Comparison of CT and MR Imaging in Staging of Neck Metastases[1]

Hugh D. Curtin, MD
Hemant Ishwaran, PhD
Anthony A. Mancuso, MD
Robert W. Dalley, MD
Daryl J. Caudry, MS
Barbara J. McNeil, MD, PhD

Index terms:
Computed tomography (CT), comparative studies
Head and neck neoplasms, CT, 276.12112, 99.12912
Head and neck neoplasms, MR, 276.12143, 99.12943
Lymphatic system, neoplasms, 997.33
Magnetic resonance (MR), comparative studies

Radiology 1998; 207:123–130

Abbreviations:
NPV = negative predictive value
PPV = positive predictive value
ROC = receiver operating characteristic

[1] From the Department of Radiology, Massachusetts Eye and Ear Infirmary, 243 Charles St, Boston, MA 02114 (H.D.C.); the Departments of Radiology (H.D.C., B.J.M.) and Health Care Policy (D.J.C., B.J.M.), Harvard Medical School, Boston, Mass; the Department of Mathematics and Statistics, University of Ottawa, Ontario, Canada (H.I.); the Department of Radiology, Shands Hospital and University of Florida College of Medicine, Gainesville (A.A.M.); the Department of Radiology, University of Washington Medical Center, Seattle (R.W.D.); and the Department of Radiology, Brigham and Women's Hospital, Boston, Mass (B.J.M.). From the 1995 RSNA scientific assembly. Received August 7, 1997; revision requested October 23; revision received November 26; accepted December 10. Supported in part by National Cancer Institute and National Institutes of Health grant UO1CA54019. **Address reprint requests to** H.D.C.

Author contributions:
Guarantors of integrity of entire study, H.D.C., H.I., A.A.M., R.W.D., B.J.M.; study concepts, H.D.C., A.A.M., R.W.D., B.J.M.; study design, H.D.C., H.I., A.A.M., R.W.D., B.J.M.; definition of intellectual content, H.D.C., H.I., A.A.M., R.W.D., D.J.C., B.J.M.; literature research, H.D.C.; clinical studies, H.D.C., A.A.M., R.W.D.; data acquisition, H.D.C., A.A.M., R.W.D.; data and statistical analysis, H.I., D.J.C.; manuscript preparation, H.D.C., D.J.C.; manuscript editing, H.D.C., H.I., D.J.C., B.J.M.; manuscript review, H.D.C., H.I., A.A.M., R.W.D., D.J.C., B.J.M.

PURPOSE: To compare the abilities of magnetic resonance (MR) imaging and computed tomography (CT) in detection of lymph node metastasis from head and neck squamous cell carcinoma.

MATERIALS AND METHODS: MR imaging and CT were performed with standard protocols in patients with known carcinoma of the oral cavity, oropharynx, hypopharynx, or larynx. Histopathologic examination was performed to validate imaging findings. Between 1991 and 1994, 213 patients undergoing 311 neck dissections were accrued at three institutions.

RESULTS: For the upper jugular and spinal accessory regions, the areas under the receiver operating characteristic curves for combined information on size and internal abnormality were 0.80 for CT and 0.75 for MR imaging. Sensitivities, specificities, negative predictive values (NPVs), and positive predictive values (PPVs) were calculated for various size criteria with and without internal abnormality information. With use of a 1-cm size or an internal abnormality to indicate a positive node, CT had an NPV of 84% and a PPV of 50%, and MR imaging had an NPV of 79% and a PPV of 52%. CT achieved an NPV of 90%, correlating with a PPV of 44%, with use of 5-mm size as an indicator of a positive node.

CONCLUSION: CT performed slightly better than MR imaging for all interpretative criteria. However, a high NPV was achieved only when a low size criterion was used and was therefore associated with a relatively low PPV.

Squamous cell carcinoma accounts for more than 90% of head and neck aerodigestive tract malignancies. The presence of metastasis to the cervical lymph nodes is crucial in the selection of treatment and in the evaluation of the prognosis for disease in patients with such primary tumors (1). Patients with palpable lymph nodes undergo surgery, radiation therapy, or a combination of the two. The appropriate treatment of patients without palpable nodes presents a dilemma, because a substantial number of these patients have nonpalpable nodal metastases (2–6). Clinical examinations of patients with primary tumors of the tongue, the floor of the mouth, oropharynx, hypopharynx, and supraglottic larynx (1–3,7–12) have an estimated false-negative rate of from 15% to more than 50%. Computed tomography (CT) and magnetic resonance (MR) imaging have been used to help detect metastasis and reduce this false-negative rate (12–35).

The Radiological Diagnostic Oncology Group, or RDOG, organized a multi-institutional cooperative effort to determine the relative accuracies of MR imaging and CT in determining the presence or absence of cervical lymphatic metastases from squamous cell carcinomas of the head and neck. We were interested in the effect of nodal size, internal architecture, or both on intra- and intermodality accuracy. Currently, all patients with tumors of a certain size and location in the head and neck are presumed to have metastasis to the neck and, therefore, their necks are treated. We evaluated whether CT or MR imaging could help reliable identification of a subpopulation of patients with these tumors that had no nodal metastases and thus obviate treatment of the node-bearing regions of the neck. We arbitrarily sought a negative predictive value (NPV) of at least 90%.

MATERIALS AND METHODS

Patient Criteria

Between 1991 and 1994, 213 patients (150 [70.4%] men and 63 [29.6%] women; age range, 18–84 years; mean age, 59 years ± 11.5 [standard deviation]) were enrolled in the study at three facilities—the University of Pittsburgh Medical Center (n = 99), Shands Hospital at the University of Florida College of Medicine (n = 75), and the University of Washington Medical Center (n = 39). All patients had squamous cell carcinoma in a region of the upper aerodigestive tract known to have a high frequency of nodal metastasis, but the clinical nodal status of their disease was not considered.

Patients were entered in the study after the decision had been made that they would undergo neck dissection (n = 311) with histopathologic examination before any radiation therapy or chemotherapy. One hundred fifty-eight dissections were performed on the right side of the neck and 153 on the left. Disease stage was determined according to the classification system of the American Joint Committee on Cancer Staging (36). Patients with the following lesions were enrolled: oral cavity (excluding lip), stages T2–T4 (n = 115 [53.9%]); oropharynx, stages T1–T4 (n = 27 [12.7%]); hypopharynx, stages T1–T4 (n = 18 [8.5%]); larynx (supraglottic), stages T2–T4 (n = 35 [16.4%]); larynx (glottic), stage T4 (stages T2 and T3 could be included if marked supraglottic extension existed) (n = 18 [8.5%]).

Patients were excluded from the study if they were less than 18 years old, had evidence of distant metastasis, had a history of irradiation of the head and neck for any reason, or had undergone surgery for a head and neck malignancy other than superficial skin cancer. Before imaging and within 5 weeks of surgery, each patient underwent chest radiography with findings that were negative for lung metastasis. Patients with contraindications to MR imaging or with an allergy to iodinated contrast material were excluded. Pregnant women, prisoners, and institutionalized individuals could not participate.

Patients were enrolled by the participating otolaryngologists or radiation oncologists. The enrolling physician performed the physical examination, determined the stage of the primary tumor, and decided if the patient met the criteria for inclusion in the study. The protocol, which had been approved by the institutional review boards, was described to each patient, and the patients signed a study-specific consent form.

Imaging Procedures

CT was performed with high-resolution systems. Iodinated contrast medium was injected so that the blood vessels were clearly differentiated from nodes. Section thickness was 3 or 4 mm, and the section interval was not more than 5 mm. Section scanning time was not more than 3 seconds, and the field of view was not larger than 18 cm. Scanning was performed from the occlusal plane to the clavicles.

MR imaging was performed with a standard protocol. Sagittal T1-weighted imaging was performed with repetition time of 650 msec or less and an echo time of 20 msec or less (≤650/≤20), 256 × 192 matrix, 24-cm field of view, and 5-mm section thickness. One or more signals were acquired, and imaging was performed from one sternocleidomastoid muscle to the other. Axial T1-weighted imaging (≤800/≤20, 256 × 192–256 matrix, 3–4-mm section thickness, ≤5-mm section interval) was performed before and after injection of a gadolinium chelate with a standard dose based on the patient's weight. Two or more signals were acquired, and imaging was performed from at least the middle ramus of the mandible to below the anterior arch of the cricoid cartilage. The imaging range was limited by the number of sections that could be acquired in an acceptable repetition time. Sequences with long repetition times were performed before administration of contrast medium (≤2,000/30, 80; 24-cm field of view; 5-mm section thickness; 6-mm section interval; 1-mm intersection gap). Two signals were acquired, and imaging was performed from the occlusal plane to the clavicles. Fast spin-echo T2-weighted sequences were used when the technology was available (1993 and later) (37,38).

The CT and MR examinations were completed within 4 weeks of each other and no more than 5 weeks before neck dissection.

Interpretation of Studies

CT and MR studies were interpreted prospectively and separately at the accruing hospital. Images obtained with each modality were interpreted by two different radiologists at each institution. In addition, each image was read by the radiologists at the other two institutions. Thus, eight radiologists read the images in each case: Four radiologists (two from the accruing institution and one from each of the other two institutions) read the CT images, and four read the MR images. All readers were trained head and neck radiologists or neuroradiologists or were fellows in the respective fields. All readers were trained by senior investigators (H.D.C., A.A.M., R.W.D.), who used sets of teaching images to illustrate various findings. The readers had no knowledge of the clinical stage of disease in the neck, findings with the other modality, or interpretations of the other readers.

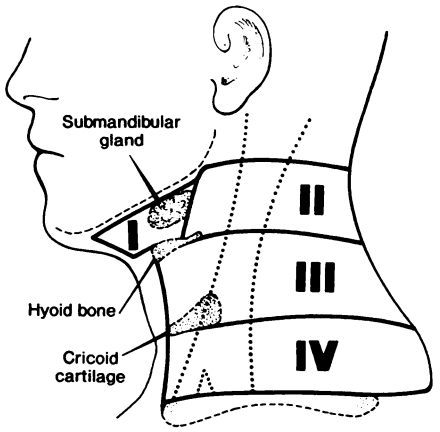

Figure 1. Diagram of the zones (*I* = zone 1, *II* = zone 2, *III* = zone 3, *IV* = zone 4) used for categorization of lymph nodes during image interpretation.

On axial images, the readers noted the largest dimension of the largest node in each zone of the neck. Measurements were made by means of comparison with a centimeter scale printed on each image.

The same interpretation protocol was used for both MR and CT images. The presence or absence within any node of abnormal signal intensity on MR images or abnormal attenuation on CT scans was indicated with a five-point scale: 0 = node definitely normal, finding definitely not present; 1 = finding probably not present; 2 = node indeterminate, unclear if finding is present or not; 3 = finding probably present; 4 = node definitely abnormal, finding definitely present.

Appearances that indicated the presence of an internal abnormality at CT included central attenuation considerably lower than that of muscle or attenuation approximately equal to that of muscle but with a markedly enhancing portion. At MR imaging, a hypointense area on the T1-weighted image that did not enhance after injection of the gadolinium chelate and a focal inhomogeneous or hyperintense area on the T2-weighted image were considered to indicate internal abnormalities.

Nodes were grouped into four zones (Fig 1), which conformed approximately but not exactly to the levels in the classification system used clinically (20,39). Zone

1 represented the submandibular and submental region and included the area anterior to the posterior margin of the submandibular gland. Zone 2 had the posterior boundary of zone 1 as an anterior boundary. All nodes in the jugular and spinal accessory chains above the inferior surface of the body of the hyoid were placed in this group. Zone 3 extended from the hyoid bone to the lower margin of the anterior arch of the cricoid. Zone 4 represented the entire neck below the level of the cricoid. In this study, zones 2–4 extended to the posterior midline of the neck, and thus there was no separate category for the posterior triangle nodes. This represents a departure from the standard clinical categorization, in which level 5 is used to describe palpable nodes posterior to the sternocleidomastoid muscle.

Quality Control and Inter-reader Reliability

All images were assessed by a quality control committee (H.D.C., A.A.M.) to ensure that all imaging parameters of the protocol had been followed and that the images were adequate. If an image was considered inferior to protocol standards, it was omitted from the analysis.

To evaluate consistency of measurements within a given institution, the average of the largest dimension of nodes in each of zones 1–3 was compared for each of the two readers. Comparisons of the two readers' measurements were made with a paired Student *t* test and with the Wilks λ statistic for a multivariate test.

Surgical Correlation and Histopathologic Examination

All patients underwent uni- or bilateral radical or modified radical dissection of the neck. The surgeon marked the specimen to identify zones 1–4. The laboratory assistant separated the nodes into zones as indicated by the surgeon. The nodes were then fixed in formalin, embedded in paraffin, sectioned, and stained with hematoxylin and eosin. The pathologist examined the nodes microscopically and indicated whether a positive node was present in each zone.

The standard of reference for a positive node was the presence of tumor at histopathologic examination. The presence of tumor in any node defined that side and zone of the neck as positive for disease.

Data Collection and Analysis

In all cases, the standard of reference was the presence or absence of tumor at histopathologic examination. Zones 1–3 were evaluated. Very few positive nodes were found in zone 4; therefore, an analysis was not performed for this region. To evaluate the entire upper jugular chain, zones 2 and 3 were combined by using the maximum nodal size and maximum internal abnormality grade over both zones. The standard of reference for this combined zone was defined as follows: The presence of tumor in any node in either zone defined the combined zone as positive for disease. If no tumor was found in any node in either zone, the combined zone was considered negative for disease. Data were aggregated over the intra- and interinstitutional readings, but the correlations that resulted from repeated readings were also taken into account in the statistical analysis.

The analysis addressed two major issues. At the test level, we used several nodal imaging criteria to compare the overall performances of CT and MR imaging in the detection of metastases to the lymph nodes. We then calculated the NPV, because of clinical interest in identifying patients who might be able to avoid neck dissection on the basis of reliable results of "negative for disease" in an imaging examination.

Comparing performance of CT and MR imaging according to various imaging criteria.—To evaluate the overall performance of CT and MR imaging and to compare the two modalities, the areas under the receiver operating characteristic (ROC) curves were determined. The ROC curves were based on different imaging criteria—on size alone and on a hybrid test that considered size in conjunction with the presence or absence of an internal abnormality. For each modality, the area under the ROC curve was derived by averaging the areas under the ROC curves from the three participating hospitals. The ROC curve for each hospital was based on observations from only that hospital. In cases in which the hospital was the accruing institution, we used the maximum value of the observations for the neck. The average area under the ROC curve for CT was then compared with the average area for MR imaging. All tests and standard errors were based on the Mann-Whitney *U* test within a multivariate framework (40).

In tests based on size alone, we presumed an increasing level of certainty of metastasis with an increase in nodal size. A test was positive for disease when the maximum nodal size at imaging was at least a specific measurement, the nodal size cutoff point. Thus, the nodal size cutoff point was the measurement of the smallest node considered to be positive for disease. Each nodal size from 5 to 20 mm was given a separate ranking, which resulted in a scale from 1 to 16. All nodes 5 mm or smaller were assigned a rank of 1, and all nodes 20 mm or larger were assigned a rank of 16. For each number on the scale, true-positive and false-positive ratios were calculated and used to plot the ROC curves.

We also defined a hybrid test that combined information on both nodal size and the presence of an internal abnormality. This test was based on an ordinal scale of 32 points. Nodes without any internal abnormality were assigned ranks 1–16, with nodes 5 mm or smaller ranked 1 and nodes 20 mm or larger ranked 16. Nodes with an internal abnormality were assigned ranks 17–32, with nodes 5 mm or smaller with an internal abnormality ranked 17 and nodes 20 mm or larger with an internal abnormality ranked 32. Again, ratios were calculated and ROC curves were plotted.

To test the validity of this ranking algorithm, a logistic regression analysis (STATA; Computing Resource Center, Los Angeles, Calif) was performed in which the dependent variable was the standard of reference (positive or negative), and the predictor variables were nodal size and the presence of an internal abnormality at imaging. An internal abnormality was said to be present if the reader assigned a rating of 3 or 4 (probably present or definitely present). With this approach, we determined the relationship between nodal size and the presence of an internal abnormality. Specifically, we calculated the ratios of the regression coefficients for size and internal abnormality for each modality. The standard errors for this relationship were calculated on the basis of the Taylor approximation of the ratio, with use of the covariance matrix of the coefficients (41).

Calculating NPV, positive predictive value, sensitivity, and specificity according to various imaging criteria.—NPVs and positive predictive values (PPVs) were calculated on the basis of combined information about nodal size and the presence of internal abnormality for zones 2 and 3. This second test was different from the hybrid test, but it was considered to be more consistent with actual clinical practice. The hybrid test was appropriate for comparison of overall performance, but it could not be easily applied in patient evaluation. If a node was smaller than the specified nodal size cutoff point and had no internal abnormality (internal abnor-

mality rating, 0–2), it was considered a negative node. If a node was at least the size of the nodal cutoff point, it was considered a positive node. If a node had an internal abnormality rating of 3 or 4, it was considered positive regardless of its size. The presence of a positive node defined the zone as positive for disease. Sensitivity and specificity values were calculated with the same interpretative criteria.

In this study, sensitivity (TP/[TP + FN], where TP = true-positive results and FN = false-negative results) is the percentage of necks with histopathologically proved positive nodes that were correctly interpreted at imaging as positive for disease. Specificity (TN/[TN + FP], where TN = true-negative results and FP = false-positive results) is the number of necks correctly interpreted at imaging as negative for disease divided by the total number of necks with no histopathologically proved positive nodes. The NPV (TN/[TN + FN]) is the percentage of necks interpreted at imaging as negative for disease that had no histopathologically proved positive nodes. The PPV (TP/[TP + FP]) is the percentage of necks interpreted at imaging as positive for disease that had histopathologically proved positive nodes.

TABLE 1
Number of Neck Dissections and Patients Who Underwent Neck Dissection

Zone	Right Side of Neck	Left Side of Neck	Total No. of Neck Dissections	Total No. of Patients
1	75 (17)	78 (26)	153 (22)	121 (26)
2	139 (34)	130 (33)	269 (34)	196 (40)
3	136 (24)	130 (25)	266 (24)	194 (27)
4	100 (14)	91 (13)	191 (14)	151 (15)

Note.—Numbers in parentheses are percentage of true-positive results.

RESULTS

Not all dissections included all zones. Table 1 presents the number of neck dissections in specific zones and the percentage of dissections with positive histopathologic results. The majority of dissections with positive histopathologic results involved zones 2 and 3 of the neck (and, to some extent, zone 1). In only three cases were the histopathologic results positive for disease in zone 4 but not in zones 2 and 3.

Comparisons between CT and MR Imaging

Table 2 presents the areas under the ROC curves (A_z) and corresponding P values (with U statistics) for CT and MR imaging with nodal size alone and with the hybrid test (nodal size and the presence of an internal abnormality). In general, CT performed best (A_z = 0.80) in the hybrid tests and was markedly better than MR imaging (A_z = 0.75) for zones 2 and 3. On the basis of the size of the data set, there was a 76% power to discriminate the 5% difference between CT and MR imaging with the hybrid test. (This power was calculated on the basis of a test with a 5% level of significance). The performance of MR imaging, as defined on the basis of the area under the ROC curves, was not changed significantly by the addition of information on internal abnormalities (A_z = 0.73 for tests of size alone versus A_z = 0.75 for the hybrid tests). ROC curves for CT and MR imaging for both tests are shown in Figure 2.

TABLE 2
Area under the ROC Curves

Zone	CT	MR Imaging	*P* Value
	Size Alone		
1	0.67 ± 0.06	0.72 ± 0.05	.13
2	0.73 ± 0.06	0.70 ± 0.05	.15
3	0.71 ± 0.04	0.72 ± 0.04	.66
2 and 3	0.77 ± 0.03	0.73 ± 0.03	.04
	Size and Internal Abnormality (hybrid test)		
1	0.76 ± 0.06	0.72 ± 0.05	.38
2	0.77 ± 0.06	0.71 ± 0.05	.02
3	0.72 ± 0.04	0.72 ± 0.04	.91
2 and 3	0.80 ± 0.03	0.75 ± 0.03	.008

Note.—Data are means ± standard errors.

The validity of the ordinal scale used in our hybrid test was strongly supported for both modalities. The logistic regression model indicated that the presence of an internal abnormality is equivalent to an 8.89-mm increase in nodal size (standard error, 2.93) for CT and a 10.48-mm increase for MR imaging (standard error, 3.58). These values show that the presence of an internal abnormality is truly dominant in the determination of positive nodal disease. Classifying a node of any size with an internal abnormality as positive for disease was appropriate with either CT or MR imaging.

NPV, PPV, Sensitivity, and Specificity

Table 3 contains NPVs and PPVs for CT and MR imaging on the basis of tests of size alone and size with internal abnormality for zones 2 and 3 combined. For example, with CT on the basis of nodal size alone with the largest node of 12 mm, the NPV is 0.79. The addition of information on internal abnormality increased the NPV to 0.83. In the latter case, 17% of nodes classified at CT as negative for disease would in fact be positive. Table 4 presents sensitivity and specificity values for the same imaging criteria used to generate the NPVs and PPVs in Table 3. (Note: this approach is not the same as the hybrid test but would be closer to clinical practice.)

Inter-reader Reliability for Evaluation of Nodal Size

To measure inter-reader variability in the measurement of nodal size, we compared the size recorded by the first reader at the patient's institution to that recorded by the second reader at that institution. The average size of the node with the largest dimension on the right side of the neck and that on the left side of the neck were recorded for each of the two CT readers and for each of the two MR image readers. The mean ± standard error of the differences (in millimeters) for CT images were 0.70 ± 0.36 in zone 1, 0.25 ± 0.24 in zone 2, and 0.11 ± 0.33 in zone 3 and for MR images were 0.56 ± 0.35 in zone 1, 0.26 ± 0.31 in zone 2, and 0.02 ± 0.33 in zone 3. We used a paired Student t test to determine whether any of these differences was significantly different from zero. For CT, the P values from these tests were 0.05, 0.29, and 0.74, respectively, and for MR imaging were 0.11, 0.40, and 0.95, respectively. Therefore, with the possible exception of zone 1 for CT, no difference appeared to exist between the first and second readers' measurements of nodal size. An overall multivariate test was applied to compare the differences in the two readings for both CT and MR imaging. The Wilks λ for this test had a P value of 0.51, indicating no overall difference.

DISCUSSION

The primary goal of the study of the Radiological Diagnostic Oncology Group

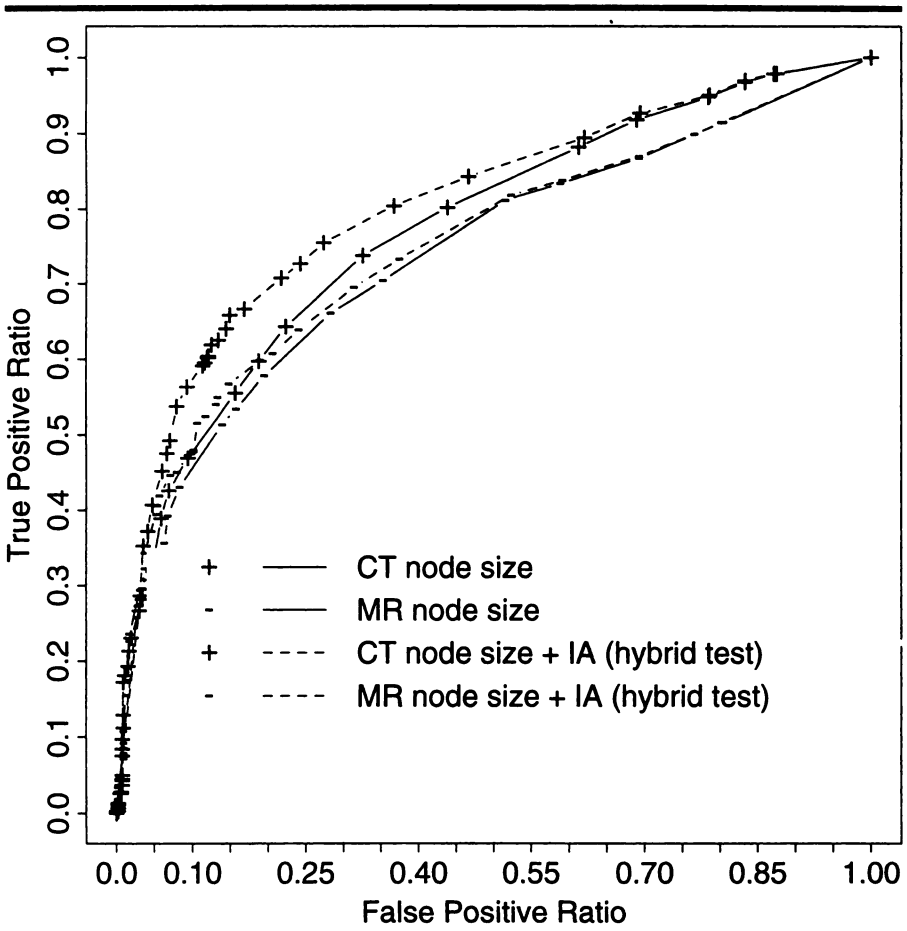

Figure 2. ROC curves for CT and MR images based on nodal size alone and on nodal size in conjunction with internal abnormalities (*IA*) (hybrid test). No information on the presence or absence of internal abnormalities was included for the curves based on size alone, but this information was included for the curves based on hybrid test results.

TABLE 3
PPVs and NPVs for CT and MR Imaging

Size (mm)	Size Alone				Size and Internal Abnormality			
	CT		MR Imaging		CT		MR Imaging	
	NPV	PPV	NPV	PPV	NPV	PPV	NPV	PPV
5	0.90	0.44	0.77	0.44	0.90	0.44	0.77	0.44
7	0.88	0.44	0.77	0.45	0.89	0.45	0.77	0.45
8	0.86	0.45	0.77	0.47	0.86	0.45	0.77	0.47
9	0.85	0.48	0.78	0.50	0.86	0.48	0.79	0.50
10	0.83	0.50	0.79	0.52	0.84	0.50	0.79	0.52
11	0.81	0.56	0.76	0.58	0.83	0.55	0.77	0.58
12	0.79	0.61	0.75	0.62	0.83	0.61	0.77	0.61
15	0.73	0.71	0.72	0.72	0.80	0.69	0.75	0.69

Note.—All values represent zones 2 and 3 (not hybrid test).

was to compare CT and MR imaging in the detection of metastases of squamous cell carcinoma to the nodes of the neck on the basis of the size of the nodes and the presence of internal abnormalities. In general, our results indicate that CT is superior to MR imaging for this purpose, particularly when information on nodal structure is considered. Such information did not significantly improve the performance of MR imaging. Information on internal abnormalities might have a greater effect at CT because findings of internal abnormality at CT have been described well whereas those at MR imaging have not been as well documented (13,14,16,17,19,20,27,29,33).

Although the addition of information on nodal structure improved the overall performance of CT, the advantage in patient treatment is not so obvious. For CT, the ROC curve (for combined zones 2 and 3) based on nodal size and internal abnormality information (the hybrid test) is statistically different from that based on size alone. However, much of the separation between the two curves occurs in the region where the sensitivity is between 50% and 75%. A test with this level of sensitivity would not be considered useful in clinical evaluation. The curves converge as sensitivity approaches 90%.

Information on internal abnormalities appears to affect sensitivity significantly at only relatively high nodal sizes. This fact suggests that the finding of a small node with an internal abnormality in the neck is rare when no larger node, with or without an internal abnormality, is present. Nodal size cutoff points of 1.0–1.5 cm are frequently used in clinical practice (20,27,42–46). For nodal size cutoff points of 1 cm or less, no significant improvement in sensitivity or specificity was achieved by adding information on internal abnormality to the information on size (Table 4). As the nodal size approaches 1.5 cm, specificity increases, but this improvement is accompanied by a steep drop in sensitivity. Use of nodal size in the range where the presence of an internal abnormality makes a difference would be of limited clinical value because of the high false-negative rate. If patients were not treated on the basis of negative imaging findings with these interpretative criteria, the tumor burden or "recurrence" rate would probably be unacceptably high.

From the perspective of treatment of our particular patient population, we wanted to understand the predictive capabilities of these modalities. We considered the NPV to be the most clinically relevant expression of the effectiveness of imaging in this clinical situation. The error rate of the negative examination is expressed as 1 − NPV and indicates the percentage of patients with nodes classified as negative at imaging but who still have tumor. If the neck is not treated in these cases, the residual nodal tumor would presumably progress, causing future morbidity and lowered survival. If the NPV is 90%, the examinations read as negative would be correct in nine of 10 patients, and one patient in 10 would have undetected residual nodal tumor. In

our study, CT achieved a NPV of 90% but only with a nodal size cutoff point of 5 mm. With use of this size criterion, the PPV would be 44%, that is, 56% of patients identified as having positive nodes would have no tumor. With use of a 1-cm nodal size cutoff point, the NPV would be 84% and the PPV 50%. Presumably, if disease in the neck in patients with negative imaging findings was left untreated, 16% of tumors would "recur." Of course, this would not be a true recurrence but rather a further growth of residual tumor. Regardless of interpretative criteria, MR imaging did not achieve a NPV of 90%.

TABLE 4
Sensitivity and Specificity for CT and MR Imaging

Size (mm)	Size Alone				Size and Internal Abnormality			
	CT		MR Imaging		CT		MR Imaging	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
5	0.98	0.13	0.92	0.20	0.98	0.13	0.92	0.20
7	0.97	0.17	0.90	0.23	0.97	0.17	0.90	0.23
8	0.95	0.22	0.87	0.31	0.95	0.21	0.87	0.31
9	0.92	0.31	0.83	0.41	0.93	0.31	0.84	0.41
10	0.88	0.39	0.81	0.48	0.90	0.38	0.82	0.48
11	0.80	0.56	0.70	0.65	0.84	0.53	0.73	0.63
12	0.74	0.67	0.66	0.72	0.80	0.63	0.70	0.69
15	0.56	0.84	0.51	0.86	0.71	0.78	0.60	0.81

Note.—All values represent zones 2 and 3 (not hybrid test).

Predictive probabilities vary with the prior probability of disease (nodal, in this case). Our patients came from three large institutions, and we believe that the prior probability of disease in our study is representative of patients with tumors in these locations, but we cannot be sure of this. These prior and predictive probabilities cannot be applied to other clinical situations (ie, the differentiation of a normal from an abnormal node in a patient without the types of squamous cell carcinoma we studied).

This study has several limitations, some of which are related to design and others to analysis. With regard to design and data collection, the nodes were measured by means of comparison with a scale printed along the image margin. Although measurements with electronic calipers would likely have been more accurate, their use would not correspond to routine clinical practice. Under the circumstances of this study, the error rate was approximated at ±1 mm. With regard to the interpretation of an internal abnormality (as well as the certainty of its presence), some variability existed among readers (not reported), even though each reader was trained before the start of the study.

For a given nodal size, the NPV of MR imaging was lower than that of CT. In many cases, a similar NPV could be achieved with MR imaging but at a lower nodal size cutoff point. This fact suggests that the nodal measurements were not the same for the two modalities; smaller measurements were possible with MR imaging than with CT. When we compared the maximum nodal size in each region in each neck in the primary data set, the average measurement at CT was 1.4 mm larger than that at MR imaging. This difference was significant (P = .0001). Our readers measured nodal size on T1-weighted MR images. The nodes were bordered by hyperintense fat, so the window settings might affect the measurement. Use of narrow window settings can cause the hyperintensity to "spill" into the lower signal intensity that represents the node. This would lower the measured size of the structure with intermediate signal intensity. Even with use of a very wide window setting, the high signal intensity of fat may diminish the conspicuity of the margin, resulting in a lower measurement. Appropriate window settings were defined in the protocol as wide enough to show septal detail within the fat, but further analysis of the boundary definition between node and contiguous fat was not addressed in this study.

Several problems occurred with the assignment of nodes to specific zones at both imaging and surgery. Although we would have liked to use the same level classification system that is used in clinical practice, we made slight modifications because it was difficult to apply the clinical system to imaging examinations. For example, the anterior margin of level 5 in the clinical classification is the posterior border of the sternocleidomastoid muscle, and many nodes are deep to this muscle. Thus, instead of creating an artificial separation between zone 2 or 3 and zone 5, we defined the zones in our study on the basis of readily identifiable imaging landmarks.

In our analysis, findings in zone 2 were combined with those in zone 3 partly because of the difficulty of assigning nodes into one zone or the other at both imaging and histopathologic examination. At imaging, nodes frequently spanned the border of zones 2 and 3. In addition, all imaging landmarks were in the anterior portion of the neck, and the exact plane for axial imaging could not be specified. Variability of angulation was present that led to ambiguity regarding the exact border between zones 2 and 3. During surgery, efforts were made to mark the various zones. When the surgical field is opened, however, the anatomy is distorted and some variability in defining the exact separation of zones 2 and 3 occurs.

Another study limitation exists with regard to the standard of reference. In this study, each node was cut for assessment only once. Findings in several studies have shown that additional positive nodes can be identified if a more detailed examination is carried out by examining multiple sections or by performing an immunologic assessment (47,48). Another means of evaluating for false-negative disease in patients with an untreated neck in whom imaging findings were negative involves close follow-up after a preset time interval (49). Even with this approach, however, we could not be certain whether disease that arose during the observation period was a result of undetected disease or was new disease.

The evolution of technology continues. These cases were accrued over a 3-year period. Protocols were designed on the basis of the state of the art at the beginning of the study. No important changes were made in the CT protocol. However, fast spin-echo imaging and fat suppression represented important advances in MR imaging of the head and neck. Fast spin-echo imaging was incorporated into the study after 2 years. Findings in a separate study indicated that the visibility of nodes was improved with fast spin-echo T2-weighted imaging, so the new sequence was used (37). If this affected the data in any way, the change would presumably improve the performance of MR imaging. Fat suppression after injection of a gadolinium chelate has also been an important advance. This technology was not completely evaluated

in time to be included in a major portion of the study, however, so it was not used.

Our analytic approach validated the importance of internal abnormality over that of nodal size. It is possible that if we had used more refined measures of internal architecture, perhaps techniques from feature analysis, the relationship between internal architecture and size would have been altered. Such an analysis would necessitate re-readings with prospectively designed templates that describe and quantify each of the possible features associated with the internal architecture. This approach, with use of radiologists who see a high volume of CT and MR images of the head and neck, would give us better information about the relative contributions of these modalities.

CT performed slightly better than MR imaging for all interpretative criteria, but a high NPV was achieved only when the small-size criterion was used and, therefore, was associated with a relatively low PPV. CT performed slightly better than MR imaging, particularly when internal abnormalities were considered in addition to nodal size. However, the ability of either modality to achieve a NPV of 90% was disappointing. Any size criterion greater than 5 mm was associated with a large number of "misses." For CT, when the presence of a node 5 mm or larger or of a node with an internal abnormality was considered a positive result, the NPV was 90% and the PPV was 44%. With these test criteria, 10% of patients classified as having negative results at imaging would have residual tumor in the neck. When a positive test was defined as the presence of a node 10 mm or larger or of a node with an internal abnormality, the NPV was 84% and the PPV was 50%. With these test criteria, 16% of the patients with negative imaging results would actually have disease, and 50% of the patients with positive imaging results would not have metastasis and would undergo unnecessary treatment. If the patients with negative imaging results were not treated, the presumed rate of "recurrence" would be 16%.

Acknowledgments: The authors thank Alan Zaslavsky, PhD, Department of Health Care Policy, Harvard Medical School, for helpful advice regarding the statistical aspects of the study and Constantine A. Gatsonis, PhD, Center for Statistical Science, Brown University, for help with the study design. The authors also thank Kathy Parkhurst, BA, JoAnn Stetz, RN, RTT, and Cynthia Olson, MBA, MHS, of the American College of Radiology for data management. Finally, the authors thank research assistants William Cox, Joy Wilcox, and Sharon Elliot, as well as the many image readers who participated in the study.

References

1. Johnson JT. A surgeon looks at cervical lymph nodes. Radiology 1990; 75:607–610.
2. Ogura JH, Biller HF, Wette R. Elective neck dissection for pharyngeal and laryngeal cancers. Ann Otol Rhinol Laryngol 1971; 80:646–650.
3. Bocca E, Calearo C, DeVincentiis I, Marullo T, Motta G, Ottaniani A. Occult metastases in cancer of the larynx and their relationship to clinical and histological aspects of the primary tumor. Laryngoscope 1984; 94:1086–1090.
4. Berger DS, Fletcher GH, Lindberg RD, Jesse RH. Elective irradiation of the neck lymphatics for squamous cell carcinomas of the nasopharynx and oropharynx. Am J Roentgenol Radium Ther Nucl Med 1971; 111:66–72.
5. Fletcher GH. Elective irradiation of subclinical disease: cancers of the head and neck. Cancer 1972; 29:1450–1454.
6. Marks JE, Breaux S, Smith PG. The need for elective irradiation of occult lymphatic metastases from cancers of the larynx and pyriform sinus. Head Neck Surg 1985; 8:3–8.
7. Sako K, Pradier RN, Marchetta FC, Pickren JW. Fallibility of palpation in the diagnosis of metastasis to cervical nodes. Surg Gynecol Obstet 1964; 118:989–990.
8. Lutz CK, Johnson JT, Wagner RL, Myers EN. Supraglottic cancer: patterns of recurrence. Ann Otol Rhinol Laryngol 1990; 99:12–17.
9. Snyderman NL, Johnson JT, Schramm VL Jr, Myers EN, Bedetti CD, Thearle P. Extracapsular spread of carcinoma in cervical lymph nodes. Cancer 1985; 56:1597–1599.
10. Nason RW, Sako K, Beecroft WA, Razack MS, Bakamjian VY, Sheed DP. Surgical management of squamous cell carcinoma of the floor of the mouth. Am J Surg 1989; 158:292–296.
11. Zeitels SM, Comanowski GF, Vincent ME, Malhotra C, Vaughan CW. A model for multidisciplinary data collection for cervical metastasis. Laryngoscope 1991; 101: 1313–1317.
12. Friedman M, Mafee MF, Pacella BL, Strorigl TL, Dew LL, Toriumi DM. Rationale for the elective neck dissection in 1990. Laryngoscope 1990; 100:54–59.
13. Mancuso AA, Maceri D, Rice D, Hanafee W. CT of cervical lymph node cancer. AJR 1981; 136:381–385.
14. Reede DL, Bergeron RT. CT of cervical lymph nodes. J Otolaryngol 1982; 11:411–418.
15. Dillon WP, Mills CM, Kjos B, DeGroot J, Brant-Zawadzki M. Magnetic resonance imaging of the nasopharynx. Radiology 1984; 152:731–738.
16. Dooms GC, Hricak H, Crooks LE, Higgins CB. Magnetic resonance imaging of the lymph nodes: comparison with CT. Radiology 1984; 153:719–728.
17. Dooms GC, Hricak H, Moseley MR, Bottles K, Fisher MR, Higgins CB. Characterization of lymphadenopathy by magnetic relaxation times: preliminary results. Radiology 1985; 155:691–697.
18. Stevens MH, Harnsberger HR, Mancuso AA. Computed tomography of cervical lymph nodes: staging of head and neck cancer. Otolaryngol 1987; 111:1307–1310.
19. Mancuso AA, Harnsberger HR, Muraki A, Stevens MH. Computed tomography of cervical and retropharyngeal lymph nodes: normal anatomy, variants, and applications in staging head and neck cancer. II. Pathology. Radiology 1983; 148: 715–723.
20. Som PM. Lymph nodes of the neck. Radiology 1987; 165:593–600.
21. Steinkamp HJ, Hosten N, Richter C, Schedel H, Felix R. Enlarged cervical lymph nodes at helical CT. Radiology 1994; 191: 795–798.
22. Close LG, Merkel M, Vuitch MF, Reisch J, Schaefer SD. Computed tomographic evaluation of regional lymph node involvement in cancer of the oral cavity and oropharynx. Head Neck 1989; 11:309–317.
23. Stern WB, Silver CE, Zeifer BA, Persky MS, Heller KS. Computed tomography of the clinically negative neck. Head Neck 1990; 12:109–113.
24. Hillsamer PJ, Schuller DE, McGhee RG, Chakeres D, Young DC. Improving diagnostic accuracy of cervical metastases with computed tomography and magnetic resonance imaging. Arch Otolaryngol Head Neck Surg 1990; 116:1297–1301.
25. Feinmesser R, Freeman JL, Feinmesser M, Noyek A, Mullen JB. Role of modern imaging in decision-making for elective neck dissection. Head Neck 1992; 14:173–176.
26. Feinmesser R, Freeman JL, Noyek AM, Birt D, Gullane P, Mullen JB. MRI and neck metastases: a clinical, radiological, and pathological correlative study. J Otolaryngol 1990; 19:136–140.
27. Som PM. Detection of metastasis in cervical lymph nodes: CT and MR criteria and differential diagnosis. AJR 1992; 158:961–969.
28. Van den Brekel MW, Stel HV, Castelijns JA, et al. Cervical lymph node metastasis: assessment of radiologic criteria. Radiology 1990; 177:379–384.
29. Van den Brekel MW, Castelijns JA, Stel HV, et al. Detection and characterization of metastatic cervical adenopathy by MR imaging: comparison of different MR techniques. J Comput Assist Tomogr 1990; 14:581–589.
30. Van den Brekel MW, Castelijns JA, Croll GA, et al. Magnetic resonance imaging vs palpation of cervical lymph node metastasis. Arch Otolaryngol Head Neck Surg 1991; 117:666–673.
31. Friedman M, Roberts N, Kirschebaum GL, Colombo J. Nodal size of metastatic squamous cell carcinoma. Laryngoscope 1993; 103:854–856.
32. Moreau P, Goffart Y, Collignon J. Computed tomography of metastatic cervical lymph nodes: a clinical, computed tomographic, pathologic correlative study. Arch Otolaryngol Head Neck Surg 1990; 116: 1190–1193.
33. Yousem D, Som PM, Hackney DB, Schwaibold F, Hendrix RA. Central nodal necrosis and extracapsular neoplastic spread in cervical lymph nodes: MR imaging versus CT. Radiology 1992; 182:753–759.
34. Don DM, Anzai Y, Lufkin RB, Fu YS, Calcaterra TC. Evaluation of cervical lymph node metastases in squamous cell carcinoma of the head and neck. Laryngoscope 1995; 105:669–674.
35. August M, Nguyen M. Evaluation of metastatic neck disease by computed tomography. Int J Oral Maxillofac Surg 1994; 23:290–293.
36. Beahrs OH, Henson DE, Hutter RVP, et al, eds. Committee on cancer: manual for staging of cancer. 3rd ed. Philadelphia, Pa: Lippincott, 1988.

37. Lewin JS, Curtin HD, Ross JS, Weissman JL, Obuchowski NA, Tkach JA. Fast spin-echo imaging of the neck: comparison with conventional spin-echo, utility of fat suppression, and evaluation of tissue contrast characteristics. AJNR 1994; 15:1351–1357.
38. Yousem DM, Hurst RM. MR of cervical lymph nodes: comparison of fast spin-echo and conventional spin-echo T2W scans. Clin Radiol 1994; 49:670–675.
39. Suen JY, Goepfert H. Standardization of neck dissection nomenclature (editorial). Head Neck Surg 1987; 10:75–77.
40. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837–845.
41. Skinner GJ, Holt D, Smith TMF. Analysis of complex surveys. Chichester, NY: Wiley, 1989.
42. Friedman M, Shelton VK, Mafee M, Bellity P, Grybauskas V, Skolnik E. Metastatic neck disease: evaluation by computed tomography. Arch Otolaryngol 1984; 110: 443–447.
43. Stevens MH, Harnsberger HR, Mancuso AA, Davis RK, Johnson LP, Parkin JL. Computed tomography of cervical lymph nodes: staging and management of head and neck cancer. Arch Otolaryngol 1985; 111:735–739.
44. Watkinson JC, Todd GEC, Paskin L, et al. Metastatic carcinoma in the neck: a clinical and radiological scintigraphic and pathological study. Arch Otolaryngol 1991; 16:187–192.
45. Feinmesser R, Freeman JL, Nojek AM, Birt BD. Metastatic neck disease: a clinical/radiographic/pathologic correlative study. Arch Otolaryngol Head Neck Surg 1987; 113:1307–1310.
46. Bergman SA, Ord RA, Rothman M. Accuracy of clinical examination versus computed tomography in detecting occult lymph node involvement in patients with oral epidermoid carcinoma. J Oral Maxillofac Surg 1994; 52:1236–1239.
47. Van den Brekel MW, Stel HV, Van der Valk P, van der Waal I, Meyer CJ, Snow GB. Micrometastases from squamous cell carcinoma in neck dissection specimens. Eur Arch Otorhinolaryngol 1992; 249:349–353.
48. Saka SM, MacDonald DG. Sampling of jugulo-digastric lymph nodes in oral cancer. J Oral Pathol Med 1989; 18:123–124.
49. August M, Gianetti K. Elective neck irradiation versus observation of the clinically negative neck of patients with oral cancer. J Oral Maxillofac Surg 1996; 54:1050–1055.