Marilyn J. Siegel, MD
Hemant Ishwaran, PhD
Barry D. Fletcher, MD
James S. Meyer, MD
Fredric A. Hoffer, MD
Diego Jaramillo, MD
Ramiro J. Hernandez, MD
Susan E. Roubal, DO
Barry A. Siegel, MD
Daryl J. Caudry, MS
Barbara J. McNeil, MD, PhD

Index terms:
Computed tomography (CT), comparative studies, 60.1211, 80.1211
Magnetic resonance (MR), comparative studies, 60.1214, 80.1214
Neoplasms, staging
Neuroblastoma, 60.3251
Radionuclides, comparative studies, 60.1216, 80.1216

Published online before print
10.1148/radiol.2231010841
Radiology 2002; 223:168–175

Abbreviations:
RMSE = root-mean-square-error
ROC = receiver operating characteristic

[1] From the Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S Kingshighway Blvd, St Louis, MO 63110 (M.J.S., B.A.S.); Department of Biostatistics, Cleveland Clinic Foundation, Cleveland, Ohio (H.I.); Department of Diagnostic Imaging, St Jude's Children's Hospital, Memphis, Tenn (B.D.F., F.A.H.); Department of Radiology, Children's Hospital of Philadelphia, Pa (J.S.M.); Departments of Radiology (D.J.) and Health Care Policy (D.J.C., B.J.M.), Harvard Medical School, Boston, Mass; Department of Radiology, C. S. Mott Children's Hospital, Ann Arbor, Mich (R.J.H.); Department of Pediatric Imaging, Children's Hospital, Detroit, Mich (S.E.R.); and Department of Radiology, Brigham and Women's Hospital, Boston, Mass (B.J.M.). Received April 25, 2001; revision requested June 15; revision received August 3; accepted September 17. Supported by grant CA59403 from the National Cancer Institute. **Address correspondence to** M.J.S. (e-mail: *siegelm@mir.wustl.edu*).

Author contributions:
Guarantors of integrity of entire study, M.J.S., H.I., D.J.C., B.A.S. The complete list of author contributions appears at the end of this article.

Staging of Neuroblastoma at Imaging: Report of the Radiology Diagnostic Oncology Group[1]

PURPOSE: To compare the accuracies of computed tomography (CT), magnetic resonance (MR) imaging, and bone scintigraphy in staging disease in patients with neuroblastoma.

MATERIALS AND METHODS: Ninety-six children with newly diagnosed neuroblastoma were enrolled in a multicenter prospective cohort study. CT, MR, and bone scintigraphy were used to evaluate tumor stage. Sensitivity and specificity values and receiver operating characteristic (ROC) curve analyses were used to compare the accuracy of CT, MR, and scintigraphy for tumor staging.

RESULTS: Eighty-eight patients were eligible for staging analysis, and 45 patients who underwent surgery at initial diagnosis were eligible for analysis of local tumor extent. CT and MR had sensitivities of 43% and 83%, respectively ($P < .01$), and specificities of 97% and 88%, respectively ($P > .05$), for detection of stage 4 disease. Areas under the ROC curves for CT and MR were 0.81 and 0.85, respectively ($P = .06$); that for scintigraphy was 0.83. Addition of scintigraphy to both CT and MR increased the areas under the ROC curves to 0.90 and 0.88, respectively. Accuracy of CT and MR for staging disease confined to the chest or abdomen (stages 1, 2, and 3) was poor.

CONCLUSION: MR alone and CT and MR combined with bone scintigraphy enable the accurate detection of stage 4 disease. Both CT and MR perform poorly for local tumor staging.

Neuroblastoma accounts for 8%–10% of all childhood cancers in the United States (1,2). The treatment of neuroblastoma is determined by the stage of the tumor at time of diagnosis (1–5). Disease that is regionally limited is potentially resectable, while disease that is locally extensive or disseminated is usually not resectable. The staging and monitoring of disease in patients with neuroblastoma are major areas of application for diagnostic imaging methods. Thus, diagnostic studies that enable accurate tumor staging should help in treatment planning and reduce unnecessary surgery.

The current staging evaluation of patients with neuroblastoma consists of computed tomography (CT) or magnetic resonance (MR) imaging of the primary tumor, a skeletal survey and either bone scintigraphy or metaiodobenzylguanidine scintigraphy for skeletal metastases, and bone marrow aspiration and biopsy for marrow disease (6–14). CT has a reported accuracy of about 80% in tumor staging; when CT is complemented with scintigraphy or bone marrow aspiration, the accuracy has been reported to increase to 97% (15). MR imaging has also yielded high sensitivities (85%–100%) in the detection of abdominal disease and distant metastases (6,7,13,16,17). However, these imaging studies were performed with older-generation equipment, were retrospective, and each imaging modality was evaluated in isolation.

To our knowledge, no large prospective studies have compared CT, MR imaging, and scintigraphy in the staging of neuroblastoma. The purpose of our multiinstitutional Radiology Diagnostic Oncology Group study, sponsored by the National Cancer Institute,

was to compare the relative accuracies of CT, MR, and bone scintigraphy in the staging of disease in patients with neuroblastoma.

MATERIALS AND METHODS

Study Subjects and Setting

Six institutions participated in this study: Mallinckrodt Institute of Radiology, Washington University School of Medicine (St Louis, Mo); St Jude's Children's Hospital (Memphis, Tenn); C. S. Mott Children's Hospital (Ann Arbor, Mich); Boston Children's Hospital (Mass); Children's Hospital of Philadelphia (Pa); and Children's Hospital of Detroit (Mich). The study was approved by each of the corresponding institutional review boards.

Children under 18 years of age who were suspected of having neuroblastoma of the chest or abdomen and pelvis on the basis of results of chest radiographs or abdominal sonograms were eligible for inclusion in the study. Patients with a history of other malignancies and those with newly diagnosed tumors who had received chemotherapy or radiation therapy for longer than 48 hours before the imaging examinations were ineligible for the study. Other exclusion criteria were contraindications to sedation, a history of major allergic reaction to intravenous contrast material, and the presence of cardiac pacemakers or intracranial vascular clips. The parents gave written informed consent prior to each patient's participation in the study. After parental permission was obtained and the consent form was signed, each participating patient was enrolled by telephone at the American College of Radiology office in Philadelphia, Pa.

Study Protocol

The age, sex, and race of all patients were recorded on demographic forms.

All patients were to undergo CT, MR imaging, bone scintigraphy, and bone marrow aspiration and biopsy at the time of initial presentation. All studies were to be performed within 2 weeks of each other and within 2 weeks of any surgical procedure. Patients who underwent delayed surgery after diagnosis were included in the analysis of staging but not in the analysis of the extent of local disease.

Imaging Protocols

CT protocol.—All patients underwent CT of the chest or abdomen and pelvis, depending on the site of the primary tumor. The osseous structures of the pelvis, including the femoral heads, were routinely included in examinations of the abdomen. The thoracic spine, the ribs, and the shoulders were examined in patients with chest tumors. The original study design did not provide for evaluation of the pelvis and proximal femurs in patients with chest disease.

The CT scans were obtained with fourth-generation scanners (Somatom Plus or Plus-S, Siemens Medical Systems, Iselin, NJ; or 9800 GE Advantage or HiSpeed Advantage, GE Medical Systems, Milwaukee, Wis). Patients under 5 years of age underwent sedation with either oral chloral hydrate or intravenous pentobarbital sodium (Nembutal; Abbott, North Chicago, Ill). All patients received nonionic intravenous contrast medium at a dose of 2 mL per kilogram of body weight, given either by a rapid push by hand or by a mechanical injector. The examinations were begun after 80% of the intravenous contrast material had been injected. Chest examinations extended from the lung apices to the lower edge of the liver. Abdominal examinations extended from the level of the diaphragm to the pubic symphysis. Scans were obtained with 8–10-mm collimation, an 8–10 mm/sec table speed, and an 8–10-mm reconstructed section thickness, except in neonates and in infants under 2 years of age. In these patients, the scan parameters were reduced by half. Scan time was 1 second per section. The time for the entire CT examination was approximately 30 seconds.

All CT studies were photographed at window width and level settings that best enabled evaluation of soft-tissue structures (ie, +300 to +400 HU width; +30 to +40 HU level), lungs (ie, −1,450 to −1,200 HU width; −450 to −250 HU level), and bones (ie, +2,000 HU width; +300 HU level). The exact settings were individualized for each patient with each scanner.

MR imaging protocol.—MR imaging studies were performed with 1.5-T units (Signa, GE Medical Systems; or Magnetom, Siemens Medical Systems). All patients underwent MR imaging of the bone marrow (pelvis, proximal femora, and spine) and the chest or abdomen and pelvis, depending on the site of the primary tumor. The pelvis was routinely imaged as part of the abdominal examination. Either the thoracic or the lumbosacral spine was studied, depending on the site of the primary tumor.

Examination of the pelvic and femoral marrow included images obtained from the iliac crests to the middle of the femurs with coronal T1-weighted spin-echo (repetition time msec/echo time msec, 500–650/10–15) and transverse T2-weighted fat-saturated fast spin-echo (3,000–4,000/90–100) sequences. The lumbosacral spine was imaged with a sagittal fast short inversion time inversion-recovery sequence (repetition time msec/echo time msec/inversion time msec, 4,000/10/150). The chest and abdomen were examined with coronal T1-weighted spin-echo and transverse fat-saturated T2-weighted fast spin-echo sequences (performed with same parameters as for pelvic and femoral marrow) and a transverse T1-weighted gradient-echo sequence (25/5–10; flip angle, 30°–60°). The T1-weighted sequence (650–700/10–15) was repeated with fat suppression after intravenous injection of 0.2 mL/kg of gadopentetate dimeglumine (Magnevist; Berlex Laboratories, Wayne, NJ). Section thickness varied from 5 mm with a 1-mm intersection gap to 8–10 mm with a 2-mm gap and was determined by patient size and the desire to cover the entire tumor in a limited number of sequences. Head or Helmholtz coils were used for neonates and infants under 2 years of age; body coils were used for older children.

Images were recorded on film with brightness scaled by machine. The time for the entire MR imaging examination was approximately 60 minutes.

Scintigraphy protocols.—The study protocol required that bone scintigraphy be performed in all patients and allowed for additional metaiodobenzylguanidine scintigraphy as an option at the discretion of the clinicians and surgeons at each institution. Metaiodobenzylguanidine imaging was performed in patients who had a negative bone marrow aspirate or biopsy or equivocal bone scans. An equivocal bone scan was defined as one that showed fewer than three lesions.

For bone scintigraphy, technetium 99m methylene diphosphonate was given at a dose of 280 μCi per kilogram (10.4 MBq/kg) of body weight (minimum dose, 2 mCi [74 MBq]), and imaging was begun 2 hours after injection. Examinations were performed with a large-field-of-view gamma camera, with a high-resolution collimator in children over 2 years of age, and a high-resolution or converging collimator in younger children. Multiple overlapping spot images were obtained of the entire body.

Quality Control

Copies of the CT, MR, and scintigraphic images were submitted to study headquarters for storage and subsequent review. A quality-control committee composed of the study chair (M.J.S) and two principal investigators (J.S.M., F.A.H.), met three times during the period of the study to review the submitted images for image quality and adherence to study protocols. If the CT or MR images did not adhere to technical standards defined in the protocol or were of poor technical quality, they were excluded from the final analysis. Technical exclusions were related to incomplete anatomic coverage or incomplete imaging sequences. Unacceptable quality referred to images that were degraded by patient motion or that had insufficient intravenous contrast enhancement or poor opacification of the bowel.

Interpretation of Studies

Three radiologists, one for each modality, read the CT, MR, and scintigraphic examinations separately and prospectively at each participating site. The readers had access to the clinical history, but were unaware of the results of other imaging modalities, surgery, or histopathologic examinations.

To assess interreader variability, the initial CT and MR images were reread by participating radiologists (M.J.S., B.D.F., J.S.M., F.A.H., D.J., R.J.H., S.E.R.) from other institutions. The rereading was accomplished by sending the imaging examinations to all six institutions. The radiologist who performed the rereadings had access to the patient's clinical history. The design of the original study did not require that bone scintigrams be reread.

All eligible cases were assessed for the location of the primary tumor and for features relevant to local and distant staging. The location of the primary tumor was recorded on an imaging evaluation form. The diagnosis of distant spread was based on the presence of metastases in bone, bone marrow, liver, lung, and/or nonregional lymph nodes.

The diagnosis of bone metastases was based on the following criteria: areas of cortical bone destruction on CT or MR images, diffuse or focal changes in the bone marrow signal intensity on images obtained with any or all of the MR sequences, and the presence of two or more focal areas of either increased or decreased activity on skeletal scintigrams. The finding of at least two focal lesions was required for diagnosis to minimize the likelihood that benign lesions would falsely be considered to represent metastatic disease.

Lung metastases were defined as foci of soft-tissue attenuation at CT or abnormal signal intensity at MR imaging (ie, intermediate intensity on T1-weighted images and high intensity on T2-weighted images). Liver metastases were defined as foci of abnormally low attenuation at CT or abnormal signal intensity at MR imaging (ie, low intensity on T1-weighted images and high intensity on T2-weighted images), with or without contrast material enhancement. Any visualized nonregional lymph node was considered abnormal. This decision was based on a consensus agreement of the principal investigators that it is highly unusual to see lymph nodes, regardless of size, in prepubertal children on CT or MR images.

The diagnosis of local tumor extension was based on findings of midline extension or regional lymph node involvement. These two parameters were chosen because they affect staging. Vascular encasement and intraspinal extension were not evaluated in the analysis of local tumor extent. They do not affect staging, although they do influence treatment. Midline extension was defined as tumor originating on one side and crossing the midline to or beyond the opposite side of the vertebral body. Local lymph node involvement was defined as visualization of regional lymph nodes regardless of their size.

Radiologists were instructed to score the presence of each of the aforementioned findings by using a five-point confidence scale: 0 = finding definitely not present, 1 = finding probably not present, 2 = indeterminate, 3 = finding probably present, and 4 = finding definitely present. The readers of the CT and MR images also recorded the predicted stage (stage 1–4) on the basis of all the imaging findings in the study.

Final Diagnoses

The standard of reference for the final staging of the tumors was based on a combination of surgical and histopathologic results. The staging reference was the International Neuroblastoma Staging System (Table 1). In all patients included in the local-extent analysis, results of histopathologic examinations were available at the time of initial diagnosis. Histopathologic findings were recorded on a pathology examination form. If the tumor was unresectable, the patient underwent chemotherapy and subsequent imaging evaluation. These patients underwent a delayed-response staging operation and were not included in the local-extent analysis.

Patients who were eligible for the staging analysis underwent marrow aspiration and biopsy of both iliac crests. A positive bone marrow aspiration or biopsy result was considered proof of distant tumor (stage 4). Hepatic, lung, or pleural lesions or nonregional lymph nodes were selected for biopsy or resection when the diagnosis of stage 4 disease was equivocal at marrow examination. Histocytologic findings were also considered proof of distant disease in patients who underwent needle biopsy and/or resection of lymph nodes or liver lesions.

Statistical Analysis

CT and MR images obtained at the time of initial evaluation were used in the analysis of staging and detection of distant metastases. The following analyses were conducted: *(a)* comparison of CT, MR, and bone scintigraphic images alone and CT and MR images in combination with scintigraphic images in the detection of stage 4 disease; *(b)* comparison of CT, MR, and scintigraphic images in the detection of bone and bone marrow metastases; and *(c)* comparison of the positive predictive values of CT and MR images in the detection of stage 1, 2, 3, and 4 disease. Because skeletal metastases account for the majority of stage 4 disease, we believed it important to analyze separately the detection of bone and bone marrow lesions and the detection of all stage 4 disease.

We used the predicted stage provided by the reader to create receiver operating characteristic (ROC) curves to compare the accuracy of the imaging modalities in the detection of all stage 4 disease. The confidence scale ratings were used to create ROC curves to compare the accuracy of the modalities in the detection of stage 4 disease characterized only by bone or bone marrow metastases. Areas under the ROC curves, tests for differences in areas, and confidence intervals were derived by means of a multivariate bootstrap approach (19,20). Because the area under the ROC curve is related to the Mann-Whitney *U* statistic, the technique amounts to a multivariate bootstrap technique for *U* statistics (21). In this approach, we bootstrapped the original ROC data by patient and then computed the ROC curve for each reader from the boot-

TABLE 1
International Neuroblastoma Staging System

Stage	Description
1	Localized tumor confined to the area of origin; complete gross resection with or without microscopic residual disease; identifiable ipsilateral and contralateral lymph nodes negative macroscopically.
2A	Localized tumor with incomplete gross excision; identifiable ipsilateral and contralateral lymph nodes negative microscopically.
2B	Unilateral tumor with complete or incomplete gross resection with positive ipsilateral regional lymph nodes; contralateral lymph nodes negative microscopically.
3	Tumor infiltrating across the midline with or without regional lymph node involvement; unilateral tumor with contralateral regional lymph node involvement; or midline tumor with bilateral regional lymph node involvement.
4	Dissemination of tumor to distant lymph nodes, bone, bone marrow, liver, or other organs (except as defined in stage 4S).
4S	Localized primary tumor (as defined for stage 1 or 2A or 2B) with dissemination limited to skin, liver, or bone marrow (<10% tumor cells, and metaiodobenzylguanidine scan negative in the marrow). Limited to infants <1 year of age.

strapped data. Each bootstrap sample generated one ROC curve and one ROC area for each reader. From the resulting vector of areas for the readers, we computed an overall averaged ROC area for the imaging modality, using weights inversely proportional to the number of cases read by each particular reader. The same method was also used to compute differences in areas. The multivariate feature of the approach automatically accounts for the correlation in the data due to the multiple readings per image. Using all readings, we also calculated the true-positive and true-negative values for CT and MR imaging for stage 4 disease. These data were also presented as the frequency (percentage) with which the disease was understaged with each of the modalities considered.

Interreader variability in accuracy was assessed for each modality by means of a root-mean-square-error (RMSE) measure of variability. This value is on the scale of the area and was computed for each bootstrap sample by taking the square root of the weighted squared difference between each reader's bootstrap ROC area and the overall bootstrap-averaged ROC area. As before, we used weights inversely proportional to the number of cases read by the particular reader. RMSE values below 10% indicate low interreader variability, while values above 20% are evidence of high variability.

The reference standard for constructing the ROC curves for accuracy of disease staging was defined as the presence or absence of stage 4 disease (including 4S disease). Sensitivity was the probability that stage 4 disease was correctly identified at each imaging modality, and specificity was the probability that stage 1, 2, or 3 disease was correctly identified at each imaging modality. In obtaining an ROC curve for the combinations of CT and scintigraphy and MR and scintigraphy, we created a new ordinal scale by adding the number 4 to each of the CT and MR ratings whenever the companion scintigraphic studies had a rating of 3 or 4. Any number greater than 4 would suffice for the purpose of denoting the greater likelihood of disease than that maximally achievable at CT alone or MR alone. For example, a patient given a rating of 1 at CT alone and 3 or 4 at bone scintigraphy would be given a rating of 5 (1 + 4) for the combination of CT and scintigraphy. Nothing was added to the CT or MR rating when the scintigraphic study rating was less than 3. The resulting ordinal scales for CT and scintigraphy and MR and scintigraphy had values ranging from 1 to 8.

The diagnostic value of CT and MR imaging in the assessment of specific sites of local disease that affected local staging (eg, midline extension and nodal involvement) was also calculated with ROC curves derived with a multivariate bootstrap approach and the confidence scale ratings for each of the specific findings. In patients with a false-negative diagnosis of stage 4 disease, the MR and CT studies were retrospectively reviewed by the study chair to determine the reason for the diagnostic errors.

RESULTS

Study Cohort

Between January 1994 and April 1997, 129 patients suspected of having neuroblastoma were enrolled into the study. Eighteen patients were excluded because they did not have neuroblastoma at pathologic examination. The diagnoses in these patients included ganglioneuroma (n = 4), rhabdomyosarcoma (n = 3), pulmonary sequestration (n = 1), peripheral nerve sheath tumor (n = 1), neuroendocrine tumor (n = 1), neuroectodermal tumor (n = 1), round blue cell tumor (n = 1), malignant rosettes (n = 1), Ewing sarcoma (n = 1), benign mesenchymal tumor (n = 1), and nonneuroblastoma or lesion of unknown histologic type (n = 3).

Of 111 patients with histologically proven neuroblastoma, 15 were ineligible for the study because they had primary tumors arising in the neck (n = 2), had undergone surgical resection prior to study entry (n = 1) or radiation or chemotherapy more than 48 hours prior to imaging (n = 10), or had technically inadequate MR examinations (n = 2). Thus, 96 patients were potentially available for the staging analysis. However, according to the inclusion criteria, all patients had to undergo imaging with all three modalities (CT, MR, and scintigraphy) and also had to have undergone bone aspiration or biopsy. An additional four patients were not eligible for the staging analysis because they did not undergo bone scintigraphy, and three more patients were ineligible because they did not undergo bone aspiration biopsy. One patient had negative bone biopsy results and positive bone scintigraphy results and was assigned a disease stage solely on the basis of imaging, leaving 88 patients who could be included in the analysis of staging. The mean difference in imaging time was 4.4 days between CT and MR, 4.7 days between CT and bone scintigraphy, and 4.4 days between MR and bone scintigraphy. Imaging studies were performed within 3 days of surgery in 84 patients and within 10 days of surgery in four patients.

Forty-five patients underwent surgery at the time of initial diagnosis. These included 38 of the 88 patients in the staging analysis cohort and seven patients who were excluded from the staging analysis but were eligible for the analysis of local determinants of stage. The reasons these seven patients did not qualify for the staging analysis were as follows: No bone scintigraphy was performed (n = 3), no bone marrow aspiration was performed (n = 3), or stage was assigned only on the basis of imaging findings (n = 1). These patients underwent surgery immediately after initial imaging

and did qualify for the analysis of extent of local disease.

There were 52 boys and 44 girls in the study population, which included 88 staging patients and 45 local-determinant patients; 38 patients qualified for both analyses. One of the 96 patients did not undergo bone scintigraphy and did not undergo surgery at the time of initial diagnosis and therefore did not qualify for either analysis. Approximately 79% were white, 13% were African American, 4% were Hispanic, and 4% were of other or unknown race. The median age was 32.2 months, and the age range was 9 days to 12.4 years; 75% of the patients were 4.1 years of age or younger.

The primary tumors were located in the adrenal glands in 50 patients (52%), in the abdominal sympathetic ganglia in 16 (17%), in the presacral region in five (5%), in the posterior mediastinum in 22 (23%), and in another location (not specified) in three (3%). The histologic nature of the tumors was undifferentiated neuroblastoma in 14 patients (15%), differentiated neuroblastoma in 44 (46%), neuroblastoma not otherwise specified in seven (7%), and ganglioneuroblastoma in 31 (32%).

Of the 88 patients in the staging cohort, 10 had stage 1 disease, 14 had stage 2 disease, 14 had stage 3 disease, and 50 had stage 4 disease. The final diagnosis in the 88 patients in the staging analysis was determined on the basis of positive bone marrow biopsy results ($n = 43$); positive pleural ($n = 3$), liver ($n = 2$), or distant lymph node ($n = 2$) biopsy results; or with surgical and pathologic findings in patients with negative bone marrow aspiration results ($n = 38$).

The initial CT and MR studies were read by one CT reader and one MR reader at the patient's institution and were then sent to be reread by one CT reader and one MR reader at each of the other five institutions, for a possible total of six CT readings and six MR readings per patient. All images were read at the patient's institution. However, images in some patients were not read at all of the other participating institutions. Therefore, some patients had fewer than six CT readings and/or fewer than six MR readings. On average, the CT images were read by 3.9 readers, and the MR images were read by 3.6 readers. There was no apparent pattern to the missing readings. They appeared to be missing at random, so adjustments were not made for bias. The absent readings were mostly problems of logistics (eg, a reader was unavailable when images arrived).

TABLE 2
ROC Results for Detection of Stage 4 Disease

Modality	Area under ROC Curve	95% CI for Area	Range of Areas	RMSE for Areas
Bone scintigraphy	0.83	0.74, 0.91	NA*	NA*
CT	0.81	0.74, 0.87	0.65–0.87	0.09
CT and bone scintigraphy	0.90	0.83, 0.96	0.77–0.95	0.06
MR imaging	0.85	0.79, 0.91	0.77–0.92	0.06
MR and bone scintigraphy	0.88	0.80, 0.95	0.81–0.92	0.04

Note.—CT and bone scintigraphy (0.90) was significantly better than CT alone (0.81), $P < .05$.
* NA = not applicable.

TABLE 3
Understaging of Stage 4 Disease at Imaging

Result	CT	CT and Bone Scintigraphy	MR Imaging	MR and Bone Scintigraphy	Bone Scintigraphy
Understaged	57	11	17	12	22
Correctly staged	43	89	83	88	78

Note.—Data are percentages.

Accuracy in Detection of All Stage 4 Disease

Fifty patients had stage 4 disease. Of the 50 patients, 43 had skeletal involvement and seven had nonskeletal involvement (involvement of the pleura in three, of the liver in two, and of distant lymph nodes in two). On the basis of the area under the ROC curves, MR imaging was slightly more accurate than CT in the detection of stage 4 disease; the areas under the ROC curves for CT and MR imaging were 0.81 and 0.85, respectively ($P = .06$) (Table 2). The true-positive values (sensitivity) for CT and MR imaging in the detection of stage 4 disease were 43% and 83%, respectively ($P < .01$), and the true-negative values for CT and MR imaging were 97% and 88%, respectively ($P > .05$). The area under the ROC curve for the 88 scintigraphy readings was 0.83. The addition of scintigraphy to both CT and MR imaging increased the areas under their respective ROC curves to 0.90 ($P < .05$ compared with CT alone) and 0.88 (not significantly different compared with MR alone).

When a simple proportion analysis was conducted, it was determined that neuroblastoma had been understaged at CT alone with a frequency of approximately 57% in patients with stage 4 disease (Table 3). With the addition of scintigraphic data, disease was understaged at CT in 11%. Disease was understaged at MR alone with a frequency of approximately 17% in patients with stage 4 disease. With the addition of scintigraphy, this value changed to 12%. The results show that MR alone is comparable to CT and scintigraphy and is superior to CT alone.

The interreader variability was greater at CT (RMSE, 0.09) than at MR imaging (RMSE, 0.06). An RMSE of 0.09 for CT indicates that CT readers have accuracies that differ by roughly 9%.

Diagnostic Value for Stage 4 Disease Involving Bone or Bone Marrow

Of the 50 patients with stage 4 disease, 43 had skeletal involvement at aspiration or biopsy. The diagnostic value of CT, MR, and bone scintigraphy in the detection of skeletal disease was analyzed separately with ROC curves.

The areas under the ROC curves for CT and MR imaging in the diagnosis of bone and bone marrow metastases were 0.59 and 0.86, respectively ($P < .05$) (Table 4). The corresponding area for bone scintigraphy was 0.85.

Errors in Diagnosis of Stage 4 Disease

A retrospective review of the false-negative diagnoses of stage 4 disease showed that disease was understaged at MR imaging in approximately 38% of patients with pleural, liver, or nodal involvement and in 15% of those with marrow replacement by tumor. Disease in 25% of patients with nonskeletal metastases and

TABLE 4
ROC Results for Detection of Bone and Bone Marrow Metastases

Modality	Area under ROC Curve	95% CI for Area	Range of Areas	RMSE for Areas
CT	0.59	0.54, 0.65	0.42–0.69	0.10
MR imaging	0.86	0.79, 0.92	0.77–0.92	0.06
Bone scintigraphy	0.85	0.76, 0.93	NA*	NA*

Note.—MR imaging (0.86) and bone scintigraphy (0.85) were significantly better than CT (0.59), $P < .05$.
* NA = not applicable.

TABLE 5
Predictive Values for CT and MR Imaging for All Stages

Final Stage	Stage 1 (%)	Stage 2 (%)	Stage 3 (%)	Stage 4 (%)
	Determined at CT			
1	39 (n = 34)	5 (n = 3)	2 (n = 2)	0 (n = 0)
2	39 (n = 34)	23 (n = 13)	12 (n = 13)	2 (n = 2)
3	3 (n = 3)	17 (n = 10)	25 (n = 27)	4 (n = 4)
4	17 (n = 15)	54 (n = 30)	61 (n = 65)	94 (n = 84)
	Determined at MR Imaging			
1	27 (n = 14)	25 (n = 12)	6 (n = 4)	1 (n = 2)
2	47 (n = 24)	34 (n = 16)	20 (n = 11)	2 (n = 4)
3	6 (n = 3)	19 (n = 9)	53 (n = 29)	7 (n = 11)
4	19 (n = 10)	20 (n = 10)	20 (n = 11)	90 (n = 146)

in 60% of those with skeletal metastases was understaged at CT.

Our study did not allow examination of the pelvis and femurs in patients with thoracic tumors, but this did not appear to have an adverse affect on lesion detection at CT. Of the total population, 22 patients had chest disease; seven (32%) of these had marrow disease on the basis of positive biopsy results. All seven patients had lesions of the ribs or thoracic spine depicted on scintigrams.

Predictive Accuracy for Stages 1, 2, 3, and 4

CT and MR imaging were approximately equally accurate in the prediction that patients had stage 4 disease (Table 5). For example, when CT results indicated stage 4 disease, the final diagnosis was stage 4 disease 94% of the time; for MR, this number was 90%. The agreement between the two modalities decreased in predictions of lower stages. When CT results indicated stage 3 disease, 61% of the time the disease was stage 4; for MR imaging, this number was 20%. For predictions of stage 2 disease, these numbers were 54% and 20%, respectively. Thus, when CT results indicated stage 2 or stage 3 disease, patients were nearly three times more likely to have stage 4 disease than when MR results indicated stage 2 or stage 3 disease. When either CT or MR predicted stage 1 disease, the chance that the patient had stage 1 or 2 disease was approximately 75%.

Overall, CT and MR imaging had statistically similar, but relatively poor, performance in the assessment of features of local disease that affected staging (Table 6). This may explain the poorer performance of CT and MR imaging in the staging of local disease versus the staging of distant disease.

DISCUSSION

Neuroblastoma is the most common small round blue cell malignancy in the pediatric population. Successful planning of an individual patient's therapy requires precise delineation of the local extent of neoplasm in soft tissues, lymph nodes, and solid organs and the evaluation of distant spread. CT, MR imaging, and bone scintigraphy have been the primary imaging modalities used in staging disease in children with neuroblastoma. Contrast-enhanced CT alone has been reported (15) to be 82% accurate in revealing tumor extent. The addition of bone scintigraphy reportedly increases the accuracy to 97% (15). In small series, MR imaging has also been shown to be sensitive in the detection of local disease (17,18).

To the best of our knowledge, ours is the first prospective multiinstitutional study to compare the performances of CT and MR imaging, alone and in combination with bone scintigraphy, in the overall staging of neuroblastoma. In this study, 69% of neuroblastomas arose in the abdomen or pelvis and 23% arose in thoracic sites, a finding that agrees with prior reports. The percentage of stage 3 and 4 tumors was high at approximately 73% (ie, 64 of 88 patients). However, the rate was equal to that reported in previous studies (1,2). Bone marrow involvement has been reported to range from 26% to 58% of cases (10). In our series, the frequency was 56%.

The imaging techniques used at the participating institutions were all state of the art. Recruitment patterns were similar, and there were no statistically significant differences in the stages of the neuroblastomas among the institutions. These facts suggest that the results are reproducible and can likely be generalized to other centers.

Our results show that MR imaging is more accurate than CT for detection of all stage 4 disease, with sensitivities of 83% and 43%, respectively. When only skeletal metastases are evaluated, MR imaging alone and scintigraphy alone yield similar estimated areas under the ROC curve for discrimination of distant spread to bone or bone marrow, with mean areas of 0.86 and 0.85, respectively. In comparison, the mean ROC curve area for CT in the staging of distant bone or bone marrow metastases is 0.59.

The considerable understaging of stage 4 disease at CT is not surprising because CT is known not to be a sensitive method for detecting small areas of cortical destruction. The finding of a poorer accuracy for CT supports the prevailing wisdom that MR imaging is better than CT for the detection of bone or bone marrow metastases. Although we recognized when we designed this study that MR imaging has been embraced as being superior to CT, this advantage had yet to be proved by direct comparison. Thus, we made a conscious decision to include it in our comparative analysis of imaging modalities.

Differences between ROC areas and simple sensitivities and specificities, as observed in this study, can occur because the area under the ROC curve takes into account discrimination among all cate-

TABLE 6
ROC Results for Detection of Local Extension

Local Extension	No. of Readings	True-Positive (%)	Area under ROC Curve	Area Range	RMSE for Areas	Positive Predictive Value (%)	Negative Predictive Value (%)
CT							
Tumor crosses midline	175	31	0.80	0.56–0.94	0.14	73	83
Nodal involvement	183	11	0.67	0.61–0.77	0.11	20	96
MR imaging							
Tumor crosses midline	164	32	0.79	0.73–0.86	0.09	81	79
Nodal involvement	173	12	0.74	0.67–0.83	0.07	19	99

gories or stages of disease (eg, the ability to differentiate among stages 1 through 4), whereas the simple proportions in this study aggregated the results, taking into account only the most clinically meaningful decision (ie, differentiation of stages 1–3 from stage 4).

One limitation of this study was that histopathologic proof, other than that obtained at aspiration and biopsy of marrow in the iliac crest, was not available for most lesions seen on MR images. Hence, the sensitivity and specificity estimates for MR imaging may be imprecise. However, this weakness in design is one confronted in most studies in which detection of bone metastases at imaging is evaluated. Although biopsy of individual lesions identified at imaging would be the most reliable method of establishing the diagnosis of metastatic disease, this is not possible for obvious reasons. The next best method to establish the nature of the lesions detected on an imaging study is to compare the findings with those of another study or to reevaluate the patients during the course of treatment. Thus, in support of the validity of our results is the finding that the results of MR and bone scintigraphy were similar.

A second limitation was that the MR examination assessed only limited areas of the marrow-bearing skeleton. Thus, the total amount of disease was not determined at MR imaging. This was based on our intent to determine if imaging can be used to help stage neuroblastoma; this design does not require that all lesions be identified. A third limitation is that gadolinium chelates were not used in the evaluation of the appendicular skeleton for marrow disease. Results of small series have suggested that contrast-enhanced MR imaging can facilitate identification of marrow lesions (22–25).

It also could be argued that the criterion used to detect metastatic disease at scintigraphy, namely, that there be two lesions, may have resulted in underdetection of bone or bone marrow disease. However, all patients in this study with stage 4 disease had more than two skeletal lesions at scintigraphy.

Another limitation in our study design was that it did not allow statistical evaluation of the performance of CT versus MR imaging for detecting metastases in the pelvis and proximal femora in patients with thoracic tumors. However, this comparison would have required more radiation exposure than we could have ethically justified. Patients did undergo MR imaging of these areas. However, this weakness of the study may not be as important as it appears. Seven (32%) of 22 patients with thoracic primary lesions in this series had bone or bone marrow metastases, and these were always in the field of view of the primary tumor. Hence, the failure to examine the pelvis and femurs with CT cannot explain the poorer lesion sensitivity of CT compared with MR imaging.

As regards the primary tumor, the intent of this study was not to evaluate the accuracy of detection of masses. Our study design did not allow for evaluation of the performance of CT and MR imaging versus that of ultrasonography (US) in the detection or staging of neuroblastoma. However, the relatively poor performance of US has been discussed in a prior report of a study that compared US with CT (15). Rather, because all of our patients had a mass detected at US or at chest radiography, the primary intent was to evaluate the abilities of CT and MR imaging in the staging of tumor.

The prevalence of determinants of local disease was relatively low in this study because patients who had extensive disease at time of entry underwent delayed surgery after induction chemotherapy. Because of the small number, positive predictive values were used to compare the accuracy of MR imaging and CT in local tumor staging (stages 1, 2, and 3). In this approach, the data were analyzed according to the way they are actually received and used by the referring clinicians. Although the numbers are small, the data suggest the following: *(a)* For stage 1 tumor, readers were more likely to stage abdominal extent correctly at CT than at MR imaging, where they often overstaged tumors; *(b)* for stage 2 and 3 tumors, readers of both CT and MR images were more likely to understage than to overstage tumor; and *(c)* for stage 2 tumor, readers were more likely to understage tumor at CT than at MR imaging. It was not the intent of this study to evaluate the relative accuracy of CT and MR imaging in revealing individual features of local extension (eg, midline extension, nodal involvement, vessel encasement).

Factors such as the relative costs and availability of these imaging modalities and physician preference undoubtedly will continue to influence whether MR imaging or bone scintigraphy is performed in an individual patient. The results of this study suggest that MR imaging can replace the combination of CT and bone scintigraphy for overall assessment of stage 4 disease in children with neuroblastoma. This approach would reduce the radiation exposure in these children and likely would reduce the number of sedation procedures required for staging. It also has the potential to decrease the use of hospital resources and the cost of imaging children with neuroblastoma.

Recent advances in MR imaging, particularly the development of fast imaging techniques, have markedly reduced imaging time without compromising image quality. Whole-body imaging with faster MR sequences can be performed in only a few minutes and has been shown to accurately depict metastases in adults (26–28). It is likely that further improvement in the results we have reported for MR imaging as a staging tool for neuroblastoma can be achieved with these recently developed techniques.

Author contributions: Study concepts and design, M.J.S., B.D.F., J.S.M., F.A.H., R.J.H., B.J.M.; literature research, M.J.S., B.D.F.; clinical studies, M.J.S., B.D.F., J.S.M., D.J.C., R.J.H., S.E.R.; data acquisition, all authors; data analysis/interpretation, M.J.S., H.I., D.J.C., B.J.M., B.A.S.; statistical analysis, H.I., D.J.C.; manuscript preparation, M.J.S., B.A.S.; manuscript definition of intellectual content, all authors; manuscript editing, M.J.S., B.A.S.; manuscript revision/review, M.J.S., B.A.S., D.J.C., B.J.M.; manuscript final version approval, all authors.

Acknowledgments: The authors thank Constantine A. Gatsonis, PhD, Center for Statistical Science, Brown University, for help with the study design. The authors also thank Thomas Caldwell, Cynthia B. Olson, MBA, MHS, Kathy Parkhurst, BA, Elaine Parkuris, and JoAnn Stetz, RN, RTT, from the American College of Radiology, Philadelphia, Pa, for data management. Finally, the authors thank Dennis Balfe, MD, and Jeffrey Brown, MD, from the Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, Mo, Michael DiPietro, MD, from C. S. Mott Children's Hospital, Ann Arbor, Mich, and John J. Crowley, MD, and Sam R. Kottamasu, MD, from Children's Hospital, Detroit, for reading the images, and the many research assistants and technologists who participated in this study.

References

1. Brodeur GM, Seeger RC, Barrett A, et al. International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. J Clin Oncol 1988; 6:1874–1881.
2. Kushner B, Cheung NK. Neuroblastoma: an overview. Hematol Oncol Ann 1993; 1:189–201.
3. Evans AE, DiAngio GJ, Propert K, et al. Prognostic factors in neuroblastoma. Cancer 1987; 59:1853–1859.
4. Brodeur GM, Pritchard J, Berthold F, et al. Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. J Clin Oncol 1993; 11:1466–1477.
5. Kiely EM. The surgical challenge of neuroblastoma. J Pediatr Surg 1994; 29:128–133.
6. Dietrich R, Kangarloo H, Lenarsky C, et al. Neuroblastoma: the role of MR imaging. AJR Am J Roentgenol 1987; 148:937–942.
7. Fletcher BD, Kopiwoda SY, Strandjord SE, Nelson AD, Pickering SP. Neuroblastoma: magnetic resonance imaging and tissue characterization. Radiology 1985; 155: 699–703.
8. Gilday DL, Ash JM, Reilly BJ. Radionuclide skeletal survey for pediatric neoplasms. Radiology 1977; 123:399–406.
9. Kornreich G, Horev C, Kaplinsky N, et al. Neuroblastoma: evaluation with contrast enhanced MR imaging. Pediatr Radiol 1991; 21:566–569.
10. Mills AE, Bird AR. Bone marrow changes in neuroblastoma. Pediatr Pathol 1986; 5:225–234.
11. Ng Y, Kingston J. The role of radiology in the staging of neuroblastoma. Clin Radiol 1993; 47:226–235.
12. Ruzal-Shapiro C, Berdon W, Cohen M, Abramson S. MR imaging of diffuse bone marrow replacement in pediatric patients with cancer. Radiology 1991; 181:587–589.
13. Shulkin BL, Shapiro B, Hutchinson RJ. Iodine-131-metaiodobenzylguanidine and bone scintigraphy for the detection of neuroblastoma. J Nucl Med 1992; 33: 1735–1740.
14. Siegel MJ, Jamroz GA, Glazer HS, Abramson CL. MR imaging of intraspinal extension of neuroblastoma. J Comput Assist Tomogr 1986; 10:593–595.
15. Stark DD, Moss AA, Brasch RC, et al. Neuroblastoma: diagnostic imaging and staging. Radiology 1983; 148:101–105.
16. Couanet D, Geoffray A, Hartmann O, et al. Bone marrow metastases in children's neuroblastoma studied by magnetic resonance imaging: advances in neuroblastoma research. Prog Clin Biol Res 1988; 271:547–555.
17. Sofka CM, Semelka RC, Kelekis NL, et al. Magnetic resonance imaging of neuroblastoma using current techniques. Magn Reson Imaging 1999; 17:193–198.
18. Tanabe M, Ohnuma N, Iwai J, et al. Bone marrow metastasis of neuroblastoma analyzed by MRI and its influence on prognosis. Med Pediatr Oncol 1995; 24:292–299.
19. Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979; 7:1–26.
20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.
21. Bickel PJ, Freedman DA. Some asymptotic theory for the bootstrap. Ann Stat 1981; 9:1196–1217.
22. Bloem JL, Reiser MF, Vanel D. Magnetic resonance contrast agents in the evaluation of the musculoskeletal system. Magn Reson Q 1990; 6:136–163.
23. Erlemann R, Vassallo P, Bongartz G, et al. Musculoskeletal neoplasms: fast low-angle shot MR imaging with and without Gd-DTPA. Radiology 1990; 176:489–495.
24. Ma LD, Frassica FJ, McCarthy EF, Bluemke DA, Zerhouni EA. Benign and malignant musculoskeletal masses: MR imaging differentiation with rim-to-center differential enhancement ratios. Radiology 1997; 202: 739–744.
25. Verstraete KL, Van der Woude HJ, Hogendoorn PCW, Deene Y, Kunnen M, Bloem JL. Dynamic contrast-enhanced MR imaging of musculoskeletal tumors: basic principles and clinical applications. J Magn Reson Imaging 1996; 6:311–321.
26. Eustace S, Tello R, DeCarvalho V, et al. A comparsion of whole-body TurboSTIR MR imaging and planar 99mTc-methylene diphosphonate scintigraphy in the examination of patients with suspected skeletal metastases. AJR Am J Roentgenol 1997; 169:1655–1661.
27. Horvath L, Burtness B, McCathy S, Johnson K. Total-body echo-planar MR imaging in the staging of breast cancer: comparison with conventional methods—early experience. Radiology 1999; 211: 119–128.
28. Johnson KM, Leavitt GD, Kayser HWM. Total-body MR imaging in as little as 18 seconds. Radiology 1997; 202:262–267.