Hypothesis: Metabolic imaging by positron emission tomography (PET) using [18F]fluorodeoxyglucose will be more accurate than anatomic imaging by computed tomography (CT) for detection of recurrent colorectal cancer. More accurate staging of recurrent tumor by PET will lead to more appropriate management decisions.

Design: Prospective blinded study comparing PET with CT, using histologic diagnosis, serial CT imaging, and clinical follow-up as criterion standards, with a fully blinded, retrospective reinterpretation of PET studies. Changes in diagnosis resulting from PET findings were correlated with subsequent treatment and surgical findings. Potential cost savings resulting from use of PET for preoperative staging were calculated.

Setting: Private practice in an outpatient tertiary referral center.

Patients: A group of 155 consecutive patients with imaging for diagnosis or staging of recurrent colorectal cancer. Twenty-one patients (14%) were excluded due to lack of a criterion standard. Computed tomographic scans were available for comparison for 115 patients.

Results: Positron emission tomographic scan sensitivity and specificity were 93% and 98%, respectively, compared with 69% and 96% for CT. Ninety-five percent confidence intervals for the differences between the modalities were 16% to 32% for sensitivity and 1% to 5% for specificity. The sensitivity of both modalities varied with anatomic site of recurrence. Positron emission tomographic scans were true positive in 12 (67%) of 18 patients with elevated serum carcinoembryonic antigen levels and negative CT findings. In 23 (29%) of 78 preoperative studies in which CT showed a single site of recurrence, PET showed tumor at additional sites. At surgery, nonresectable, PET-negative tumor was found in 7 (17%) of 42 patients who had PET evidence of localized recurrence only. Potential savings resulting from demonstration of nonresectable tumor by PET were calculated at $3003 per preoperative study.

Conclusions: Positron emission tomography was more sensitive and specific than CT for detection of recurrent colorectal cancer. Preoperative detection of nonresectable tumor by PET may avoid unnecessary surgery, and thereby reduce the cost of patient treatment.


POSITRON EMISSION tomography (PET) with [18F]fluorodeoxyglucose (FDG) has potential for demonstrating tumor metabolic activity before structural changes can be detected by computed tomographic (CT) imaging. Previous evaluations of FDG PET for diagnosis and staging of recurrent colorectal cancer have demonstrated higher sensitivity and specificity than CT.1-5 We now report a prospective evaluation of the accuracy of PET in 134 studies that were performed in patients with known or suspected recurrent colorectal cancer, together with assessment of the impact of PET findings on patient treatment and treatment cost.

RESULTS

PREVALENCE OF TUMOR RECURRENCE AND IMAGE ABNORMALITIES

One hundred fifteen patients underwent both PET and CT. The PET study was performed after CT in most patients, and the interval between PET and CT ranged from 0 to 56 days, with a median interval of 22 days. Validation procedures established a diagnosis of recurrent tumor at 157 sites in 101 patients and no tumor in 14 patients. Validated recurrence was found at 149 of 171 sites that were abnormal by

See Invited Critique at end of article

This article is also available on our Web site: www.ama-assn.org/surgery.
MATERIALS AND METHODS

PATIENT POPULATION

All 155 PET studies that were performed at the Northern California PET Imaging Center between October 27, 1992, and May 16, 1996, for evaluation of known or suspected recurrent colorectal cancer were included in the study. No independent diagnosis was available for 21 studies (14%), and these were excluded from analysis. Study subjects included 66 men and 68 women with a median age of 67 years (range, 31-93 years). Seventy-one primary tumors had been located in the ascending, transverse, or descending colon and 63 in the sigmoid colon or rectum. Median time of presentation after resection of the primary tumor was 22 months (range, 1 month to 8 years). Informed consent was obtained from all participating subjects.

An independent final diagnosis was obtained for 134 studies (86%) that were performed in 127 patients. One hundred twenty-one patients were studied once, 5 patients were studied twice, and 1 patient was studied 3 times. Excluded patients consisted of 6 patients who had less than 1 year of follow-up after studies with negative findings, 6 who were lost to follow-up, 5 who died without validation of sites of tumor involvement, and 4 who were treated by radiation or chemotherapy without further validation of findings. Computed tomography findings were available for comparison with PET findings in 115 patients.

INDICATIONS FOR PET IMAGING

Preoperative Staging

Seventy-eight studies were performed for preoperative assessment in patients who were under consideration for resection of recurrent tumor. In 76 cases, the patient had a localized CT abnormality at a single site that was considered to be potentially resectable.

Serum Carcinoembryonic Antigen Elevation

Thirty-two patients were being evaluated because of elevated serum carcinoembryonic antigen (CEA) levels. In 18 patients, prior CT findings had been negative. Fourteen patients had not been examined by CT.

Other Indications

Twenty-four studies were performed for other indications. Nine patients were undergoing restaging after surgery, 6 patients were thought to have nonresectable disease on the basis of multiple CT abnormalities, and 5 patients were being evaluated prior to radiation therapy. Four patients were studied because of symptoms or abnormal physical findings.

PET IMAGING

Imaging was performed with an ECAT EXACT 921 PET system (CTI Inc, Knoxville, Tenn). This system has an axial field of view of 16.2 cm, with intrinsic transverse resolution of 6.0 mm and axial resolution of 5.5 mm full width at half maximum. Direct and cross-plane data obtained from 24 detector rings were used to reconstruct 47 image sections, separated by 3.4 mm center to center. When images were reconstructed using a Hann filter with 0.4 cutoff, image resolution was approximately 10 mm full width at half maximum. Data acquisition, image reconstruction, and image display were performed using standard system software. Patients lasted for at least 4 hours prior to intravenous injection of FDG, 3.29 MBq/kg (0.14 mCi/kg), for an average dose of 0.37 MBq (10 mCi). Non–attenuation-corrected whole-body imaging at 4 minutes per bed position was commenced 30 minutes after injection. At the conclusion of whole-body imaging, 1 to 3 static 10- to 20-minute emission acquisitions were performed in regions of high suspicion. Transmission data were acquired for 10 to 20 minutes, depending on the activity of transmission sources. In the first 42 patients, transmission imaging was performed prior to tracer injection. In subsequent patients, transmission acquisition was performed at the end of the emission studies.

CT IMAGING

Computed tomographic imaging of the abdomen and pelvis had been performed prior to the PET study in 115 cases, usually at 1 of 7 different hospital radiology departments or outpatient imaging centers in the Sacramento region. The scanners in use at these sites were all GE 8900 or GE HiLight Advantage systems (GE Medical Systems, Milwaukee, Wis) during the first 18 months of the study. Subsequently, GE HiSpeed Advantage helical systems were used at 3 sites. The 1-cm contiguous-image sections were obtained in 2 to 4 seconds with the earlier scanners, and 7-mm sections were obtained in 1 second with the high-speed systems. Studies were performed after intravenous injection of 100 to 200 mL of contrast material. Standard scanning parameters were used.

PET IMAGE INTERPRETATION

All PET images were interpreted at the time of the study by 1 or 2 of us (P.E.V., E.A.C., T.R.P., M.K.H., or R.W.M.) who had access to clinical data. Images were viewed in axial, coronal, and sagittal planes using an interactive video display system. Since most PET studies were performed for further evaluation of CT findings, PET images were read in conjunction with CT images whenever possible.

PET, CT, or both and at 8 sites that were normal by both modalities.

A final diagnosis of no recurrence was established at 22 sites of image abnormality. Seventeen of these were abnormal by CT only, 3 were abnormal by PET only, and 2 were abnormal by both modalities. Five patients without recurrence who had normal PET and CT scans remained clinically disease free for more than 1 year after imaging.

DIAGNOSTIC ACCURACY OF PET AND CT

Sensitivity and Specificity

Sensitivity and specificity of PET and CT on a patient-by-patient basis are shown in Table 1, Table 2, and Table 3. Show sensitivity and specificity of PET and CT by site of recurrence, as well as the difference between the 2 mo-
Standardized uptake values (SUVs) and target-to-background ratios (T-B ratios) were determined for lesions that were more than 1 cm in diameter. The SUV is a semiquantitative measure of uptake that is obtained by dividing maximum activity detected in the lesion (micromicrocuries per milliliter) by injected dose corrected for body weight (microcuries per gram). Standardized uptake values were not calculated for smaller lesions because the recovery coefficient decreases rapidly with size below 1 cm. To obtain measurements that were not affected by small lesion size, only lesions that were clearly more than 1 cm in diameter in the PET images were included. Standardized uptake values and T-B ratios were also determined for 19 sites with false-positive CT findings.

To obtain tumor activity values for calculation of SUVs and T-B ratios, circular regions of interest 1 cm in diameter were placed over the region of increased uptake, centered on the point of highest visible activity, and the maximum pixel activity within the region was determined. To obtain background activity values for calculation of ratios, the region of interest was placed just outside the area of visible tumor uptake.

All images were reread at the conclusion of the study by 2 of 3 investigators (E.A.C., P.E.V., and M.K.H.) reading independently, without knowledge of patient identity, clinical data, CT findings, histologic data, or the results of the initial reading. Readings were subsequently compared and consensus was reached by discussion. At rereading, images were read by comparison of lesion activity to adjacent normal activity and were graded as positive or negative for tumor. Involvement of an organ or an anatomic region (eg, pelvis) was considered a single site of disease, regardless of the number of individual lesions. Each lung was considered as a separate site.

**CT IMAGE INTERPRETATION**

Since CT imaging had been performed at many different imaging facilities, CT scans were reinterpreted by 2 experienced readers (G.A.H. and H.B.G.), to ensure uniformity of interpretation. Computed tomographic images were read as positive or negative for each of the same anatomic regions as the PET scans. Readers were blinded to clinical, histologic, and PET data.

**VALIDATION OF IMAGING FINDINGS**

A final diagnosis was established at 171 sites that were positive for tumor by PET and/or CT. A further 8 tumor sites that were negative by PET and CT were confirmed surgically. The diagnosis was established histologically at 103 sites (57%), surgically at 93 sites, and by needle biopsy at 10 sites. Abnormal imaging findings at 45 sites (25%) were validated by demonstration of progression (38 sites) or no progression (7 sites) at a second CT imaging. In 7 patients with advanced disease, who died during the course of the study, clinical evidence of tumor progression was accepted as positive evidence at 20 sites of imaging abnormality (11%). Clinical evidence of absence of disease for 1 year or more after the PET study was accepted as negative evidence at 13 sites of imaging abnormality (7%) and in 5 patients who had normal PET and CT findings.

**DATA ANALYSIS**

For comparison of the accuracy of PET and CT, only the 115 studies in which patients had been imaged by both modalities were considered. All 134 studies were considered in evaluation of the effect of PET findings on patient treatment.

Means and SDs of SUVs and T-B ratios of lesions that were clearly greater than 1 cm in diameter were calculated. Mean SUVs and T-B ratios were also calculated for 19 sites with false-positive findings by CT. Analysis of variance was used to determine whether significant differences existed between mean SUVs and T-B ratios of tumor foci at different anatomic sites and between the false-positive CT sites and true sites of tumor. Where differences were detected, post hoc tests were used to identify sites of difference.

Sensitivity and specificity of PET and CT for the detection of tumor were calculated patient by patient and site by site. The differences in sensitivity and specificity between the 2 modalities were calculated. Ninety-five percent confidence intervals (CIs) were calculated according to the method described by Berry, to evaluate the precision of the difference values and to determine statistical significance. Sensitivity was also determined separately for sites where the final diagnosis had been established by histologic examination only.

**IMPACT OF PET FINDINGS ON SURGICAL TREATMENT AND MANAGEMENT COST**

The effect of PET findings on surgical treatment was determined by review of pre- and post-PET clinical records and discussion of those records with surgeons. Positron emission tomographic findings were correlated with subsequent surgical treatment and surgical findings. To determine the effect of the PET findings on the cost of patient treatment, it was assumed that recurrence at more than 1 site would be regarded as nonresectable. The cost of the surgical procedures that would be avoided was compared with the cost of PET imaging. These costs were determined from Medicare reimbursement rates for hospitalization, professional services (surgery and anesthesia), and pathologic examination of surgical specimens. Since there was no Medicare reimbursement for oncologic PET during the time of the study, PET costs were based on average reimbursement for whole-body PET examination at our center during the study period.
Differences were smaller than the differences in sensitivity, ranging from 1% to 6%. Only in the lungs was the difference statistically significant (Table 3). The difference was also significant when all sites were considered together. There were 5 false-positive PET results, including 3 in patients who proved to be free of tumor recurrence at all sites. Three of these false-positive foci were found in the pelvis. The positive predictive value of PET findings was 97% (146/151) on a site-by-site basis and 97% (96/99) on a patient-by-patient basis. The negative predictive value was 69% (11/16) patient by patient.

There were 19 false-positive CT findings, including 7 in patients who proved to have no tumor recurrence at any site. Eight occurred in the pelvis, 4 in the lungs, 2 in the abdomen, and 3 in the liver. Two patients had benign, enlarged adrenal glands. One false-positive finding in the pelvis was also false positive by PET; all others were negative by PET. The positive predictive value of CT findings was 85% (109/128) on a site-by-site basis and 93% (79/85) on a patient-by-patient basis. The negative predictive value was 24% (7/29) patient by patient.

**Blinded Reinterpretation of PET Images**

Blinded reinterpretation resulted in change in diagnosis in 9 (7%) of 134 studies at 9 (5%) of 178 sites of imaging abnormality.

In 5 patients who had been initially diagnosed as having recurrences at 2 or more sites, an additional site of tumor was noted at rereading. An independent diagnosis could not be established at these sites at the time of reinterpretation; however, all 5 patients were already known to have nonresectable tumors, so that the changes in interpretation would not have affected treatment decisions. In 3 patients who had no abnormality at the initial reading, solitary foci of abnormal activity were noted at rereading. In 1 of these patients, an initially false-negative site in the liver was correctly reinterpreted as positive, and this change could have affected treatment. The other 2 findings proved to be false positive on the basis of more than 1 year of disease-free follow-up. In 1 patient, an initial false-positive finding in the pelvis was correctly reinterpreted as representing bowel uptake.

**Table 1. Sensitivity and Specificity of Positron Emission Tomographic and Computed Tomographic Scans by Patients**

<table>
<thead>
<tr>
<th>Site</th>
<th>Positron Emission Tomographic Scan, No. (%)</th>
<th>Computed Tomographic Scan, No. (%)</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>54/57 (95)</td>
<td>48/57 (84)</td>
<td>11 (1 to 22)</td>
</tr>
<tr>
<td>Pelvis</td>
<td>30/31 (97)</td>
<td>21/31 (68)</td>
<td>29 (9 to 49)</td>
</tr>
<tr>
<td>Abdomen</td>
<td>22/23 (97)</td>
<td>13/23 (56)</td>
<td>31 (11 to 54)</td>
</tr>
<tr>
<td>Retroperitoneum</td>
<td>12/12 (100)</td>
<td>7/12 (58)</td>
<td>42 (10 to 74)</td>
</tr>
<tr>
<td>Lungs</td>
<td>16/17 (94)</td>
<td>16/17 (94)</td>
<td>0 (19 to 19)</td>
</tr>
<tr>
<td>Other†</td>
<td>12/12 (100)</td>
<td>4/12 (33)</td>
<td>67 (36 to 98)</td>
</tr>
<tr>
<td>Total</td>
<td>146/157 (93)</td>
<td>109/157 (69)</td>
<td>24 (16 to 32)</td>
</tr>
</tbody>
</table>

* CI indicates confidence interval.
* Other sites include hilar and mediastinum (n = 4), bone (n = 4), abdominal wall (n = 2), and adrenal glands (n = 2).

CT. Hepatic metastases were missed in 3 patients. Computed tomography findings were negative in 1 patient and positive in the other 2. There were false-negative PET findings at 6 sites in the abdomen. In one instance, a largely necrotic 4-cm lesion was diagnosed by CT but not by PET. The other 5 patients had diffuse peritoneal spread or small nodules of less than 1 cm that were negative by CT as well. One patient had a metastatic lung nodule of less than 1 cm that was negative by PET and positive by CT, and another had a 3-cm, PET-negative pelvic tumor mass that was positive by CT. Five of the 11 patients who had false-negative PET findings proved to have disease at a single site only. The other 6 false-negative findings occurred in patients who had other, PET-positive lesions.

There were false negative CT findings at 48 sites, 6 of which were also negative by PET. In 4 of 9 instances in which hepatic CT images were false negative for tumor, the images showed no abnormalities. In 5 cases, the CT images showed benign changes, including heterogeneous fatty infiltration in 2 patients, postoperative changes in 2 patients, and multiple small hepatic cysts in 1 patient; coexisting tumors were not diagnosed. In 8 of the 9 patients with false-negative CT results in the liver, the PET study was true positive. The largest number of false-negative CT findings occurred in the abdomen, pelvis, and retroperitoneum, where 30 (42%) of 71 disease sites were not detected. Twenty-one of these sites (70%) were positive by PET.

Positron emission tomography was more specific than CT at all sites except the retroperitoneum, but the differences were smaller than the differences in sensitiv-

<table>
<thead>
<tr>
<th>Site</th>
<th>Positron Emission Tomographic Scan, No. (%)</th>
<th>Computed Tomographic Scan, No. (%)</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>58/58 (100)</td>
<td>55/58 (95)</td>
<td>5 (1 to 10)</td>
</tr>
<tr>
<td>Pelvis</td>
<td>81/84 (96)</td>
<td>76/84 (90)</td>
<td>6 (2 to 13)</td>
</tr>
<tr>
<td>Abdomen</td>
<td>87/87 (100)</td>
<td>85/87 (98)</td>
<td>2 (2 to 6)</td>
</tr>
<tr>
<td>Retroperitoneum</td>
<td>103/103 (100)</td>
<td>103/103 (100)</td>
<td>0 (0 to 0)</td>
</tr>
<tr>
<td>Lungs</td>
<td>97/98 (99)</td>
<td>94/98 (96)</td>
<td>3 (1 to 5)</td>
</tr>
<tr>
<td>Other</td>
<td>103/104 (99)</td>
<td>101/104 (98)</td>
<td>1 (1 to 1)</td>
</tr>
<tr>
<td>Total</td>
<td>529/534 (99)</td>
<td>514/534 (96)</td>
<td>3 (1 to 5)</td>
</tr>
</tbody>
</table>

* CI indicates confidence interval.

©1999 American Medical Association. All rights reserved.
PET imaging was used to assess tumor recurrence in 78 patients referred specifically for preoperative assessment. Table 4 shows the standardized uptake values and target-to-background ratios for sites of tumor recurrence. Standardized uptake values and T-B ratios are shown by site of recurrence in Table 4. Only sites where there was a total of more than 10 lesions more than 1 cm in diameter were included. The mean ± SD SUV for all tumor sites was 6.17 ± 2.48, and the mean ± SD T-B ratio was 3.87 ± 2.08. The mean ± SD SUV at 19 sites of false-positive CT findings was 1.97 ± 0.81, and the mean ± SD T-B ratio was 1.26 ± 0.60. Analysis of variance of SUVs by tumor location showed no statistically significant differences between mean SUVs at different anatomic sites. Analysis of variance and post hoc tests showed that SUVs and T-B ratios of false-positive CT sites were significantly lower than SUVs and T-B ratios of tumor at any location (P < .001). Target-to-background ratios of liver lesions were lower than T-B ratios of tumor at all other sites (P < .001), as would be expected from the relatively higher uptake in normal liver compared with most other normal tissues.

There was overlap between the mean SUV of identifiable bowel, which was 2.91 ± 1.44 (range, 1.11-5.30), and the mean SUV of abdominal tumor, which was 6.35 ± 2.40 (range, 2.43-11.19). In these cases, diagnostic confusion was avoided by clear visual differences in the distribution of activity, which followed a characteristic linear pattern in the case of bowel and was focal in the case of tumor. There were 3 cases in which a false-positive diagnosis could have resulted from unrecognized focal bowel uptake.

**IMPACT OF PET FINDINGS ON PATIENT TREATMENT**

Preoperative Assessment of Tumor Recurrence

Seventy-eight patients were referred specifically for preoperative evaluation prior to attempted curative resection of recurrent tumor at a single site. In 76 patients, tumor recurrence had been demonstrated by CT. The PET findings in these patients are summarized in Table 5. In 47 of these patients (60%), PET imaging also showed a single, localized site of recurrence. In 23 patients (29%), PET demonstrated other, unsuspected sites of recurrence, and in 2 patients with hepatic metastases (3%), PET demonstrated extensive, bilobar disease, where only localized tumor had been detected by CT. Six patients (8%) had no evidence of tumor on PET. Figure 1 shows CT and PET images obtained in a patient with CT evidence of hepatic metastasis only, while PET correctly demonstrated retroperitoneal metastasis as well.

Surgery was undertaken in 42 (89%) of 47 patients with evidence of localized recurrence at a single site. At surgery, it was found that PET had underestimated the extent of tumor recurrence in 11 (26%) of 42 patients, and 7 of these patients (17%) proved to have nonresectable tumor. Two patients with known liver metastases also had diffuse peritoneal tumor and 1 had multiple small liver lesions that had not been detected by PET. Three patients with focal pelvic recurrence and 1 with focal abdominal recurrence on PET also had undetected diffuse tumor that could not be resected.

One patient each with hepatic, pelvic, abdominal, and pulmonary recurrence had more focal lesions than PET had demonstrated, which were then resected. Resection for cure was undertaken in 35 (83%) of 42 patients, representing 45% of the original group of 78 patients referred for preoperative assessment.

Thirteen of the 23 patients with PET evidence of tumor at multiple sites were subsequently evaluated at surgery, with confirmation of PET findings in 12 (92%) of 13 cases. One patient who had liver and lung abnormalities by PET and CT was found to have a resectable hepatic recurrence and a benign granulomatous lung lesion. In the other 10 patients, PET findings were confirmed by imaging and clinical follow-up. No CT-negative PET abnormalities proved to be false positive.

There were 6 patients with hepatic, pelvic, or abdominal CT abnormalities in whom PET showed no tumor. Subsequent surgery demonstrated a 3-cm pelvic tumor mass in one of these patients and a 3.5-cm hepatic metastasis in another. On PET imaging, the hepatic metastasis had shown lower FDG uptake than normal liver. Of the other 4 patients, 2 had no evidence of tumor at surgery, and all remained clinically tumor free at 14 to 32 months after PET.
Elevated Serum CEA Level

Thirty-two patients underwent PET imaging for investigation of serum CEA elevation. Eighteen had negative CT findings, and 14 did not undergo CT. Positron emission tomography findings were positive in 12 (67%) of 18 and 7 (50%) of 14 patients, respectively, and negative in the other 13. Thirteen patients had recurrence at 1 anatomic site only, but in 2 of these patients, tumor was extensive and not amenable to resection. Six patients had recurrences at multiple sites. Figure 2 shows true-positive PET findings, with false-negative CT findings in the liver of a patient with an elevated serum CEA level.

Nine patients with PET evidence of localized recurrence at a single site underwent surgery, and resection with curative intent was performed in 7 cases. One patient was found to have diffuse, nonresectable abdominal tumor as well as a PET-positive focal lesion, and 1 pelvic PET abnormality proved to be false positive. Three patients with PET evidence of tumor at multiple sites also underwent surgical evaluation, with confirmation of PET findings in each case. Follow-up for 1 year or more showed no recurrence in 11 of 13 patients with negative PET results. The other 2 patients subsequently underwent surgery 33 and 58 days after the PET scans and were both found to have extensive abdominal tumor recurrence that had not been demonstrated by PET.

In the patients who were studied because of elevated serum CEA levels, PET sensitivity was 90% (18/20) and specificity was 92% (11/12) patient by patient. The positive predictive value was 95% (18/19) and the negative predictive value was 85% (11/13).

Impact of Preoperative PET Imaging on Cost of Surgical Treatment

In a management algorithm in which recurrence at more than 1 site is treated as nonresectable, the PET findings in the 78 preoperative patients would translate into avoidance of surgery for nonresectable tumor in 25 patients (32%). Surgery would be avoided in 14 patients with hepatic, 4 with pelvic, 3 with abdominal, 2 with
Positron emission tomography was significantly more sensitive than CT for detection of recurrent tumor, with a difference of 24% (95% CI, 16% to 32%). Positron emission tomography was also more specific than CT, with a smaller difference of 3% (95% CI, 1% to 5%). The largest difference between the 2 modalities was found in the abdomen and pelvis, where more than one third of sites that were true positive by PET were false negative by CT. Computed tomography also had 10 false-positive findings in the abdomen and pelvis, compared with 3 by PET. In the liver, the difference between the modalities was smaller, at 11% (95% CI, −1% to 22%), and in some cases resulted from difficulty in differentiating benign from malignant abnormalities by CT. In the lungs, both modalities failed to detect 1 metastatic lesion, but CT had more false-positive findings.

Although PET is more sensitive than CT, PET sensitivity is also limited by minimum detectable lesion size. Detectability of tumor by PET depends on 2 factors: tumor size and uptake of FDG. In our study, PET failed to detect 7 lesions that were less than 1 cm in diameter and 4 lesions that were larger. One of these larger lesions had only a narrow rim of viable tumor around a necrotic center while the other 3 were solid and showed low tracer uptake. Also, in 5 patients with PET-positive abdominal or pelvic recurrence, PET underestimated the extent of tumor. Each of these patients had diffuse peritoneal or pelvic spread that was missed by PET and by CT.

Physiologic tracer accumulation in normal structures, particularly the intestine and urinary tract, is another possible source of error in interpretation of PET images. In 3 of our patients, false-positive findings in the pelvis may have represented focal bowel activity, and in 1 of these cases, the pelvic activity was correctly read as intestinal uptake at reinterpretation; however, in most cases, bowel activity was identifiable by its characteristic linear pattern. Artifacts resulting from high bladder activity can interfere with interpretation of non–attenuation-corrected pelvic images, but we did not find this to be a problem when attenuation correction was used. Urteric activity also did not interfere with interpretation.

Our findings regarding the accuracy of PET and CT are similar to those obtained by other investigators. Schepers et al12 compared PET findings with those obtained by conventional imaging, which included CT and ultrasound imaging of the liver, CT of the pelvis, and chest radiography. Reported sensitivities of PET and conventional imaging were 94% and 85%, respectively, for liver metastasis, and 93% and 60% for pelvic recurrence. Also, PET demonstrated unsuspected tumor at other sites in 10 (13%) of 76 patients.

Delbeke et al3 found in 61 patients that PET was more accurate than CT for detection of liver metastases (92% vs 78%) and extrhepatic metastases (92% vs 71%). Altogether, PET detected unsuspected metastases in 17 (28%) of 61 studies. Similarly, Lai et al3 demonstrated unsuspected extrhepatic tumor in 11 (32%) of 34 patients with known hepatic metastases.

An unbiased determination of the accuracy of PET and CT cannot be obtained from studies of this type, in which many patients are selected for PET imaging because of positive CT findings. This case-selection bias results in overrepresentation of positive CT findings, both true and false, with consequent overestimation of CT sensitivity and underestimation of specificity. However, study of a biased population of this type is appropriate when the purpose is to determine the impact of adding a new modality to the patient treatment algorithm and does demonstrate which test is more accurate in the relevant clinical context.

A more difficult issue in quantifying sensitivity and specificity in disseminated malignant disease arises from problems in verification of imaging findings.13 Often there are no means of detecting tumor sites that are nonsymptomatic and that are not detected by either imaging modality, so that the findings of the more sensitive technique essentially provide their own criterion standard. This results in high values for sensitivity, such as the values that were reported for CT when it was first introduced14 and the values that we are finding for PET at the present time. Specificity may also be overestimated, since it is not possible to verify all abnormalities in a patient with a disseminated tumor.

It is desirable to use histologic diagnosis as the only reference test in measuring the performance of a diagnostic modality, but this can lead to another type of verification bias in the study population.13 When patients have clinical and imaging evidence of extensive recurrence, it is usual for only one lesion, and sometimes no lesions, to be verified pathologically. If sites of

Table 6. Net Savings Resulting From Use of Positron Emission Tomographic Scans for Preoperative Staging

<table>
<thead>
<tr>
<th>Cost/Savings, $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings from avoided surgery</td>
</tr>
<tr>
<td>Total cost of positron emission tomographic studies</td>
</tr>
<tr>
<td>Net savings from avoided surgery</td>
</tr>
<tr>
<td>Net savings per patient</td>
</tr>
</tbody>
</table>
pathologically nonverified lesions are excluded from the study, positive PET and CT findings at the corresponding anatomic sites will be underrepresented, thereby reducing the measured sensitivity of both modalities in these regions. To reduce this bias, we included results at sites where a reference diagnosis was obtained by clinical and imaging follow-up, as suggested by Begg and McNeill.15

These forms of bias affected the evaluations of both imaging modalities equally, without biasing the comparison of PET and CT results. There were other possible sources of bias that could have favored PET over CT, including the interval between CT and PET imaging, unequal skills in test performance, variations in CT technology, and bias in test interpretation.

**AS A RESULT** of the interval between the CT and PET examinations, the tumor was an average of 24 days more advanced by the time of the PET study. This imaging sequence favored PET but accurately reflected the results that could be expected when PET is used for preoperative evaluation of patients with known CT abnormalities. Some degree of inequality in PET and CT performance was also likely, since the study was conducted in a PET imaging center, and the investigators had no control over the performance of the CT examinations. The PET operators were involved in the evaluation of a new imaging technology, while the CT examinations were performed as routine procedures at multiple clinical sites.15,16

The findings did not reflect the maximum CT performance that was technologically possible, since only some of the examinations were performed with helical CT scanners. Indeed, when Vitola et al17 compared PET and helical CT in 24 patients with hepatic recurrence, they found a smaller difference in sensitivity of hepatic tumor detection (90% vs 86%) than we did; however, the increase in CT sensitivity was accompanied by a decrease in specificity to 58%, indicating that there was increased detection of benign as well as malignant lesions. Positron emission tomography also detected unsuspected extrahepatic tumor in 17% of their patients. Thus, it seems that existing technologic improvements in CT could narrow the gap between the modalities but would not be likely to eliminate it.

Availability of CT images at the time of initial PET interpretation was a potential source of interpretation bias. In the preoperative patients, this was an essential part of the algorithm that was under study, according to which PET was used for further evaluation of CT abnormalities. Only negative CT images were available in the patients who were being studied for elevated CEA levels. This study was initially undertaken as a prospective field study of PET imaging rather than a fully controlled comparison of the 2 modalities. Field studies that observe conventional clinical practices are more likely to be directly generalizable than studies in a controlled environment.19 A controlled comparison of PET and CT was achieved by the retrospective, fully blinded reinterpretation of images.

In our evaluation of the impact of PET on management decisions, preoperative PET imaging correctly demonstrated unsuspected tumor at other sites in 29% of patients and more extensive nonresectable tumor at the same site in 3%. In a treatment algorithm according to which recurrence at more than 1 site is treated as nonresectable, this would translate into avoidance of nonindicated surgery by demonstration of nonresectable tumor in 32% of patients. At Medicare reimbursement rates, the cost of avoided surgery would exceed the cost of PET studies, resulting in net savings of $3003 per patient.

Similar findings were obtained in 3 previous studies that evaluated the effect of PET findings on management decisions. Beets et al17 found that PET findings affected treatment in 15 (43%) of 35 patients with liver or pelvic recurrence and demonstrated unsuspected, nonresectable tumor in 6 (25%) of 24 preoperative patients. In staging suspected liver metastasis only, Vitola et al1 found that PET altered surgical plans in 6 (25%) of 24 patients, avoiding surgery in 3 patients (13%). In the study by Lai et al,4 PET avoided inappropriate hepatic resection in 11 (32%) of 34 patients.

In published reports of surgical findings, approximately 40% to 50% of patients who were thought to have resectable hepatic metastases and 30% to 40% of patients with limited local recurrences by CT were found to have nonresectable tumor at surgery.12,18 In our study, PET reduced the proportion of patients with nonresectable tumor at surgery to 17%. Whether PET will also reduce the rate of rerecurrence by reducing the rate of undetected residual tumor after surgery remains to be determined.

In the past, serum CEA monitoring of patients with colorectal cancer has been regarded as controversial19; however, practice guidelines determined by the American Society of Clinical Oncology now recommend that CEA levels be monitored postoperatively if resection of liver metastasis would be clinically indicated.20 Similarly, the National Comprehensive Cancer Network recommends monitoring serum CEA levels to evaluate patients who have an elevated serum CEA level preoperatively and whose level has decreased after surgery.21

When serum CEA is used for patient monitoring, positive results should be further assessed by the most accurate imaging modality available, and our findings suggest that this modality should be PET. Positron emission tomography findings were true positive in 67% of patients with elevated serum CEA levels and negative CT findings. A whole-body PET study is more expensive to perform than a CT study, but PET may reduce the overall cost of patient treatment by reducing the cost of repeated imaging after false-negative results, as well as the cost of unnecessary surgery.

All findings published to date have shown that PET imaging with FDG is more accurate than CT for diagnosis and staging of recurrent colorectal cancer. This and other studies have demonstrated the effectiveness of PET for determining resectability of recurrent tumor, and our results suggest that PET will also be more effective for initial diagnosis of recurrence.
Val et al are certainly to be congratulated for their fervent enthusiasm for this new privately held technology. The skeptic might assume that since these private physicians are based in a PET imaging center, they would bias a study by design to favor PET scan; however, it is important to put aside all such potentially cynical prejudices in favor of academic interpretation of a large number of patients carefully assessed by a variety of means and numerous statistical tests. It is particularly important to do so in light of the scope of this problem. Colorectal cancer is the second most common noncutaneous malignant neoplasm, anticipated to affect approximately 165,000 people this year. It can also be anticipated that more than 40% of patients, despite seemingly curative resections, will experience a recurrence of their disease. The most common sites of recurrence are the pelvis, liver, and lungs. Over the past several decades, numerous new technologies have been introduced in an attempt to better identify and therefore more appropriately treat patients with recurrences in each of these areas. The primary tool for both preoperative and postoperative evaluation of patients with rectal carcinoma is rectal ultrasonography. The potential problem with rectal ultrasound is that whether the focal length is 3.0, 7.0, 7.5, or 10.0 MHz, the best clarity is in the layers of the rectal wall as opposed to within the rectum itself. Local recurrences tend to occur not at the suture line or anastomosis but actually in the pelvis. Therefore, any intraluminal evidence of tumor tends to show the tip of the iceberg rather than the major nidus of tumor.

The most common extrapelvic and potentially resectable tumor site is the liver. Finlay and McArdle and others have demonstrated very high rates of sensitivity and specificity of ultrasonography for liver metastasis. More recently, intrathoracic ultrasonography by either laparotomy or laparoscopy has enabled avoidance of unnecessary hepatic resection in many patients. Even with multiple bilobar lesions, a 5-year survival rate of 35% can be anticipated, provided that a curative liver resection can be accomplished. In addition to CT scan and extracorporeal and intraoperative hepatic ultrasonography, other modalities exist to evaluate the presence of any liver lesions. These tests include liver-spleen scan and, more recently, radioimmunoguided surgery. The main use of the latter technology is for detection of unresectable deposits, such as retroperitoneal tumor spread and periporal tumor involvement, both of which would preclude attempted curative hepatic resection. Lung metastasis can also be treated for cure, with an anticipated survival rate only slightly worse than that after resection of liver metastases. The optimal test for detection of these pulmonary metastases is regarded by some thoracic surgeons as helical CT scan.
These authors have elected to compare their privately held technological investment with a single modality: CT scan. There are multiple problems with this type of study design, some of which have been mentioned by the authors, and others of which have been overlooked. I would like to address these problems on an individual basis. First, the study has been designed to include as the criterion standard positives not only histopathologically proven tumor but also clinical evidence of tumor progression (11%), clinical evidence of absence of disease for 1 year or more after the PET study was accepted as negative evidence (7%), and demonstration of progression (38 sites) or no progression (7 sites) at a second imaging. Although a final diagnosis was established in 171 sites, in 78 of those sites, no histopathologic confirmation was ever obtained. Thus, the denominator of the study is skewed in favor of the interpretive design of the study. Therefore, the authors should have eliminated from their denominator all of the patients in whom histopathologic confirmation was not obtained. Had they done so, I imagine that their sensitivity rate for the remaining 93 patients would have approximated that of CT scan.

Second, the follow-up of patients in whom “clinical evidence of absence of disease for 1 year or more” was accepted was suspect. For example, slow tumor growth could have well accounted for a patient who had a 1-cm recurrence 8 years after surgery and still had a 1-cm recurrence 9 years after surgery. Had that patient been evaluated at 10 years, the tumor might have continued to slowly enlarge; 1 year is an inappropriately short interval. As a much more important issue, the length of follow-up after attempted curative surgery for recurrent disease was not even mentioned. This factor is crucial to ascertain the true accuracy of the PET scan. For example, a patient may have undergone an attempted curative pelvic exenteration for isolated pelvic disease and remained disease-free for 6 months, after which hepatic recurrence may have been noted. Therefore, the length of follow-up should have been supplied along with rigorous postresectional imaging studies to confirm the authors’ conclusions. Was re-recurrence defined as only clinically evident or was it vigorously sought?

Third, the authors have made an assumption in their definitions that if more than a single site was involved by tumor, that patient was “unresectable.” As a surgeon, I would certainly challenge this assumption. Two of many possible examples include bilateral lobe peripheral lung metastases 5 years after initial resection and a true anastomotic recurrence with a single perianastomotic tumor focus 2 years after the initial operation. The authors should have been more specific in their discussion of the positive sites and should have included all potentially resectable patients, as jointly judged by a surgeon and radiologist and not as judged in isolation by a radiologist.

Fourth, although the authors have accepted some culpability for the mean 22-day delay between CT scan and PET scan, I feel that this problem is minor relative to other flaws in the study design. I do not think that within 3 weeks, any major tumor growth would occur, and if such growth indeed were so rapid, it would most likely render the patient incurable. Once again, the problem is that the parameters seem to have been set forth exclusively by radiologists, without consultation from a surgical oncologist, colorectal surgeon, or general surgeon.

Fifth, in a similar light, the authors have assumed in their cost analysis that surgery would be performed based on a positive PET scan result without surgeons having performed other tests. Perhaps the surgeons in Sacramento are so enamored and convinced of this technology that they would not require additional studies, such as CT portography, radioimmunoguided surgery, diagnostic laparoscopy, or other modalities; however, most surgeons would indeed be loath to operate based on a single positive study result. Accordingly, the cost of those additional confirmatory investigations should have been factored into their cost analysis.

Sixth, one of the most major problems was the inequivalence in performance of the CT scan. It is well accepted in surgery that the more one performs a procedure, the better the results will be. Heald et al have quite convincingly and repeatedly demonstrated local recurrence rates after curative resection of rectal carcinoma well below 5%. By adoption of these meticulous surgical techniques and high-volume procedure performance, others have been able to reproduce these laudable results. Furthermore, the authors have compared their obviously carefully selected equipment, their standardized technique, and their impressive interpretative skills with a myriad hybrid of scans performed at an unknown number of institutions by countless radiologists and technicians. Obvious differences may exist in terms of thickness of the “cuts”; use or lack of use of intravenous, oral, or rectal contrast; and performance of additional cuts to investigate any questionable areas, contingent on the level of interest and expertise of the radiologist. In point of fact, the radiologist is often guided by a sometimes inadequate amount of clinical information supplied by the surgeon and/or oncologist. In every instance, these investigators had available to them not only their own machine and technicians using a standardized approach, but also the previously performed CT scans in every instance. Although they state that this study design is appropriate, they also admit the potential source of interpretation bias. Clearly, a more accurate study would be designed to include randomization of patients to undergo 1 of the 2 studies first, followed by the other. In both instances, the scan would be performed by “experts” using the best and newest possible machines with the highest resolution using standardized techniques and blinded to the results of the other study.

In terms of the results, the authors tout the 93% sensitivity of PET scanning against the 69% sensitivity of CT scan. They gloss over the fact that the specificity rates of the 2 studies (98% and 96%, respectively) were virtually identical. Once again, because of flaws in the study design, the sensitivity of PET scanning for the liver, pelvis, abdomen, and retroperitoneum might not have been better compared with other modalities, such as magnetic resonance imaging, transanal rectal ultrasound, radioimmunoguided surgery, or CT portography. Much in the way that the Krook et al study had an unacceptably high rate of local recurrence in the surgery alone group, the CT specificity rate shown in Table 1 was among the worst I have seen, undoubtedly due to the problems I have previously cited.

Additional problems with the results include the fact that in 11 (26%) of 42 patients, resection for cure was undertaken but found to be impossible. That cost should clearly be factored out of their cost/savings equation. The 69% negative pre-
dictive value should also be discussed as clearly less than ideal. Furthermore, in 30% of PET scan–negative patients, tumor was apparently found in the pelvis, abdomen, or retroperitoneum. This group also needs to be subtracted out of the cost/savings equation. Last, in 5 of 11 sites that were deemed negative by PET scan, the tumors were judged to be true positive by CT scan. This portion of the discussion was again not highlighted within the cost/savings equation. A greater problem is that the cost analysis is really based on Medicare reimbursement and therefore is related exclusively to local charges and in no direct way to cost. A discussion of relative value units or a distillation of indirect and direct costs would be much more meaningful in the current context of health care economics than the new reporting of Medicare reimbursement rates. Medicare reimbursement is of course exclusively related to charges but not to cost. Therefore, although throughout this article the authors used the word cost, they are in fact presenting discounted charges.

Based on this analysis, PET scan may well have a role in selected patients with suspected recurrence after attempted curative surgery for colorectal carcinoma. A prerequisite might be the availability of radiologists with the level of expertise of the authors. Regardless, I do not feel that PET scan should be widely adopted without more convincing evidence of true cost savings, or before comparison with many of the other modalities. The authors have elected to use their expertise against a single, poorly controlled modality rather than against any or all of the other tools that are currently available and in widespread use for the identical purpose of detection of recurrent disease. To their credit, they admit to this problem in the “Comment” section, where they mention the study of Vitola et al., in which PET scan was compared with helical CT scan, and the sensitivity rates were 90% vs 86%, respectively. Hopefully, more appropriately designed prospective randomized trials will better define the role of this promising new technology.

Steven D. Wexner, MD
Fort Lauderdale, Fla