–1221.
3. Mettlin CJ, Black B, Lee F, et al. Workgroup #2: screening and detection—reference range/clinical issues of PSA. Cancer 1993; 71:2679–2680.
4. Salomon L, Colombel M, Patard JJ, et al. Value of ultrasound-guided systematic sextant biopsies in prostate tumor mapping. Eur Urol 1999; 35:289–293.
5. Smith JA, Scardino PT, Resnick MI, Hernandez AD, Rose SC, Egger MJ. Transrectal ultrasound versus digital rectal examination for the staging of carcinoma of the prostate: results of a prospective multi-institutional trial. J Urol 1997; 157: 902–906.
6. Yu KK, Hricak H, Alagappan R, Chernoff DM, Bacchetti P, Zaloudek CJ. Detection of extracapsular extension of prostate carcinoma with endorectal and phased-array coil MR imaging: multivariate feature analysis. Radiology 1997; 202:697–702.
7. Huch Boni RA, Boner JA, Debatin JF, et al. Optimization of prostate carcinoma staging: comparison of imaging and clinical methods. Clin Radiol 1995; 50:593–600.
8. Yu KK, Scheidler J, Hricak H, et al. Prostate cancer: prediction of extracapsular extension with endorectal MR imaging and three-dimensional proton MR spectroscopic imaging. Radiology 1999; 213: 481–488.
9. Scheidler J, Hricak H, Vigneron DB, et al. Prostate cancer: localization with three-dimensional proton MR spectroscopic imaging—clinicopathologic study. Radiology 1999; 213:473–480.
10. Coakley FV, Kurhanewicz J, Lu Y, et al. Prostate cancer tumor volume: measurement with endorectal MR and MR spectroscopic imaging. Radiology 2002; 223: 91–97.
11. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24–0.7 cm^3) spatial resolution. Radiology 1996; 198:795–805.
12. Kurhanewicz J, Swanson M, Nelson SJ, Vigneron D. Magnetic resonance molecular imaging of prostate cancer. J Magn Reson Imaging 2002; 16:451–463.
13. Swanson MG, Vigneron DB, Tabatabai ZL, et al. Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted post-surgical prostate tissues. Magn Reson Med 2003; 50: 944–954.
14. American Urological Association. Prostate-specific antigen (PSA) best practice policy. Oncology (Huntingt) 2000; 14: 267–272, 277–268, 280 passim.
15. Kurhanewicz J, Vigneron DB, Males RG, Swanson MG, Yu KK, Hricak H. The prostate: MR imaging and spectroscopy—present and future. Radiol Clin North Am 2000; 38:115–138.
16. Star-Lack J, Nelson SJ, Kurhanewicz J, Huang LR, Vigneron DB. Improved water and lipid suppression for 3D PRESS CSI using RF band selective inversion with gradient dephasing (BASING). Magn Reson Med 1997; 38:311–321.
17. Tran TK, Vigneron DB, Sailasuta N, et al. Very selective suppression pulses for clinical MR spectroscopic imaging studies of brain and prostate cancer. Magn Reson Med 2000; 43:23–33.
18. Nelson SJ. Analysis of volume MRI and MR spectroscopic imaging data for the evaluation of patients with brain tumors. Magn Reson Med 2001; 46:228–239.
19. Costello LC, Franklin RB. Novel role of zinc in the regulation of prostate citrate metabolism and its implications in prostate cancer. Prostate 1998; 35:285–296.
20. Costello LC, Franklin RB. The intermediary metabolism of the prostate: a key to understanding the pathogenesis and progression of prostate malignancy. Oncology 2000; 59:269–282.
21. Kaji Y, Kurhanewicz J, Hricak H, et al.

Localizing prostate cancer in the presence of post-biopsy changes on MR images: role of proton MR spectroscopic imaging. Radiology 1998; 206:785–90.
22. Podo F, de Certaines JD. Magnetic resonance spectroscopy in cancer: phospholipid, neutral lipid and lipoprotein metabolism and function. Anticancer Res 1996; 16:1305–1315.
23. Ackerstaff E, Pflug BR, Nelson JB, Bhujwalla ZM. Detection of increased choline compounds with proton nuclear magnetic resonance spectroscopy subsequent to malignant transformation of human prostatic epithelial cells. Cancer Res 2001; 61:3599–3603.
24. van der Graaf M, Schipper RG, Oosterhof GO, Schalken JA, Verhofstad AA, Heerschap A. Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. MAGMA 2000; 10:153–159.
25. Swanson MG, Vigneron DB, Tran TK, Sailasuta N, Hurd RE, Kurhanewicz J. Single-voxel oversampled J-resolved spectroscopy of in vivo human prostate tissue. Magn Reson Med 2001; 45:973–980.
26. Males RG, Vigneron DB, Star-Lack J, et al. Clinical applications of BASING and spectral/spatial water and lipid suppression pulses for prostate cancer staging and localization by in vivo 3D ^{1}H magnetic resonance spectroscopic imaging. Magn Reson Med 2000; 43:17–22.
27. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174.
28. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148:839–843.
29. Pickett B, Vigneault E, Kurhanewicz J, Verhey L, Roach M. Static field intensity modulation to treat a dominant intraprostatic lesion to 90 Gy compared with seven field 3-dimensional radiotherapy. Int J Radiat Oncol Biol Phys 1999; 44:921–929.
30. Zaider M, Zelefsky MJ, Lee EK, et al. Treatment planning for prostate implants using magnetic resonance spectroscopy imaging. Int J Radiat Oncol Biol Phys 2000; 47:1085–1096.