Clare M. Tempany, MD • Xiao Zhou, PhD • Elias A. Zerhouni, MD • Matthew D. Rifkin, MD[2]
Leslie E. Quint, MD • Catherine W. Piccoli, MD • James H. Ellis, MD
Barbara J. McNeil, MD, PhD

Staging of Prostate Cancer: Results of Radiology Diagnostic Oncology Group Project Comparison of Three MR Imaging Techniques[1]

PURPOSE: To assess accuracy of three different magnetic resonance (MR) imaging techniques, including the endorectal coil, in staging prostate cancer.

MATERIALS AND METHODS: MR imaging was performed in 213 patients with prostate cancer with a conventional body coil, with fat suppression and a body coil, and with an endorectal coil. Radiologists identified tumor invasion into periprostatic tissues, neurovascular bundles, and seminal vesicles. Each technique was evaluated separately, and in a subset of 74 patients the three techniques were evaluated together. Images obtained with the two body-coil techniques were read in combination with images obtained with the endorectal coil (combination A) and alone (combination B).

RESULTS: Overall accuracy for conventional body-coil, fat-suppressed body-coil, and endorectal-coil MR was 61%, 64%, and 54%, respectively. Overall group accuracy for combinations A and B was 57% and 61%. Considerable interreader variability was found for combination A.

CONCLUSION: No technique was highly accurate for staging early prostate cancer. Individual radiologists did achieve a high degree of staging accuracy with the endorectal-coil and body-coil combination.

Index terms: Diagnostic radiology, observer performance, 844.121411, 844.121415 • Prostate, MR, 844.121411, 844.121415, 844.320 • Prostate, neoplasms, 844.32, 844.33, 99.33

PROSTATE carcinoma is the most common form of nonskin malignancy and the second leading cause of cancer-related death in American men. Approximately 165,000 new cases were diagnosed in 1993, and approximately 38,000 patients will die of prostate cancer this year (1). The prognosis is poor once the tumor has spread into the extracapsular periprostatic tissues and seminal vesicles. Thus, early diagnosis and treatment of prostate carcinoma, when the lesion is limited to the gland and potentially curable, is crucial.

In 1987, the Radiology Diagnostic Oncology Group (RDOG) multicollaborative study began comparing transrectal ultrasound (US) with conventional body coil magnetic resonance (MR) imaging for staging early prostate carcinoma. Results from the first 230 patients found an accuracy of 69% for MR imaging in distinguishing stage A and B lesions from stage C and D lesions (2). The results for transrectal US were lower (58%), but the difference was not statistically significant. Approximately 2 years after the study began, new MR imaging techniques and pulse sequences became available. The endorectal coil was the most important of these and has direct bearing on the goals of the study.

The prototype endorectal coils were first designed for imaging the prostate gland, and in two small studies the coil provided reliable, high-resolution images of the prostate gland and periprostatic region (3,4). After initial feasibility tests and alterations of the coil design were performed, testing prior to U.S. Food and Drug Administration (FDA) approval began. Initial data demonstrated increased spatial resolution and increased accuracy for detection of tumor invasion of the gland capsule, neurovascular bundles (NVBs), and seminal vesicles (5,6). In a series of 22 patients, the overall accuracy of endorectal MR imaging was 82% in differentiating stage B from stage C cancer (5).

The value of this technique remained to be defined and tested with a larger number of patients and radiologists, however, so the RDOG multicollaborative effort was redesigned with this intent (7). The main objective of the revised protocol was, as before, to determine the accuracy of MR imaging in staging prostate cancer. The added goals were to evaluate the new techniques (specifically, the endorectal probe and fat-suppression pulse sequence) in comparison with the conventional body-coil technique and to determine the incremental gain, if any, of all three techniques together compared with the combination of the two body-coil sequences.

MATERIALS AND METHODS

Patient Population and Clinical Evaluation

Initially, four institutions participated in the study: the Cleveland Clinic Foundation (Ohio), Johns Hopkins University Medical Institutions (Baltimore, Md), Thomas Jefferson University Hospital (Philadelphia, Pa), and the University of Michigan (Ann Arbor). One of the four institutions was unable to contribute patients to the study but did provide radiolo-

gists for the rereading study. The study dates were January 1990 to July 1992, and the entry criteria were the same as those in the first study (2). All patients underwent a physical examination by the referring urologist, and clinical information regarding the estimated location, size, and stage of the tumor was recorded. Patients were enrolled in the study only if they had biopsy-proved prostate carcinoma. On the basis of clinical findings, all patients were thought to have surgically resectable tumors (ie, stage A or B).

Patients were excluded if they had a cardiac pacemaker or had undergone pelvic surgery for colorectal or prostate cancer. Patients with a history of inflammatory bowel disease (Crohn or ulcerative colitis), external-beam radiation therapy to the pelvis, or severe hemorrhoids were excluded from the endorectal study in order to reduce the risk of irritation to the rectal mucosa. If the patient and clinical service agreed to treatment with radical prostatectomy, informed consent was obtained from the patients. All MR studies had to be performed within 5 weeks prior to surgery.

From 1990 to 1992, 213 (85%) of the total 252 patients were eligible for inclusion in the study. The total numbers of patients from each of the three participating institutions were 111, 73, and 29. The 39 ineligible patients included 37 for whom complete pathologic data were not available (this includes 11 patients who did not undergo surgery) and two patients for whom the time between imaging and surgery exceeded 5 weeks.

All patients underwent a prostate biopsy prior to entry into the study. The mean time from biopsy to MR imaging was 98 days (range, 9 days to 1.5 years). On the basis of the clinical information available for 244 of the 252 total patients, a palpable nodule was present in 160 (65%). An abnormal prostate-specific antigen level was found in 183 (75%); 90 (36%) had levels of 4–10 ng/mL, and the remaining 93 (39%) had levels greater than 10 ng/mL. Normal levels were found in 50 patients (20%), and in the remaining 11 (5%) the levels were not recorded.

Pathologic examination of specimens in these 213 patients revealed 91 patients (43%) had localized disease and 122 (57%) had advanced disease, which included six patients with metastatic lymph node disease. These six patients underwent radical prostatectomy because the metastatic node disease was not diagnosed at the time of frozen-section analysis. In the group of 91 patients with localized disease, 18 had stage A lesions and 73 had stage B lesions. In the group of 122 patients with advanced disease, periprostatic invasion was found in all 122 patients (100%), and 16 patients (13%) had seminal vesicle invasion. Of the 122 patients with periprostatic invasion, 52 (43%) had NVB invasion.

MR Imaging

All MR examinations were performed on a 1.5-T whole-body imager (GE Medical Systems, Milwaukee, Wis). The goal was to perform all three techniques in every patient. The endorectal coil was used first whenever possible because this coil is relatively invasive and patients were more compliant if it was used first rather than at the end of a long study (the two body-coil sets). All eligible patients were required to be examined with at least two of the three techniques.

The endorectal coil (Medrad, Pittsburgh, Pa) is a receive-only coil; that is, it does not transmit radio-frequency waves. The whole-body coil acts as the radio-frequency transmitter. The coil is inside an inflatable latex balloon, which is positioned in the rectum. The balloon is inflated with 50–70 cm³ of air, which places the coil immediately posterior to the prostate gland.

For each patient, an initial sagittal body-coil localizer sequence was performed to ensure that the probe was in the correct position. The protocol used with the endorectal probe was as follows: conventional spin-echo (SE) T1-weighted axial imaging with repetition time (TR) of 600 msec, minimum echo time (TE), 256 × 128 matrix, two signals averaged, 4-mm section thickness, and 1-mm gap. The conventional SE T2-weighted sequence involved 2,500–3,000/20–30, 80 (TR msec/TE msec); 256 × 128 matrix; two signals averaged; 4-mm section thickness; and 1-mm gap. Both used a field of view of 10–12 cm. During axial imaging with the endorectal probe, the phase-encoding gradient was in the right-to-left direction to prevent phase-encoding artifacts from being superimposed on the prostate gland. We also routinely administered glucagon (1 mg, intramuscular) prior to imaging, to reduce peristalsis.

The conventional body-coil imaging protocol has been previously described and was not altered for the patients in this study (2). These axial conventional SE images were obtained with the following parameters: T2 weighted with 2,500–2,800/20–60, 80–120; 256 × 128 matrix; two signals averaged; 5-mm section thickness (interleaved with no gap); and 24–28-cm field of view; T1 weighted with 600/20, 256 × 256 matrix, two signals averaged, and 5-mm section thickness, and 1.5-mm gap.

The body-coil fat-suppression sequence was either a standard T2-weighted frequency-selective presaturation sequence (ChemSat, GE Medical Systems), as available on 1.5-T Signa units, or a hybrid sequence (8,9). These fat-suppression techniques were not always exactly the same, but the differences were judged to be minor. The parameters were 2,500–2,800/20, 80, 256 × 128 matrix, two signals averaged, and 5-mm section thickness, interleaved. These were also performed at a 24–28-cm field of view in the axial plane. The imaging volume was centered at the middle of the gland and was large enough to cover it entirely.

The protocol thus included axial body-coil conventional SE T1-weighted and T2-weighted sequences, axial body-coil conventional SE T2-weighted sequences with fat suppression, and axial endorectal-coil conventional SE T1-weighted and T2-weighted sequences. Because of the length of the study (almost 2 hours), the images were all obtained in the axial plane. All images were filmed from superior to inferior, and the endorectal-coil images were filmed without any postprocessing. The window width and level were set for each image to reduce the near-field high signal intensity, which can result from the proximity of the coil to the prostate.

Image Analysis

Images obtained with each of the three techniques on each patient were separated and read by one of three radiologists at the institution where the patient was treated. Each radiologist knew the patient had a clinical diagnosis of early-stage prostate cancer but did not know the results of the other MR imaging techniques. Neither did the radiologist know any other details, such as the prostate-specific antigen level in blood serum or biopsy results (location or Gleason score). Cases were randomly assigned to the readers; each radiologist read an equal number of cases obtained with each technique.

Subsequently, 75 cases in which images were obtained with all three techniques were reread several months after the end of the study (over one weekend) in one of two combinations: combination A, images obtained with all three techniques; and combination B, images obtained with the two body-coil techniques excluding those obtained with the endorectal probe. Nine radiologists, two from each of the four participating institutions and one from an external institution, reread the 75 cases. The two radiologists from each institution were randomly assigned to read either combination A or B for each case. The external radiologist was one of the original investigators who was involved with the endorectal coil design and was perceived by the group to have a wide range of experience in MR imaging with the endorectal coil. For this reason, this radiologist read only the combination A set of images. This reader, therefore, read twice as many combination A sets as the other readers.

In all image interpretations, the following features were analyzed by the radiologists: lesion identification and its axial and sagittal location. The size of the lesion, its relative signal intensities, and the margin distinction of each lesion were recorded. The degree of suspicion of malignancy for each lesion was scored on a five-point scale for receiver operating characteristic (ROC) analysis. The scores were 0, definitely or almost definitely benign; 1, probably benign; 2, possibly malignant; 3, probably malignant; and 4, definitely or almost definitely malignant.

The right and left NVBs were identified. Suspicion of invasion through the capsule (defined as extension of tumor beyond the capsule on T1- and T2-weighted images) into the NVBs, seminal vesicles, or both, was recorded and scored on a five-point

scale. The criteria used to identify invasion of the tumor through the capsule or NVBs were (a) direct tumor extension beyond the prostate, (b) decreased signal intensity on the T1- and T2-weighted images in the periprostatic fat adjacent to the capsule nearest the tumor, and (c) focal contour bulge of the capsule nearest the tumor (5,6,10). The same scores were used for periprostatic invasion and NVB invasion: 0, definitely or almost definitely not invaded; 1, probably not invaded; 2, possibly invaded; 3, probably invaded; and 4, definitely or almost definitely invaded. The presence or absence of abnormal lymph nodes was graded in a similar fashion.

Surgical Procedure

Within 3 weeks of imaging, 201 (94%) patients underwent a radical retropubic prostatectomy. Eight patients underwent surgery within 4 weeks, and another four underwent surgery within 5 weeks of imaging. Surgery was usually performed after the lymph nodes had been sampled and examined for metastases at frozen-section analysis (11). If nodal metastases were found at this time, the radical prostatectomy was not performed. Details of the prostatectomy and the findings at surgery were recorded by the surgeon for each patient. For this study, specific attention was paid to the status of the NVBs at surgery and whether they were resected.

Pathologic Analysis

All specimens were uniformly prepared with fixation in formalin. They were then coated with India ink and sectioned in the axial plane at intervals of 5 mm (whole-mount sections) or 2–3 mm. The entire gland was sectioned and methodically examined in all cases. Sections were designated so that the location of each lesion within the prostate could be accurately noted.

In each case, the pathologist recorded the size of all cancers larger than 5 mm and the location in both the axial and sagittal plane. In all cases, the primary, secondary, and combined (sum of the primary and secondary grades) Gleason grades of the tumor were determined. The pathologist noted the presence of any capsular penetration (defined as extension of tumor into the periprostatic soft tissue) and its location. The maximal depth of any invasion was measured in millimeters. The presence or absence of NVBs in the resected specimen was recorded. The presence of seminal vesicle invasion (defined as tumor extension into the muscle wall of the seminal vesicle) and invasion in any metastatic lymph nodes was recorded. The final staging did not differentiate microscopic from macroscopic invasion of the capsule.

Statistical Methods

Data from pathologic studies were used to separate patients into two groups: those with periprostatic invasion of tumor and spread of disease to the seminal vesicles and lymph nodes and those with stages A and B disease. The data analysis was performed on the two data sets, the initial prospective readings for the three techniques, and the rereading for combinations A and B.

Diagnostic sensitivity, specificity, and accuracy of the three techniques were initially evaluated by using pooled data from all institutions, and then the data were further evaluated on the basis of hospital and of reader. The ROC scale ratings were grouped for the accuracy calculations: All probably abnormal and definitely abnormal findings (scores of 3 and 4) were considered positive. The remaining three scores of 0, 1, and 2 (possibly abnormal, probably normal, and normal or almost definitely normal) were considered negative. Comparisons of the three techniques were made by using the McNemar test. A pair-wise analysis compared each set of sequences in pairs for the patients who underwent imaging with both sequences of the pair. To compare diagnostic accuracy in detecting invasion of the prostatic capsule with each of the three techniques, empirical ROC curves were made for each reader who read at least seven cases. The areas under the ROC curves and their corresponding standard errors were computed (12).

For the rereading data, the diagnostic sensitivity, specificity, and accuracy were initially evaluated by using pooled data from all institutions and then for individual reader. Empirical ROC curves for detection of periprostatic invasion and invasion of the NVBs for combinations A and B were plotted for each reader. Nonparametric estimates of the associated areas and the errors were computed (12). Smooth ROC curves were calculated by using computer programs developed by Metz and his collegues at the University of Chicago (13).

RESULTS

The combined techniques rereading group included only 75 of the original 213 patients because each patient had to undergo imaging with all three techniques and have complete pathologic analysis data available at the time of the rereading. We report the results for the combination of techniques first and then the individual techniques, with the overall staging results followed by ROC analysis for periprostatic invasion and NVB invasion. Because seminal vesicle invasion and lymph node metastases were so uncommon in this patient population, they were not evaluated in the rereading data set, but the results of seminal vesicle invasion are given for each technique. In each group, the results are given for the overall group and the individual readers. The number of patients in each technique group differs because not all patients were imaged with all three techniques; only 75 had all three types of images at the time of the rereading.

Staging Accuracy

For the pooled group (excluding the external radiologist), the mean sensitivity, specificity, and accuracy for readings of combination A (all three techniques combined) and combination B (two body-coil sequences alone) showed no overall difference between the two (Table 1).

Sensitivities and specificities for each radiologist for each combination are shown in Table 2. The highest accuracy (79%), sensitivity (83%), and specificity (67%) were achieved by reader 5 with combination A, which included the endorectal coil. This reader was from the institution that contributed the largest number of cases to the study ($n = 111$). The results for the external radiologist were the second highest, with accuracy of 69%, sensitivity of 78%, and specificity of 55% with combination A. In general, readers tended to undercall the presence of advanced disease; all readers had greater specificity than sensitivity with combination B, but with combination A three readers had greater sensitivity (83%–59%) than specificity and two of these readers achieved the highest accuracy.

The mean staging results of the group for the conventional body-coil, fat-suppression, and endorectal-coil images, when read separately and at the patient's institution, were 62%, 63% and 52%, respectively (Table 3). Among the 116 patients with advanced disease who underwent conventional MR, 71 diagnoses were correctly identified for a sensitivity of 61%. Among the 85 patients with localized disease, conventional MR correctly identified 53, yielding a specificity of 62%. Among the 106 patients with advanced disease, who underwent fat-suppression MR imaging, diagnoses in 68 were correctly identified, giving a sensitivity of 64%. Among the 80 patients with localized disease, fat-suppression MR correctly identified disease in 49, giving a specificity of 61%. Among the 106 patients with advanced cancer who underwent endorectal-coil MR, the study correctly identified disease in 63, giving a sensitivity of 60%. Among the 78 patients with localized cancers, endorectal-coil MR correctly depicted 33, giving a specificity of 42%. The readings for

the individual techniques did not show the same low sensitivities or degree of undercalling of advanced disease as was shown for the combination-technique rereadings.

The individual readings of all cases, pooled for all readers, yielded no statistically significant differences in the sensitivities and specificities of conventional, fat-suppressed, and endorectal-coil MR imaging. No statistically significant differences were found for any of the pairs. The staging accuracies for each of the three techniques varied among the readers, however, with ranges for conventional body coil of 50%–73%, for fat suppression of 52%–73%, and for the endorectal coil of 38%–63%. Sensitivities and specificities were calculated for each individual reader and for each individual pulse sequence, but the small numbers make formal statistical estimates unreliable.

Periprostatic Invasion

Of the 213 total patients, 122 had periprostatic invasion at pathologic examination. The ROC curves for all radiologists pooled for combination A yielded an area of 0.61 (standard error = 0.12) and for combination B an area of 0.63 (standard error = 0.13). For the four institutions separately, the highest area under the curve was 0.72 for combination A and 0.71 for combination B (not significant). The empirical ROC areas by reader for combinations A and B varied enormously (Table 4). For combination A, the areas under the empirical curves ranged from 0.80 to 0.49; six radiologists, including the external radiologist, had areas above 60%. For combination B, areas ranged from 0.79 to 0.57, and again six radiologists had areas above 60%. The variability among readers was greater for combination A than combination B (Fig 1, Table 4). Because of the high standard errors for each of the areas (due to small sample size), the differences among readers were not statistically significant. However, reader 5 and the external reader, both of whom had considerable experience with endorectal-coil imaging, had the highest ROC areas for combination A (0.80 and 0.74).

The associated fitted ROC curves for the pooled data from all three institutions for individual techniques had areas of 0.70, 0.71, and 0.48 for conventional, fat suppression, and endorectal coil, respectively (Fig 2). Overall accuracies for detection of periprostatic invasion for each tech-

nique, as read separately at the three institutions, varied slightly. For conventional body-coil MR imaging, accuracy scores were 59%, 61%, and 63%; for fat suppression, they were 56%, 69%, and 62%; and for endorectal coil, they were 45%, 54%, and 33%. For patients who underwent both conventional and fat suppression imaging (173 patients), the two modalities were not statistically significantly different for detection of periprostatic invasion. The difference was, however, statistically significant between conventional and endorectal-coil MR imaging (164 patients) and between fat-suppression and endorectal-coil MR imaging (153 patients) in detec-

tion of periprostatic invasion. The conventional body coil provided more accurate results than the endorectal coil ($P = .02$), as did fat suppression ($P = .005$).

NVB Invasion

At pathologic examination, specimens from 122 patients showed periprostatic invasion, and 52 (43%) of these patients had NVB invasion. The pooled rereading accuracy for both combination A and B yielded similar areas under the ROC curve: 0.56 and 0.57 for combination A and B, respectively, both with a standard error of 0.13. When the institutional ROC

Table 1
Pooled Data at Rereading for Eight Radiologists in Staging Early Prostate Cancer

Imaging Combination*	Sensitivity (%)	Specificity (%)	Accuracy (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
A	51	67	57	71	46
B	47	82	61	81	49

* Combination A is all three pulse sequences and combination B is two body-coil sequences as described in Materials and Methods; $n = 75$ cases for each combination.

Table 2
Staging Results for MR Imaging Combination A and Combination B for Each Radiologist

Reader	Sensitivity (%) A	Sensitivity (%) B	Specificity (%) A	Specificity (%) B	Accuracy (%) A	Accuracy (%) B
1	33	27	100	90	52	57
2	55	63	70	67	63	64
3	50	36	56	80	52	57
4	45	58	75	89	60	67
5	83	45	67	75	79	60
6	50	50	50	89	50	61
7	33	32	100	90	52	60
8	59	63	45	67	52	64
External*	78	na	55	na	69	na

Note.—Each reader number represents the same radiologist in Tables 2 and 4.
* The external reader only read combination A images. na = not available.

Table 3
Pooled Staging Results for Images Obtained with Each MR Imaging Technique Read Prospectively at Patient's Institution

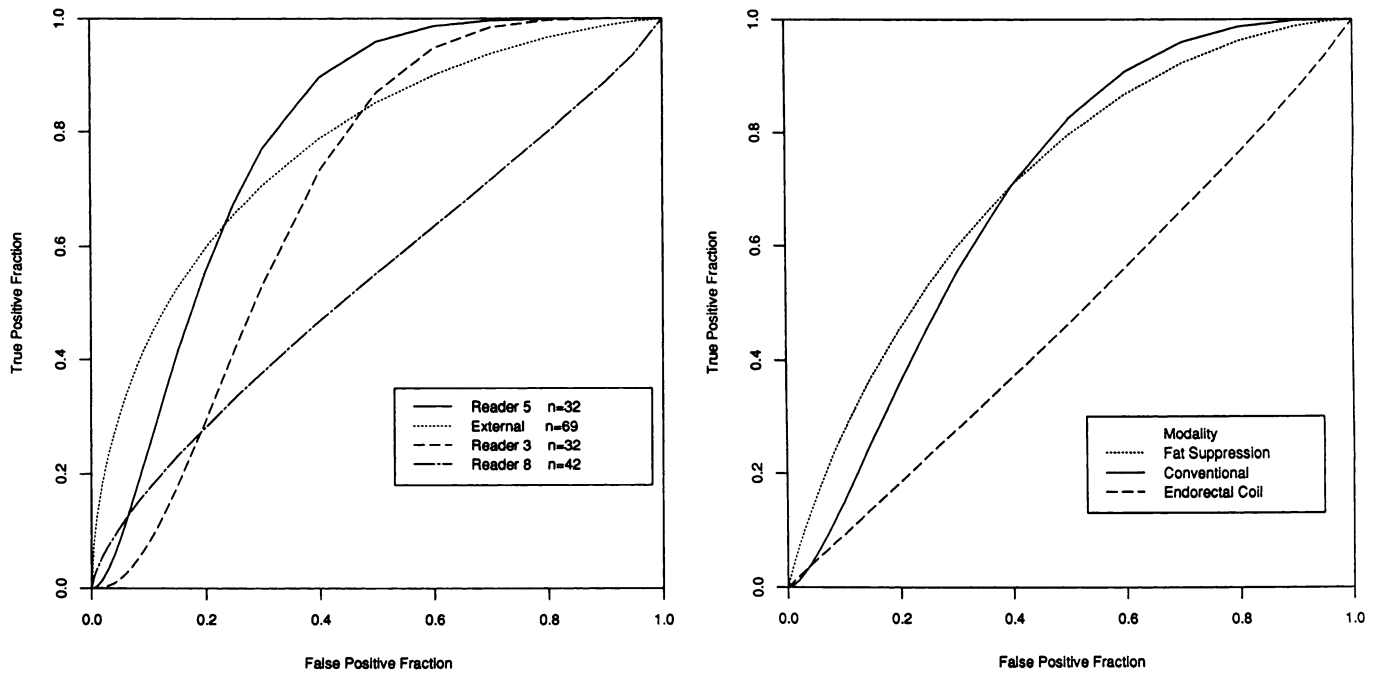| Technique	Sensitivity (%)	95% CI	Specificity (%)	95% CI	Patients with Advanced Disease (%)	Accuracy (%)
Conventional body coil ($n = 201$)	61	0.5–0.7	62	0.5–0.8	58	62
Fat-suppression body coil ($n = 186$)	64	0.6–0.8	61	0.5–0.8	57	63
Endorectal coil ($n = 183$)	60	0.5–0.7	42	0.3–0.6	57	52

Note.—Numbers of patients for each technique varies because not all patients underwent all techniques. CI = confidence interval.

Figures 1, 2. (1) Fitted ROC curves for combination A (conventional body-coil, fat-suppression, and endorectal-coil MR imaging) for four readers. Graph represents the interreader variability for detection of periprostatic invasion in 75 cases. (2) Fitted ROC curves for periprostatic invasion with each technique as read prospectively in each patient's institution. The readings from all three institutions are included.

Table 4
Empirical ROC Curve Areas for Each Radiologist for Detection of Periprostatic Invasion

Reader	Combination A	Combination B
1	0.65 (0.18)	0.58 (0.09)
2	0.60 (0.10)	0.64 (0.20)
3	0.62 (0.20)	0.57 (0.09)
4	0.65 (0.10)	0.79 (0.19)
5	0.80 (0.18)	0.62 (0.10)
6	0.49 (0.10)	0.73 (0.20)
7	0.67 (0.18)	0.61 (0.09)
8	0.55 (0.11)	0.70 (0.25)
External	0.74 (0.08)	

Note.—Numbers in parentheses are standard errors.

curves were assessed, the highest area among the four institutions was 0.79 for combination A and 0.88 for combination B. The areas for the two combinations with the two-sided tests were not statistically significantly different.

The overall accuracies for detecting NVB invasion were 65%, 70%, and 61% for conventional, fat-suppression, and endorectal-coil technique, respectively. The comparisons for the pairs of sequences showed no statistically significant differences. The ROC curves for conventional, fat-suppression and endorectal-coil MR imaging produced areas of 0.55 (standard error = 0.08), 0.70 (standard error = 0.07), and 0.49 (standard error = 0.09), respectively, when read alone; these

are not statistically significantly different.

Seminal Vesicle Invasion

Of the 213 patients, specimens from 16 had pathologic evidence of tumor invasion into the seminal vesicles. This number is too small for ROC analysis, but the three techniques produced high specificities: 85% for conventional, 91% for fat-suppression, and 85% for endorectal-coil MR imaging. All three also yielded similarly low sensitivities in detection of invasion: 21% for conventional ($n = 196$), 29% for fat-suppression ($n = 180$), and 21% for endorectal-coil ($n = 177$) MR imaging. Because few patients had lymph node metastases, these data were not analyzed for this study.

DISCUSSION

The need for improved staging methods for prostate cancer is again evident from this study. The preoperative clinical staging system that was used in these patients caused understaging in many cases of prostate cancer. Of the 213 patients in this study, 122 (57%) had advanced disease at pathologic examination and 91 (43%) had localized disease; all had been thought clinically to have localized disease. This clinical understaging was found in a similar patient

population in the initial RDOG study and by others (2,14).

By late 1989, two important facts had become apparent and resulted in this revised protocol. First, technical developments had occurred that had direct bearing on MR imaging of the prostate and the goals of this study. Namely, the endorectal probe was undergoing initial testing for FDA approval, and preliminary results supported its importance in prostate cancer staging (15). Second, the results of the initial prostate cancer RDOG staging study were somewhat disappointing (2), and rather than continue the study for an additional 2–3 years, a change was indicated. Continuing the existing protocol was expected to yield similar results, a fact that was borne out in this study with the conventional body-coil sequence. The protocol was thus revised to include the endorectal-coil and fat-suppression techniques.

Our results remain disappointing, however, and show that none of the three techniques is highly accurate for staging of early prostate cancer. The overall results show no apparent improvement compared with the previous report from this group. The group accuracy for staging with all three techniques combined was 57% and with the two body-coil techniques alone was 61%. The accuracy for each technique alone was 52%–63%, which is not statistically significantly differ-

ent from the originally reported results (2).

Some of the individual radiologist results with combination A are more in keeping with the expected improvement. The best results show reasonably high accuracy rates (69%–79%) for staging early prostate cancer. However, the results vary considerably across the nine readers in our study. Many radiologists undercalled advanced disease when reading images obtained with the techniques in combination. This led to the decreased sensitivity and accuracy rates. This degree of undercalling advanced disease was not true of the two radiologists with the best results.

Several reasons may explain the wide interreader variability, including the experience of the individual radiologists. Several differences are evident among the readers. As can be seen from the case-accrual data, the number of cases from the three institutions differ greatly, ranging from 111 to 29. Two of the radiologists involved in the rereading study came from the fourth institution, which accrued no cases. Two readers were from the institution contributing the largest number ($n = 111$) of cases to the study. The first reader, active in the RDOG group since its beginning, had the best results and highest ROC area (ie, 0.79) for the detection of periprostatic invasion. The second reader, with only 1 year of experience in MR imaging after residency, had the lowest area (ie, 0.49). The external reader, who was involved in the development of the endorectal probe, had over 5 years of experience with the probe and had the second highest accuracy with combination A (69%) and second largest area (ie, 0.74) under the ROC curve for the detection of periprostatic invasion.

These results suggest the obvious—experience may improve accuracy, but even a 1-year fellowship in MR imaging is not enough to reach the top performance. Further experimental work to understand the mechanism by which experience makes an impact is necessary. Such a study will be difficult to perform, however, because of the logistic difficulties of having many readers with different backgrounds read enough cases to reach statistical significance. In addition, the timing of this assessment is critical. It should be early enough to make a difference in the diffusion of a new technology but not so early that readers have not gained enough experience to read the images thoughtfully and accurately. Whether this timing

differs for full-time readers of a given modality versus part-time readers is unclear, but it likely does. Whether the timing varies with the extent of change in the modality (eg, the addition of fat suppression vs introduction of the probe) is also unclear.

Several other reasons help explain why results for the three imaging approaches differ. Differences in technical factors and appearances of the images are much greater for the endorectal-coil sequence than for the two body-coil sequences. For example, both body-coil sequences are performed with the same fields of view and parameters such as section thickness. In contrast, the endorectal-coil images are obtained at smaller fields of view with increased spatial resolution and are more technically demanding and require more operator input (eg, careful positioning of the coil and its localization) to obtain high-quality images. In many ways, the endorectal images are as operator-dependent as transrectal US techniques. Their interpretation is also more demanding and several pitfalls can occur. Thus, inexperience may have led to difficulty in interpreting the findings on the endorectal-coil images, particularly at the capsule. The MR features used in this study were limited to identification of a suspicious lesion, its size, and its location. Then the likelihood of invasion by each lesion was based on the features described in Materials and Methods. The analysis of the capsule and periprostatic fat was not very detailed in this study, and a more detailed feature analysis (such as the shape or integrity of the capsule, NVBs, and seminal vesicles) might result in greater accuracy.

A phase-encoding artifact horizontal to the plane of the coil can cause difficulties in interpretation, especially when trying to determine extracapsular spread of tumor. We routinely used glucagon to suppress peristaltic motion, but we did not use other medications such as diazepam (Valium; Hoffman-LaRoche, Nutley, NJ), which can reduce motion artifact further. Glucagon is more acceptable medically because many patients were imaged just before surgery. Other interpretation problems, such as interference from hemorrhage and scar from prior biopsy, are possibly more frequent due to the higher spatial resolution of the coil. The effects of hemorrhage and biopsy are difficult to quantify and cannot be identified at pathologic analysis (Epstein JI, oral communication, 1990). These ef-

fects, however, are visible on MR images and no doubt cause problems in interpretation. The average time from biopsy to imaging in this patient population was over 3 months, which is probably longer than in routine clinical practice. Since imaging of patients prior to biopsy is impractical, these interferences will routinely be present regardless of when the imaging is performed, unless it is performed well beyond the usual clinical time frame. It seems reasonable, however, to delay MR imaging as long as practically possible to avoid interference on the images.

Because of the sensitivity profile of the endorectal coil, interpretation may be difficult if the images are incorrectly filmed. Near-field brightness due to the proximity of the coil to the capsule and posterior prostate gland (and hence filming difficulties) can occur. No specific filming protocol was used in this study, but all images were subject to quality control, and poor quality (including poorly processed) cases were rejected from the study. No correction methods were used to adjust for the sensitivity profile of the coil in this study. These methods may be useful in reducing artifacts and improving image quality.

Our results varied when the images were read together (combination A or B) versus separately (Tables 1, 4). The combined readings were more controlled in that all readers read the same 75 cases, whereas the separate readings involved reading different cases prospectively on site at the patient's home institution. Nonetheless, some explanations may be possible. The combination readings were less accurate than the individual readings for the group as a whole. The endorectal coil did not appear to provide any increase in accuracy for most readers. For two readers, the endorectal coil helped in the combination setting; one showed an improvement from 54% accuracy for the coil alone to 78% when used in combination with the two body-coil techniques.

For the combination readings, the radiologists had been instructed to give their overall best judgments on the basis of the three (combination A) or two (combination B) sequences. This requirement to use more than one piece of data may have caused readers to become more stringent in their interpretation criteria, that is, to diagnose spread of disease less often. Thus, sensitivity was lower and specificity higher. This discrepancy between the sensitivity and specificity

results did not occur when the same readers were reading the single techniques prospectively at their own institutions.

In three recent preliminary studies based on endorectal coil imaging alone, the staging accuracy rates ranged from 68% to 83% for 47–120 patients (16–18). All three studies are as yet published only as abstracts. The results from two of these were not as high as those originally reported by the same group (5,16,18). These two studies suggested that accuracy rates are higher when stage C disease is subdivided into microscopic and macroscopic invasion groups (16,18). None of these studies evaluated interreader variability, and all were performed at single institutions.

The staging results in this study were based on the designation of intraglandular or extraglandular tumor at pathologic analysis. We did not, therefore, separate the patients according to the depth of tumor penetration into or through the prostatic capsule (ie, microscopic or macroscopic invasion). Recent reports from the surgery literature support the value of this separation (19). The presence of periprostatic invasion in general terms is thought to be a contraindication to radical surgery. In fact, the depth of tumor invasion into or through the capsule may alter the patient's prognosis. There does appear to be a difference between patients with microscopic (several millimeters) invasion and those with frank macroscopic (several centimeters) invasion. Rosen et al (19) found extracapsular penetration in 62 (63%) of their patients with stage B disease, and positive surgical margins were found in 23 (23%). In fact, as recently noted by Schiebler et al (20), many cases of subtle capsular penetration were previously classified as stage B disease, and microscopic stage C disease has been underreported in the literature. The patients with so-called limited stage C disease can be successfully treated with surgery (21) and adjuvant therapy (22).

Thus, it appears that as imaging, surgical, and pathologic data are further reviewed, it may be inadequate to simply determine that the patient has extraglandular tumor spread. The depth of invasion becomes more important when evaluating the NVBs and planning the surgical approach (ie, nerve sparing or not). Penetration of tumor to the NVBs at 1 mm or greater has been used to define pathologic invasion of the NVBs (10).

As the spatial resolution of MR imaging continues to improve, shallower penetration should become detectable and measurable. Probably no imaging modality, however, will be able to depict microscopic invasion reliably.

Since 1990, when this study was designed, more improvements have been made in MR imaging technology. The most significant improvement is the fast SE sequence, which considerably reduces the acquisition time for T2-weighted imaging. This development not only reduces motion artifacts but also provides the option to obtain additional images in alternate planes. In this study, we assessed the prostate gland only in the axial plane for two reasons: to maintain consistency for technique comparison and to keep the imaging time (over 2 hours per patient) within practical limits. The coronal and sagittal planes can add information about the apex of the gland and especially the seminal vesicles. Further improvements in spatial resolution and greater spatial coverage of the anterior prostate gland have been recently described with combination of the endorectal coil and phased-array coil (23).

In this study we assessed the accuracy of MR imaging alone. We deliberately did not allow the radiologists to have any knowledge of the patient's health, other than the diagnosis of cancer, prior to reading the images. If we included clinical findings and the prostate-specific antigen measurement, the results might have been better. This availability of data is recommended in a clinical setting.

In summary, the results of this study show that none of the three MR imaging techniques tested is highly accurate for staging early prostate cancer. The differences in overall staging accuracy among the three techniques were not statistically significant, and these results are not better than those originally reported (2). One reader did achieve accuracy as high as 79% for staging prostate cancer when the endorectal coil was used in combination with the other techniques, but considerable interreader variability was evident.

The results continue to suggest a need to improve the current staging systems, both clinical and image based. Several methods that may help improve overall staging accuracy include the combination of image and nonimage information to enhance image interpretation and the use of fast SE, multiple imaging planes, and a combination of the endorectal coil and the phased-array techniques. ∎

References

1. Boring CC, Squires TS, Tong T, Montgomery S. Cancer statistics. CA 1994; 44:7–26.
2. Rifkin MD, Zerhouni EA, Gatsonis CA, et al. Comparison of magnetic resonance imaging and ultrasound in staging early prostate cancer: results of a multi-institutional cooperative trial. N Engl J Med 1990; 323:621–626.
3. Martin JF, Hajek P, Baker L, Gylys-Morin V, Fitzmorris-Glass R, Mattrey RR. Inflatable surface coil for MR imaging of the prostate. Radiology 1988; 167:268–270.
4. Schnall M, Lenkinski RL, Pollack HP, Imai Y, Kressel HY. Prostate: MR imaging with an endorectal surface coil. Radiology 1989; 172:570–574.
5. Schnall MD, Imai Y, Tomaszewski J, Polack H, Lenkinski R, Kressel HY. Prostate cancer: local staging with endorectal surface coil MR imaging. Radiology 1991; 178:797–802.
6. Tempany CMC, Zerhouni EA, Epstein JI, Walsh PC. Neurovascular bundle identification and invasion by prostate cancer: comparison of MR imaging with conventional body and endorectal coil techniques (abstr). Radiology 1990; 177(P):266.
7. Gatsonis C, McNeil BJ. Collaborative evaluations of diagnostic tests: experience of the Radiology Diagnostic Oncology Group. Radiology 1990; 175:571–575.
8. Vinitski S, Mitchell DG, Szumowski J, Burke DL, Rifkin MD. Variable flip angle imaging and fat suppression in combined gradient and spin echo (GREASE) techniques. Magn Reson Imaging 1990; 8:131–139.
9. Szumowski J, Eisen JK, Vinitski S, Haake PW, Plewes DB. Hybrid methods of chemical shift imaging. Magn Reson Med 1989; 9:379–388.
10. Tempany CMC, Rahmouni AD, Epstein JI, Walsh PC, Zerhouni EA. Invasion of the neurovascular bundle by prostate cancer: evaluation with MR imaging. Radiology 1991; 181:107–112.
11. Walsh PC. Radical retropubic prostatectomy. In: Walsh PC, Gittes RF, Perlmutter AD, Stanley TA, eds. Cambell's urology. 5th ed. Philadelphia, Pa: Saunders, 1986.
12. Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. Biometrics 1988; 44:837–845.
13. Metz CE, Kronman HB, Shen JE, Wang PL. CORROC2: a program for ROC analysis of correlated, inherently categorical, rating-scale data. Chicago, Ill: Department of Ra-

diology and the Franklin McLean Research Institution, University of Chicago, June 1989.

14. Catalona WJ, Stein AJ. Staging errors in clinically localized prostate cancer. J Urol 1982; 127:452–456.

15. Schnall MD, Imai Y, Pollack HM, Lenkinski RE, Kressel HY. Local staging of prostate carcinoma with endorectal surface MR imaging (abstr). Radiology 1989; 173(P):142.

16. Cheung LP, Schnall MD, Chelsky MJ, Harkaway RC, Wein AJ, Kressel HY. Current clinical utility of endorectal surface coil MR imaging of prostatic carcinoma (abstr). Radiology 1992; 185(P):275.

17. Krebs TL, Silverman JM. Clinical utility of endorectal surface coil MR imaging of the prostate gland (abstr). Radiology 1992; 185(P):275.

18. Chelsky MJ, Schnall MD, Seidmon EJ, Pollack HM. Use of endorectal surface coil for local staging of prostate cancer (abstr). J Urol 1993; (p):287A.

19. Rosen MA, Goldstone L, Lapin S, Wheeler T, Scardino PT. Frequency and location of extracapsular extension and positive surgical margins in radical prostatectomy specimens. J Urol 1992; 148:331–337.

20. Schiebler ML, Schnall MD, Pollack HM, et al. Current role of MR imaging in the staging of adenocarcinoma of the prostate. Radiology 1993; 189:339–352.

21. Walsh PC, Epstein JI, Lowe FC. Potency following radical prostatectomy with wide unilateral excision of the neurovascular bundle. J Urol 1987; 138:823–827.

22. Zincke H, Utz D, Taylor WF. Bilateral pelvic lymphadenectomy and radical prostatectomy for clinical stage C prostate cancer: role of adjuvant treatment for residual cancer and in disease progression. J Urol 1986; 135:1199–1205.

23. Schnall MD, Connick T, Hayes CE, Lenkinski RE, Kressel HY. MR imaging of the pelvis with an endorectal-external multi-coil array. JMRI 1992; 2:229–232.