A Meta-Analysis of the Literature for Whole-Body FDG PET Detection of Recurrent Colorectal Cancer

Ralf H. Huebner, Kenneth C. Park, James E. Shepherd, Judy Schwimmer, Johannes Czernin, Michael E. Phelps, and Sanjiv S. Gambhir

Crump Institute for Biological Imaging; Department of Molecular and Medical Pharmacology, Division of Nuclear Medicine; and Department of Biomathematics, University of California, Los Angeles, School of Medicine, Los Angeles, California

A meta-analysis of the literature for the use of FDG PET in the detection of recurrent colorectal cancer (CRC) was conducted to evaluate the quality of the reported studies. Overall values for the sensitivity and specificity of whole-body FDG PET and an overall FDG PET–directed percentage change in management were also determined through this analysis. **Methods:** Guidelines to evaluate the articles were formulated on the basis of the U.S. medical payer source criteria for assessing studies that report information on usage of new medical technology. A meta-analysis was conducted using methodology described in the peer-reviewed literature. **Results:** On the basis of the guidelines established for our review, the availability of necessary information for assessing the reliability of the FDG PET data for diagnosing recurrent CRC was less than ideal. Through a meta-analysis of 11 articles, we determined, within a 95% confidence level, an overall sensitivity of 97% (95% confidence level, 95%–99%) and an overall specificity of 76% (95% confidence level, 64%–88%) for FDG PET detecting recurrent CRC throughout the whole body. Furthermore, through pooling of the change-in-management data, an overall FDG PET–directed change in management was calculated to be 29% (95% confidence level, 25%–34%). **Conclusion:** Our review suggests that improvements can be made to more effectively report the results of these FDG PET studies. The overall values determined through the meta-analysis indicate the potential benefits of using FDG PET as a diagnostic or management tool. Furthermore, these values should prove to be useful to assess the cost-effectiveness of using FDG PET in the management of patients with recurrent CRC.

Key Words: FDG PET; colorectal; cancer; meta-analysis; recurrent

J Nucl Med 2000; 41:1177–1189

Colorectal cancer (CRC) is the fourth most prevalent carcinoma in the United States and continues to be a major health problem worldwide. In 1998, there were 131,600 new cases of CRC (*1*). Of these, approximately 56,500 patients with CRC died, making CRC the second most frequent cause of death from cancer in the United States (*1*). About 70% of the patients with CRC who do not succumb to primary CRC are suitable for potentially curative surgery of their primary CRC. However, after surgical intervention, CRC recurs in approximately 40% of these patients within 2 y (*2–4*).

On the basis of conventional presurgical staging techniques, such as CT, approximately 30% of patients with recurrent CRC have apparently limited recurrent CRC (*5*). However, it has been shown that only 25% of these patients with recurrent CRC are actually curable by surgery (*6*). In looking at hepatic metastases only, over 50% of patients are found to have nonresectable disease at surgery (*7,8*). Furthermore, the 5-y disease-free survival rate after attempted curative resection is only 20%–40% (*7,9*). These facts show that many of these patients with recurrent CRC must harbor unrecognized tumor foci, making the cancer noncurable through surgical resection. Therefore, to prevent most of these patients from undergoing unnecessary surgery, it is essential to improve the accuracy of preoperative detection of recurrent tumors.

Whole-body FDG PET may provide a more accurate means of determining whether patients with recurrent CRC are indeed suitable for curative resection surgery. FDG PET localizes tumors by identifying cells in the body that have higher levels of FDG uptake and trapping relative to metabolically normal cells (*10–12*). Also, in some situations, metabolic activity can be detected by FDG PET before structural changes can be diagnosed by anatomic imaging techniques (e.g., CT, MRI, radiography) (*13*).

FDG PET continues to gain acceptance in clinical oncology. In many centers, FDG PET is becoming an essential tool in the diagnosis and management of many different cancers (*13*). Studies have shown the potential cost-effectiveness of using FDG PET imaging in the management of these cancers (*3,14–18*). Before a detailed cost-effectiveness analysis study can be performed for FDG PET in recurrent CRC, the ability of FDG PET imaging to detect recurrent CRC must be determined.

Received May 19, 1999; revision accepted Nov. 1, 1999.

For correspondence or reprints contact: Sanjiv S. Gambhir, MD, PhD, Crump Institute for Biological Imaging, University of California, Los Angeles, School of Medicine, A-222B CIBI, 700 Westwood Plaza, Box 951770, Los Angeles, CA 90095-1770.

As a starting point for our meta-analysis, a set of guidelines was created to review the currently published studies on FDG PET in recurrent CRC. We reviewed each article on the basis of our guidelines for ideal reporting of results from an FDG PET imaging study. For each article, we determined the availability of necessary information for assessing the results' reporting reliability. We should emphasize that our study did not attempt to determine or question the validity and reliability of the actual data content within the articles but to evaluate the method and extent of available data reporting and presentation. After the review, overall values for the sensitivity and specificity of FDG PET in detecting recurrent CRC and percentage change in management resulting in its use in patients with recurrent CRC were determined. A meta-analysis was necessary because of the difficulty in assessing the potential usefulness of FDG PET for patients with recurrent CRC on the basis of the individual results of these studies. Meta-analysis is a process that allows the results of different studies to be combined. The methodology for meta-analysis has been well defined in the literature (*19–21*), and its relevance to diagnostic imaging studies has been shown (*22,23*). By integrating studies, the results of the individual studies can be strengthened and overall estimates for the values of interest can be provided.

MATERIALS AND METHODS

Article Selection

A literature search was performed using the MEDLINE database for January 1990 through April 1999. It should be noted that other potential sources and databases for articles in this area were considered but not used because of the relatively comprehensive nature of the MEDLINE database.

The literature search through MEDLINE incorporated 3 search strategies. Strategy 1 included PET, positron, positron emission, or fluorodeoxyglucose as keywords; these were matched with colorectal as a keyword. Strategy 2 used FDG and colon as keywords. Strategy 3 matched the keywords FDG and rectal. A total of 81 articles were produced from the 3 search strategies.

The main selection criteria for choosing articles for this review were the availability of FDG PET imaging results of recurrent CRC or change-in-management data on patients with CRC being monitored for recurrence after their primary CRC surgical resection. Any article that did not present this information was excluded. Furthermore, only FDG PET studies that focused on postsurgical CRC patients were included. Articles providing only a general review were excluded. Special attention was given to all general review articles to verify the uniqueness of this study's literature review. All general review articles differed substantially from our study in both purpose and reviewing methodology. Published abstracts for presentation at national meetings were also excluded because they lack the necessary details of the presented data and the study's methodology. Only articles written in English were used in this study.

A total of 11 articles were included in our meta-analysis on the basis of this criteria (*2–4,24–31*). One of the studies [Ruhlmann et al. (*30*)] included in our meta-analysis was evaluated after direct contact with 1 of the study's authors, to clarify the presentation of their text and table results.

Literature Review

U.S. medical payer (e.g., Blue Cross, Blue Shield, Kaiser Permanente) source criteria for assessing studies that report information on the use of new medical technology (*32–34*) were used as a basis for creating guidelines to evaluate the current literature on FDG PET in recurrent CRC. The shared premise was that studies must present enough information to determine the reliability of the results. In evaluating the results, it should have been clear if any factors outside the tested technology had an influence on the final results. Our guidelines were modified to address specific issues of reporting FDG PET imaging results in recurrent CRC for the purposes of a meta-analysis and for any future cost-effectiveness studies.

Guidelines

Guideline 1: Description of Study Design and Patient Selection Criteria. Inherent biases can exist in different study designs, making it necessary to describe the type of study used. There must also be an explicit description of the patient selection criteria used to gather the study population.

1. Study design. The type of study that was conducted should be clearly presented (e.g., prospective, randomized, controlled, blinded). Multicenter, randomized control studies may be ideal but are not always practical when assessing the role that FDG PET may play in management algorithms.
2. Patient selection criteria. The criteria used to select the population of patients with CRC for the study should be clearly explained. The study may accrue referral or sampling bias during this selection process. Furthermore, if a study is selecting patients for specific types of recurrent CRC, such as hepatic involvement only, this should be clearly stated. It is likely that FDG PET imaging performance differs in various regions of the body.
3. Exclusion of patients from the study's final analysis. Any patients who were excluded from the final analysis of the study should be clearly explained. The ability to exclude patients from a study represents a potential way for a researcher to directly affect the final study results.

Guideline 2: Characteristics of the Patient Population Ultimately Studied. These include any characteristics of the patients that could influence the final reported imaging results. These are also characteristics that may be related to, or have an effect on, the distribution or prevalence of recurrent CRC in the study population. As more information is gathered, these characteristics may become indications for recurrent CRC.

1. Mean age with range and gender. The risk of developing CRC increases in people older than 40 y. Although there is currently no connection between the probability of recurrent CRC and age, the age group that the study is investigating should be known for future reference. Similarly, the distribution of men and women should be stated even though recurrent CRC has not been shown to be related to sex.
2. Comorbid conditions. Any patient with CRC in the study population who suffers from comorbid conditions should be identified and the disease should be described. For example, inflammatory disease may lead to misinterpretation (false-

positive [FP]) of FDG PET scans because of increased FDG uptake at sites of inflammation.

3. Location of primary CRC. The location of the primary tumor should be stated. Studies have shown that there is a relationship between the location of the primary CRC and the location of recurrent CRC (*35–37*).
4. Special institution characteristics. Although special institution characteristics are not directly a feature of the patient characteristics, they might reflect important information about the study population. For example, a study conducted in a tertiary-care center will generally have patients with more extensive and defined health problems than a study conducted in a primary-care center. These issues are categorized as centripetal biases (*38*).

Guideline 3: Patient Indications Leading to FDG PET Use. Patient indications differ from patient characteristics because indications are the medical evidence for suspicion of recurrence and are the reason for referral to FDG PET imaging.

1. Reasons for the use of FDG PET imaging. The clinical test results or diagnostic tool findings (e.g., elevated carcinoembryonic antigen levels, abnormal radiologic scan) used to refer the patient for further evaluation of recurrent CRC using FDG PET must be stated.
2. Reasons for use of FDG PET imaging corresponding to specific FDG PET findings. Not only should the initial reasons for FDG PET be stated, but if there are several reasons in a single study for referring patients for FDG PET, they should be presented so the specific patient indication and final FDG PET finding can be compared (note: the reasons for items 1 and 2 are covered in guideline 7).
3. Stage of the primary CRC. The stage of the primary CRC should be stated for each patient in the study. The stage of the primary tumor is related to the likelihood of recurrence and the extent of recurrent disease (*35–37*).

Guideline 4: Details of the Technologies Used During the Study and Image-Interpretation Issues. With the rapid advancement in medical imaging today, it is necessary to specify the parameters of the imaging modalities used in the study along with any special issues concerning the interpretation of the scans. For example, with many different CT scans available (with different resolutions and tumor-detection capabilities), to merely state that CT was performed is not adequate.

1. Imaging techniques used in the study and their resolutions. The imaging modalities used in the study should be explicitly described. In addition, the resolution parameter(s) should be stated. If comparisons between 2 different techniques or studies are being made, it should be clear whether potential differences in the results were caused by incomparable resolutions or other characteristics of the techniques themselves.
2. Patient preparation. Special methods for preparing the patients for the imaging modality should be described (e.g., time of fasting for FDG PET). The preparation of the patients can directly influence the quality and reliability of the imaging results. Also, when comparing studies, these methods need to be carefully considered.
3. Enhancement of imaging technique. It should be noted whether attenuation correction was used for FDG PET. Studies tend to agree that attenuation correction is important for quantifying tracer uptake (*39,40*). However, it is debatable whether attenuation correction is necessary at a clinical level that relies mainly on detection of lesions (*40*). For CT, it should be noted if the study was conducted using a contrast agent.
4. Explanation of special characteristics of the interpreters. To understand any potential interpreter bias or variability, the number of interpreters and their experience in reading FDG PET scans should be stated. It should also be stated whether the interpreters of the FDG PET imaging results were blinded to the initial indications or other imaging findings. This information must also be shared for interpreters of other imaging modalities used in the study.
5. Definition of positive and negative FDG PET imaging findings. The way the interpreters of the FDG PET scan defined the imaging findings should be explained (e.g., positive, equivocal, intermediate, negative). Also, it must be stated whether the definitions were based on the subjective opinion of each interpreter or if quantitative methods such as standardized uptake values (SUVs) and target-to-background ratios were used (*41*). These factors represent the study's threshold. Because these imaging studies often operate at different thresholds, it is important to consider this when combining results of different imaging studies.
6. Additional scans. Any patient in the study population who underwent additional FDG PET scanning or other imaging interventions should be described. In this description, the reason for the additional scan (e.g., technical error, more information needed) should be clear. In addition, if patients with additional scans were used in the final results, it should be understood which of the scans were used, which ones were excluded, and why.

Guideline 5: Final Diagnostic Confirmation. Imaging results are limited by the final confirmation technique used to verify the imaging findings.

1. Final confirmation. The technique used to verify the FDG PET imaging findings or any other imaging intervention used in the study should be described. Reference standard bias may exist if many different confirmation techniques were used to verify the same set of imaging results.
2. Association between a specific FDG PET finding and the final confirmation technique used. Not only should the confirmation methods be explicitly described, each specific imaging finding needs to correspond to a specific final confirmation method.
3. Histopathologic confirmation. Histopathologic confirmation is considered the gold standard for confirming imaging study results. However, to say a finding is confirmed histopathologically is not sufficient and needs further clarification (e.g., by surgery, by biopsy), because the accuracy of these techniques can vary.
4. Nonhistopathologic confirmation. When histopathologic confirmation cannot be used, the other techniques performed should be explained. For confirmation by clinical follow-up (e.g., carcinoembryonic antigen–level monitoring over time, serial radiologic imaging), the amount of time spent following up on the patient and the diagnostic tools used should be described.

5. Patients lost to clinical follow-up. All patients who are lost during follow-up should be mentioned, and it should be clear as to whether their imaging results were included in the final analysis.

Guideline 6: Sensitivity and Specificity Data.

1. True-positives (TPs), FPs, true-negatives (TNs), and false-negatives (FNs). TP, FP, TN, and FN values need to be quantitatively described in a table for all imaging interventions analyzed in the study. This is necessary to show how the sensitivity and specificity values were determined. Listing the values also helps to prevent errors that may occur when trying to explain all the findings in the text and allows for a clear, straightforward presentation of the essential imaging results. If an imaging intervention is interpreted through a nonbinary system, an appropriate table needs to be included.
2. Qualitative explanation(s) of FPs and FNs. Errors made by the imaging modality (FP, FN) need to be explained. There are no fixed definitions for these values. This makes it necessary for authors of each study to provide qualitative explanations of how the FP and FN findings were defined in terms of the confirmation technique used.
3. Specific region studied by FDG PET. When a study reports FDG PET imaging findings in a specific region of the body, the location needs to be clearly stated. It must be clear whether the imaging results were calculated according to a specific anatomic location or if the results were calculated for the whole body. For example, when FDG PET is evaluated in the liver region, a patient who tests TN in the liver but has several malignant lesions outside the liver can be considered TN.
4. Number and location of recurrences in lesions or patients. The prevalence of disease in the study population should be clearly described or evident in the final results. As the prevalence of disease increases, the positive predictive value of a diagnostic intervention increases, whereas the negative predictive value decreases. Interpretation of the imaging results may be misleading if the prevalence of disease is not known. Also, it should be clear if only patients with a specific site of recurrence are being studied.
5. Confidence intervals. Confidence intervals should be included in all applicable values reported. This allows for an understanding of the uncertainty level of the imaging results.
6. Equivocal findings. In situations in which equivocal FDG PET findings exist, researchers should be clear as to how the findings were defined and if they were used in the final analysis of the results.
7. Data reported in patients and in corresponding lesions. Complete studies should have the imaging findings described both on a per patient and a per lesion basis.

Guideline 7: Change-in-Management Information. Change-in-management information for FDG PET is clinically important to assess the actual impact this imaging modality has on the population of patients with CRC being monitored for recurrent CRC.

1. FDG PET–directed change in management. Management changes that are a direct result of FDG PET imaging should be explicitly described. If a study evaluates change in management in the patient population, it should be clear which patients underwent these management changes.
2. Diagnostic tool used to make initial treatment decision. The diagnostic tool(s) and corresponding result(s) that were responsible for making the initial treatment decision must be described. In doing so, both the strengths and weaknesses of FDG PET can be assessed compared with other modalities.
3. Diagnostic tool used to make final treatment decision. The diagnostic tool and corresponding result responsible for making the final treatment decision must be clearly described.
4. Medical treatment decision made both initially and after FDG PET. The medical treatment decision made initially and after FDG PET should be explicitly stated. This information is necessary to evaluate, in terms of medical treatments, the impact FDG PET has on the management decisions for patients with recurrent CRC.
5. Correct or incorrect FDG PET–directed management change. If the final decision was based on FDG PET, the researcher(s) should state whether this decision was ultimately correct or incorrect. Because the accuracy of FDG PET is not 100%, this modality will inevitably lead to some incorrect management decisions.
6. Upstaging and downstaging. The extent of disease diagnosed before FDG PET and the extent of disease depicted by FDG PET must be clearly presented. Only then can the change-in-management data be summarized as upstaging or downstaging the extent of disease in a patient.
7. FDG PET in the management algorithm. Ultimately, it must be determined whether FDG PET can be of benefit to patients being monitored for recurrent CRC after primary CRC resection. The role that FDG PET might play in the management algorithm should be discussed and supported by the management data presented in the study.

Literature Analysis

All articles were evaluated for adherence to each item of the 7 guidelines. An objective score of "adequate" (A), "partial" (P), "not addressed" (N), or "not applicable" (N/A) was given for how thoroughly each article addressed each guideline item. An A score signified that an item was described sufficiently, whereas a P score indicated an incomplete presentation of an item. For the items in our guidelines that were relevant to the study's objective but not included in the article, a score of N was awarded. A score of N/A was assigned when a study did not focus on a particular aspect of our guidelines. For example, a study that did not conduct a change-in-management analysis was given an N/A score for the items in guideline 7. Again, it is important to understand that the results of this review do not reflect the reliability of the actual data presented but the amount of available information provided by each article to describe the data compared with what we feel would be ideal.

Every article was independently assessed by 5 reviewers. Each reviewer was initially unaware of the evaluations made by the others. The reviewers were aware of the authorship of the articles. After all of the articles were evaluated, the reviewers compared their results. For situations in which there was not 100% concordance among reviewers, a discussion was conducted to decide on a consensus score. Table 1 presents the scores agreed on by all 5 reviewers. The percentages of As, Ps and Ns were calculated for each article independently, as well as for each guideline across all articles.

A summary of the adherence scores for each article can be seen

in the percentages in the bottom 3 rows of Table 1. These values represent the total number of each score (A, P, N) divided by the difference of total number of items and the number of N/A scores for each article. The percentages in the last column of Table 1 represent a summary of the adherence scores to each guideline, across all articles. These values were calculated by taking the total number of each score (A, P, N) and dividing them by the difference of total number of scores given for each guideline and the number of N/A scores.

Meta-Analysis

Relevant FDG PET imaging results from all studies were included in the calculations for the overall sensitivity and specificity. These data were categorized into 4 groups: whole body in patient data, hepatic in patient data, hepatic in lesion data, and local/pelvic in patient data. Three methods were used to determine an overall value for sensitivity and specificity for each of these groups. First, we attempted to fit each set of data to a summary receiver-operating-characteristic (sROC) curve on the basis of methods found in Littenberg and Moses (*42*). An sROC curve is used when the slope of the linear regression is within a prespecified range (−0.5 to 0.5). When applicable, the mean threshold for each group of studies was determined and the sensitivity and specificity at that point on the curve were provided. Overall values were also obtained by pooling of datasets, along with determining weighted averages for each of these sets of data.

The datasets were pooled by adding the TP, FP, TN, and FN results from all relevant studies and finding the sensitivity and specificity for the combined data. A 95% confidence interval was constructed for these estimates by assuming that each of the sensitivity and specificity results were a simple proportion from a normal distribution (*43*). For example, the estimate of the variance of sensitivity equals the following: sensitivity × (1 − sensitivity)/ number of truly positive patients. The 95% confidence interval would then be calculated as: sensitivity ± 1.96 × the square root of the variance. Overall weighted values for sensitivity and specificity were calculated for comparison with the results of the pooled data.

Statistical tests were used to determine if the change-in-management data could be properly summarized by pooling the data from several studies. A χ^2 test was used to check the probability that all proportions of management change were drawn from the same distribution (*44*).

For results that were <5% probability that all proportions were from a single distribution, we re-examined the article containing the data that gave the largest error from that expected by the pooled proportion to understand why the article was significantly different and to determine whether we would include this study in our final percentage change-in-management calculation.

To calculate an overall value for the percentage change in management made by FDG PET, the data that were shown statistically to be from the same distribution were pooled and variance was calculated as a binomial proportion with enough samples to approximate a normal distribution.

RESULTS

Literature Review

Table 1 summarizes the consensus scores of all reviewers for the articles included in this review. When combining the scores of all articles for each guideline, only guideline 7 received an overall A adherence score of >75%. Guideline 7, the management-information guideline, was adequately adhered to by 80%, partially by 12%, and not addressed by 8%. Guidelines 2, 3, 4, and 6 all scored a greater percentage of N items than P items.

Along with the summary of adherence to each guideline across all articles, percentages of A, P, and N scores for each article were also determined. The range of A adherence scores for each article was 36% (10/28) (*28*) to 86% (30/35) (*2,3,25*), with a mean of 66%. The P adherence scores for each article had a mean of 13%, with a range from 0% (*25*) to 37% (13/35) (*31*). The mean percentage for N items for each article was 21%, with a range of 6% (2/35) (*31*) to 43% (*28*). Three out of the 11 articles received a percentage A adherence score of >75% (*2,3,25*). All articles had more N items than P items, except for the article by Flamen et al. (*31*).

Meta-Analysis

The sensitivity and specificity of FDG PET for recurrent CRC from each article are presented in Table 2. The overall combined sensitivity and specificity are given in Table 3 for the whole body in patient data, the liver in patient data, the liver in lesion data, and the local/pelvic region in patient data. These values were calculated by direct pooling of the data and by determining a weighted average. Ninety-five percent confidence intervals are included in Table 3 for the overall values determined by pooling the data. No valid results were obtained from the sROC analysis. In 10 of the 11 studies, authors quantitatively presented descriptive details of the imaging results (*2–4,24–28,30,31*). The study by Beets et al. (*29*) was the only study that did not quantitatively describe the imaging results because the study focused only on change-in-management data. The sensitivity and specificity of FDG PET depicting CRC recurrences in the whole body, liver (hepatic involvement), and local/pelvic areas were considered for our meta-analysis.

In 5 of the studies, authors reported sensitivity and specificity data for FDG PET depicting CRC recurrences in the whole body in patient data (*2,3,24,27,30*). The number of patients included in this group was 281. It should be noted that the patient group included from the article by Valk et al. (*3*) consisted only of the 115 patients who underwent both CT and FDG PET. As in the other 4 studies, the data for these patients were based on detection of recurrent tumor throughout the entire body. The data for the total patient population (127 patients) were based on totaling the results of individual organ sites. The method of combining the individual organ-site data to ascertain values for the performance of FDG PET in the whole body was not consistent with the methods used in the other 4 studies. The total sensitivity and specificity in the whole body of these patients by pooling the data within a 95% confidence interval were 97.0% (95% confidence level, 95%–99%) and 75.6% (95% confidence level, 64%–88%), respectively. The overall weighted averages for the sensitivity and specificity in the whole body were 97.1% and 77.12%, respectively. These weighted averages both fell within the range of the 95% confidence interval of the pooled data. No information from the sROC

TABLE 1
Review of Studies with Formulated Guidelines

Guideline	Author name (reference no.)											Guide-line percent-ages
	Beets et al. (29)	Schiepers et al. (28)	Vitola et al. (27)	Lai et al. (26)	Delbeke et al. (24)	Keogan et al. (25)	Ogunbiyi et al. (4)	Ruhlmann et al. (30)	Flanagan et al. (2)	Valk et al. (3)	Flamen et al. (31)	
Guideline 1: description of study design and patient selection criteria												
1. Study design	P	P	P	A	P	A	A	P	P	A	A	A-72.72%
2. Patient selection criteria	P	A	A	A	A	A	A	A	A	A	A	P-24.24%
3. Exclusion of patients from study's final analysis	P	A	A	A	A	A	A	N	A	A	A	N-3.03%
Guideline 2: characteristics of patient population ultimately studied												
1. Mean age with range and gender	N	P	A	N	A	A	A	A	A	A	A	A-38.64%
2. Comorbid conditions	N	N	N	N	N	N	N	N	N	N	A	P-6.82%
3. Location of primary CRC	A	N	N	N	N	A	N	A	A	A	A	N-54.54%
4. Special institution characteristics	N	N	N	A	N	N	N	P	N	A	P	
Guideline 3: patient indications leading to use of FDG PET												
1. Reasons for use of FDG PET imaging	A	P	A	A	A	A	A	A	A	A	A	A-51.52%
2. Reasons for use of FDG PET imaging correlated to specific FDG PET findings	A	N	P	P	N	A	P	A	A	A	P	P-15.15%
3. Stage of the primary CRC	N	N	N	N	N	N	N	A	A	N	N	N-33.33%
Guideline 4: details of technologies used during study and image-interpretation issues												
1. Imaging techniques used in the study and their resolutions	P	P	A	A	A	A	P	P	A	A	P	A-69.7%
2. Patient preparation	A	A	A	A	A	A	A	N	A	A	A	P-10.61%
3. Enhancement of imaging technique	A	A	A	N	A	A	A	A	A	A	P	N-19.7%
4. Explanation of special characteristics of the interpreters	N	N	N	A	A	A	A	N	P	A	A	
5. Definition of positive and negative FDG PET imaging findings	N	N	A	A	A	A	A	N	N	A	A	
6. Additional scans*	A	N	A	A	N	A	A	A	A	N	A	
Guideline 5: final diagnostic confirmation												
1. Final confirmation	A	A	A	A	A	A	A	A	A	A	P	A-72.73%
2. Association between specific FDG PET finding and the final confirmation technique used	P	A	P	N	P	A	P	A	A	P	P	P-21.82%
3. Histopathologic confirmation	A	P	A	P	A	A	A	A	A	A	P	N-5.45%
4. Nonhistopathologic confirmation	A	N	A	A	A	A	A	P	A	A	P	
5. Patients lost to clinical follow-up confirmation†	A	N	A	A	A	A	A	A	A	A	A	
Guideline 6: sensitivity and specificity data												
1. TP, FP, TN, FN	N/A	A	A	N	A	A	A	A	A	A	P	A-71.43%
2. Qualitative explanation of FP and FN	N/A	A	A	A	A	A	A	N	A	A	A	P-5.71%
3. Specific region studied by FDG PET	N/A	A	A	A	A	A	A	A	A	A	A	N-22.86%
4. Number and location of recurrences in lesions and/or patients	N/A	P	A	A	A	A	A	A	A	A	A	
5. Confidence intervals	N/A	N	N	N	N	N	N	N	N	A	N	
6. Equivocal FDG PET findings‡	N/A	A	A	A	A	A	A	N	A	A	P	
7. Data reported in patients and in corresponding lesions	N/A	N	A	N	A	A	N	N	A	A	P	

TABLE 1 (*Continued*)

Guideline	Author name (reference no.)											Guideline percentages
	Beets et al. (*29*)	Schiepers et al. (*28*)	Vitola et al. (*27*)	Lai et al. (*26*)	Delbeke et al. (*24*)	Keogan et al. (*25*)	Ogunbiyi et al. (*4*)	Ruhlmann et al. (*30*)	Flanagan et al. (*2*)	Valk et al. (*3*)	Flamen et al. (*31*)	
Guideline 7: change-in-management information												
1. FDG PET directed change in management	A	N/A	A	A	A	N/A	A	N/A	N/A	A	A	A-79.59%
2. Diagnostic tool used to make initial treatment decision	A	N/A	P	P	A	N/A	A	N/A	N/A	A	P	P-12.24%
3. Diagnostic tool used to make final treatment decision	A	N/A	A	A	A	N/A	A	N/A	N/A	A	A	N-8.16%
4. Medical treatment made both initially and after FDG PET	A	N/A	A	A	A	N/A	A	N/A	N/A	A	P	
5. Correct or incorrect FDG PET–directed management change	P	N/A	N	N	N	N/A	N	N/A	N/A	P	A	
6. Upstaging and downstaging	A	N/A	A	A	A	N/A	A	N/A	N/A	A	A	
7. FDG PET in management algorithm	A	N/A	A	A	A	N/A	A	N/A	N/A	A	A	
Percentage of scores for each article	A-57.14% P-21.43% N-21.43%	A-35.71% P-21.43% N-42.86%	A-68.57% P-11.42% N-20.00%	A-62.86% P-8.57% N-28.57%	A-71.43% P-5.71% N-22.86%	A-85.71% P-0% N-14.29%	A-71.43% P-8.57% N-20.00%	A-53.57% P-14.29% N-32.14%	A-78.57% P-7.14% N-14.29%	A-85.71% P-5.71% N-8.57%	A-57.14% P-37.14% N-5.71%	

P = partial adherence; A = adequate adherence; N = not addressed; CRC = colorectal cancer; TP = true positive; FP = false positive; TN = ; FN = false negative; N/A = not applicable.

*"A" when no additional scans were obtained.

†"A" when no patients lost to follow-up.

‡"A" when no equivocal FDG PET findings.

TABLE 2
Reported Prevalence, Sensitivity, and Specificity

Author (reference no.)	Type	Patients/ lesions/studies	n	Prevalence	Sensitivity	Specificity
Schiepers et al. (*28*)	Local/pelvic*	Studies	83	57% (47/83)	96% (45/47)	97% (34/35)
	Hepatic	Studies	83	42% (35/83)	94% (33/35)	100% (48/48)
Vitola et al. (*27*)	Whole body	Patients	24	79% (19/24)	95% (18/19)	80% (4/5)
	Hepatic	Lesions	55	71% (39/55)	90% (35/39)	100% (16/16)
Lai et al. (*26*)	Hepatic	Patients	34	82% (28/34)	100% (28/28)	67% (4/6)
Delbeke et al. (*24*)	Whole body	Patients	61	90% (55/61)	98% (54/55)	83% (5/6)
	Hepatic	Lesions	127	82% (104/127)	91% (95/104)	96% (22/23)
Keogan et al. (*25*)	Local/pelvic	Patients	18	72% (13/18)	92% (12/13)	80% (4/5)
Ogunbiyi et al. (*4*)	Hepatic	Patients	58	40% (23/58)	96% (22/23)	100% (35/35)
	Local/pelvic	Patients	47	45% (21/47)	90% (19/21)	100% (26/26)
Ruhlmann et al. (*30*)	Whole body	Patients	59	78% (46/59)	100% (46/46)	69% (9/13)
Flanagan et al. (*2*)	Whole body	Patients	22	68% (15/22)	100% (15/15)	71% (5/7)
Valk et al. (*3*)	Whole body	Patients	115	88% (101/115)	95% (96/101)	79% (11/14)
	Hepatic	Patients	115	50% (57/115)	95% (54/57)	100% (58/58)
	Local/pelvic	Patients	115	27% (31/115)	97% (30/31)	96% (81/84)
Flamen et al. (*31*)	Local/pelvic*	Patients	103	33% (34/103)	94% (31/33)	100% (69/69)
	Hepatic	Patients	103	45% (46/103)	98% (45/46)	100% (57/57)

*One indeterminant finding not used in calculation.

analysis was obtained because a valid sROC curve was not described with these data.

The imaging results of FDG PET in the liver were reported in 7 articles. In 5 studies, authors reported the data in number of patients (*3,4,26,28,31*). The total number of patients in this category was 393. The 83 FDG PET studies of Schiepers et al. (*28*) that were conducted using 76 patients were considered to be 83 patients in this total because there was no description of why some patients had additional FDG PET scans. The overall sensitivity and specificity of recurrences in the liver within a 95% confidence interval were 96.3% (95% confidence level, 93.6%–99.0%) and 99.0% (95% confidence level, 97.7%–100%), respectively, by pooling of data. The overall weighted averages for the sensitivity and specificity of this dataset were 96.0% and 97.1%, respectively. Both weighted averages were within the range of the 95% confidence interval of the pooled data. No information from the sROC analysis was obtained because a valid sROC curve was not described with these data. In 2 studies, authors reported FDG PET hepatic results in lesions (*24,27*). In 182 liver lesions, the combined sensitivity and specificity within a 95% confidence interval were 90.9% (95% confidence level, 86%–96%) and 97.4% (95% confidence level, 92%–100%), respectively, by pool-

TABLE 3
Meta-Analysis of Sensitivity and Specificity Data

Type	Calculation method	Patients/ lesions	n	TP	FP	TN	FN	Combined sensitivity (95% confidence interval)	Combined specificity (95% confidence interval)
Whole body	Pooled data	Patients	281	229	11	34	7	97.03% (94.87%–99.20%)	75.56% (63.00%–88.11%)
	Weighted average							97.13%	77.12%
	sROC curve not applicable							—	—
Hepatic involvement	Pooled data	Patients	393	182	2	202	7	96.30% (93.6%–98.99%)	99.02% (97.67%–100%)
	Weighted average							96.04%	97.12%
	sROC curve not applicable							—	—
Hepatic involvement	Pooled data	Lesions	182	130	1	38	13	90.91% (86.20%–95.62%)	97.44% (92.48%–100.00%)
	Weighted average							90.86%	96.97%
	sROC curve not applicable							—	—
Local/pelvic	Pooled data	Patients	366	137	5	214	8	94.48% (90.77%–98.2%)	97.72% (95.74%–99.7%)
	Weighted average							94.71%	97.25%
	sROC curve not applicable							—	—

ing the data. The weighted average for the sensitivity was calculated as 90.9%, whereas the weighted average for specificity was found to be 97%. Again the weighted averages for sensitivity and specificity were within the range of the 95% confidence interval of the pooled data. No information from the sROC analysis was obtained because a valid sROC curve was not described with these data.

Finally the imaging results in local/pelvic recurrences were considered in 5 articles (*3,4,25,28,31*). From a total of 366 patients and within a 95% confidence interval, an overall sensitivity of 94.5% (95% confidence level, 90.8%–98.2%) and specificity of 97.7% (95% confidence level, 95.7%–99.7%) were determined by pooling the data. The 83 FDG PET studies of Schiepers et al. (*28*) that were conducted with 76 patients were also considered to be 83 patients in this total. Overall weighted averages were 94.7% for the sensitivity and 97.3% for the specificity. Both of these values were within the range of the 95% confidence interval of the pooled data. Again no information from the sROC analysis was obtained because a valid sROC curve was not described with these data.

Table 4 is a summary of the change-in-management decisions made by FDG PET that were ultimately correct. Table 5 provides a detailed description of these FDG PET management decisions described as upstaging or downstaging. Seven of the 11 articles had a portion of the study dedicated to FDG PET change in management (*3,4,24,26,27,29,31*). The percentage change-in-management determined from the data in the article by Flamen et al. (*31*) had the largest error from that expected by the pooled proportion. After a thorough evaluation of this study, we decided to include the results in our final analysis because of the similar patient population characteristics and study methodology. The data from these 7 studies were pooled to determine an overall percentage management change of 29% (82/246). At a 95% confidence level, the range of this value fell between 25% and 34%.

DISCUSSION

On the basis of our review, the current studies that focus on FDG PET in recurrent CRC sufficiently provide the information necessary for future cost-effectiveness studies.

TABLE 4
Percentage Change in Management

Author (reference no.)	No. of patients	Change in management
Beets et al. (*29*)	35	40% (14/35)
Vitola et al. (*27*)	24	25% (6/24)
Lai et al. (*26*)	34	29% (10/34)
Delbeke et al. (*24*)	52	33% (17/52)
Ogunbiyi et al. (*4*)	23	44% (10/23)
Valk et al. (*3*)	78	31% (24/78)
Flamen et al. (*31*)	103	20% (21/103)
Pooled management change	349	29% (102/349)
95% confidence interval		25%–34%

However, although enough information was presented overall, these studies failed to report many of the ideal items in our formulated guidelines, suggesting that improvements in reporting these imaging results can be made. In future cost-effectiveness studies, it will be necessary to validate these missing items through sensitivity analysis. When reviewing the results of our literature review, the reader should understand that the adherence scores are not an indication of the validity of the conducted studies. Rather, the scores are a representation of the presence or absence of items we feel are necessary when reporting these types of imaging studies for the purposes of a cost effectiveness analysis.

As an example, consider guideline 2, which had the highest percentage of N items. With 55% of the items not addressed across all articles, the concern is which items are not being reported. In this guideline, comorbid disease in the study population was not addressed by all articles except for the 1 by Flamen et al. (*31*). Before conducting FDG PET, the practitioner should know if any of the patients have comorbid diseases that might have an effect on the imaging study (see guideline 2, item 2 for reasoning). If an article states that the patients studied are all postsurgical CRC patients, readers cannot assume that these patients lack other diseases. Likewise, to infer from abnormal imaging results that concurrent diseases existed before FDG PET imaging is not very dependable.

When evaluating the reliability of these studies, there should be nothing that has to be assumed or inferred. The problem with most of these studies is that, often, relevant information is either omitted or not given the appropriate attention or mention within the study. For example, in guideline 3, which had the second-most N items, only 2 of the studies stated the stage of the primary CRC. As explained in the Materials and Methods section, as the stage of the primary CRC becomes higher, the chance of recurrence increases. If one compares 2 hypothetical study populations in which 1 includes only patients who have high-stage primary CRC and the other includes only patients with low-stage primary CRC, the study population with patients who have a high-stage primary CRC will most likely exhibit a relatively greater number of recurrences. Because this outcome can potentially lead to a greater prevalence of recurrent CRC, the positive predictive value for the study with higher-stage primary CRC patients might be greater than the study with lower-stage primary CRC patients. Items such as this that can have a direct effect on the results of FDG PET are essential for a complete understanding of an imaging study's results. When these items are omitted, all potential factors that influenced the study's results cannot be reliably identified.

Another problem found through this meta-analysis of the published literature was that many times the information presented was not clear. For example, the technique used to ultimately confirm an FDG PET imaging finding often cannot be determined through the reported information. In

TABLE 5
Detailed Management Data

Author (reference no.)	n	% Management changes	No. change in location	No. surgery*	Upstaged†	Downstaged‡	Surgery avoided§	Upstaged‖	Downstaged¶
Beets et al. (*29*)	35	40 (14/35)	0	7	4	3	7	7	0
Vitola et al. (*27*)	24	25 (6/24)	0	2	2	0	4	1	3
Lai et al. (*26*)	34	29 (10/34)	0	3	3	0	7	7	0
Delbeke et al. (*24*)	52	33 (17/52)	1	6	4	1	11	9	2
Ogunbiyi et al. (*4*)	23	44 (10/23)	2#	6	4	0	3	3**	1
Valk et al. (*3*)	78	31 (24/78)	0	0	0	0	24	24	0
Flamen (*31*)	103	20 (21/103)	8††	1	0	1	12	9	3

*Number of patients who correctly underwent surgery through FDG PET–directed change in management.
†Number of patients who were upstaged to surgery by FDG PET (e.g., CT normal, FDG PET resectable lesions).
‡Number of patients who were correctly downstaged to surgery by FDG PET (e.g., CT unresectable recurrence, FDG PET resectable lesions).
§Number of patients who correctly avoided surgery through FDG PET–directed change in management.
‖Number of patients who were correctly upstaged to avoid surgery (e.g., CT resectable lesions, FDG PET nonresectable lesions).
¶Number of patients who were downstaged to avoid surgery (e.g., CT resectable lesions, FDG PET no recurrence).
#Change from CT hepatic to true pelvic.
**Six patients upstaged, 2 patients downstaged, no indication of final management decision.
††One patient upstaged to chemotherapy, 2 patients upstaged to extensive disease and died.

most of the articles we reviewed, the final confirmation was either explained very briefly in the text or was mixed in with descriptions of some but not all of the imaging findings. For researchers or physicians who are evaluating the usefulness of FDG PET through these studies, it is critical that the confirmation techniques used for each imaging finding be presented in an unambiguous manner.

For the purpose of determining overall values for sensitivity and specificity by meta-analysis, we used data from all of the relevant articles we evaluated in the literature review. Unlike the change-in-management data, no statistical test was performed to quantitatively justify combining the data. Some tutorial articles for meta-analysis of diagnostic tool studies suggest that descriptions of each study's patient population characteristics should be presented to show that combining the results of the different studies is reasonable (*19–21*). Our justification for combining the studies relied on the fact that all of them focused on patients with CRC who had possible recurrence of the cancer. Furthermore, because of our specific study selection criteria and our extensive literature review, any study that differed from the others in regard to the type of patients studied would have been detected and excluded.

Another important point from these tutorial articles (*19–21*) is that publication bias should always be considered when selecting articles for meta-analysis. Publication bias occurs because there can be a tendency to only publish articles that present favorable results. Manual searches through physical databases as well as the determination of the presence of unpublished articles can prevent publication bias from influencing the results. With the relatively small number of facilities capable of performing FDG PET and the even smaller number of researchers performing FDG PET studies on patients with recurrent CRC, we did not find it feasible or necessary to perform these types of additional searches. Although this does not dismiss the potential for publication bias in our meta-analysis, the probability of discovering other studies that would meet our selection criteria is minimal. It should also be noted that most of these FDG PET studies are conducted by highly trained experts in the nuclear medicine field. If this imaging modality is incorporated into mainstream clinical practice, it is possible that interpreter errors made by physicians with less training might lead to the production of studies in which performance results are less accurate than those in our study. However, our results should not be considered to be inflated values but should be interpreted as a true indication of the ability of FDG PET to detect recurrent CRC in an optimal clinical setting. This situation exemplifies the importance of including information, such as special interpreter characteristics, as stated in our review guidelines.

Despite the issues of the quality by which the imaging data were reported, these studies reflect the potential applicability of FDG PET to a clinical setting for patients with recurrent CRC. In this meta-analysis, an overall sensitivity within a 95% confidence interval of 97% (95% confidence level, 95%–99%) and specificity of 76% (95% confidence level, 63%–88%) were calculated for FDG PET in the whole body (Table 3). In the regions of the liver and pelvis, the specificity was much higher. When considering the specificity of FDG PET in the whole body, the potential for FP findings is greater than isolated organ specificity values. The relatively lower whole-body FDG PET specificity can be attributed to FPs, which resulted from factors such as increased bladder, ureteral, and heart activity; sites of inflammation; and aggregates of benign cells such as granulomas in the lung. Furthermore, because all of these studies focused on patients who we believed may have been

experiencing a recurrence of cancer, the number of patients who were cancer-free was small relative to the number of patients who had cancer. This prevalence issue might have potentially led to an underestimation of the specificity and, in contrast, an overestimation of the sensitivity. Confidence intervals become increasingly important for reporting results such as these.

Using an sROC curve analysis is an ideal way to present overall values by combining sensitivity and specificity data from different studies. When combining these performance measures, it is important to consider the interdependence of sensitivity and specificity. Unfortunately, we were not able to fit our data to an sROC curve, possibly because each group of data only contained, at most, 5 studies. With a limited number of studies, it is difficult to get a distribution of points that will converge into an sROC curve. Furthermore, the sROC curve relies on these studies to operate at different thresholds. It is possible that for the studies included in our meta-analysis, interpretation of FDG PET recurrent CRC images was done at a similar threshold. As more data in this area become available to plot on this sROC curve, a better assessment using this method can be made.

The inability to fit our data to an sROC curve does not undermine the overall values that were determined by pooling the data. With all weighted averages falling within the 95% confidence interval of the pooled data, we think that the method of pooling data from these various FDG PET studies provides a reliable estimate of the combined sensitivity and specificity. Furthermore, pooled data and weighted averages tend to underestimate overall combined values (*22,23*).

There are several other issues that must be addressed when evaluating sensitivity and specificity data. The inherent nature of noninvasive imaging prevents sensitivity and specificity values from describing completely valid information. The problem is that there is no true gold standard for confirming FN and FP results. It is common practice for FN imaging findings to be verified by follow-up confirmation because it would be unacceptable to perform invasive verification techniques. There is always some uncertainty with follow-up confirmation because the diagnostic tools used are not 100% accurate and the amount of time spent monitoring a patient can be relatively short. Another example illustrating the problems with verifying FN imaging findings is when an FDG PET study might have been read as negative, but a subsequent follow-up examination showed cancer cells. To classify the FDG PET finding as FN, the cancer cells must have been present at the time of the original FDG PET scan. However, it is also possible that the cancer cells developed during the time interval between the FDG PET scan and the follow-up examination, which would make the FDG PET finding TN. These errors usually lead to an underestimation of the true sensitivity. Another key factor that influences sensitivity is the lesion size. Lesion size should always be reported so that future studies can determine the decrease in sensitivity as a function of lesion size.

Similarly, although there is not as much uncertainty, FP results can also be unintentionally misinterpreted. As with FN imaging results, it is not always ethical to subject patients to further confirmation of positive imaging findings. If, for example, it has been verified that a patient has untreatable disease because of widespread metastasis, it would not be in the patient's best interest to have all positive FDG PET findings confirmed. This could lead to an overestimation of the specificity because potential FP findings are not being verified.

For studies that focus on the use of FDG PET in a particular region of the body, researchers commonly classify a patient's findings to be TN if the person is disease free in the anatomic region of interest, regardless of the presence of disease in other areas of the body. For example, when a value for the specificity in the liver is to be determined, a patient without liver lesions but with positive extrahepatic findings is considered to be TN. If this is done in a study, the sensitivity and specificity in the other regions should be determined for the same study population. To get an estimate of whole-body FDG PET values, it would be much more accurate to combine, for example, liver-only and local/pelvic-only data from the same study than liver and local/pelvic data from different studies. Regional information is very important for initial assessments of FDG PET. Ultimately, however, the incorporation of FDG PET in the management protocol for patients with recurrent CRC will depend on its accuracy in detecting metastases throughout the entire body.

The dilemma with confirmation is present in all imaging studies. Therefore, this issue should not be a concern when comparing different imaging modality studies. However, what is more relevant is that FDG PET be studied for the purpose of actual clinical application. Providing data on FDG PET–directed change in management for patients with recurrent CRC may be the first step toward assessing FDG PET's true clinical usefulness. Regardless of the problems of confirmation that could limit sensitivity and specificity data, change-in-management studies primarily focus on the medical decision made for a patient before and after FDG PET. In doing so, there is a direct comparison with the diagnostic tools originally used for assessment. In addition, the straightforward measure of FDG PET benefits is simply whether the change-in-management decision made using FDG PET was correct, which can be defined as one in which an FDG PET finding leads to a medical decision that is different from the initial decision (which was based on the findings from other diagnostic tools) and is ultimately determined to be the most appropriate.

By pooling the data from all studies that provided change-in-management information, there was a 29% (95% confidence level, 25%–34%) change in management made by FDG PET (Table 4). This percentage is very informative for the potential incorporation of FDG PET into the management algorithm of patients with recurrent CRC. By combin-

ing the data from the various studies, a more realistic overall representation of the actual recurrent CRC patient population can be made. In individual studies, the indications and prognosis made before FDG PET tend to be restricted to 1 type. For example, 1 study might focus only on patients who have negative conventional imaging scans and elevated carcinoembryonic antigen levels, whereas another might look exclusively at patients whose liver CT scans were positive. Although these studies independently provide valuable information for specific patient groups, a more generalized representation of the overall clinical impact of FDG PET can be seen with the combined results.

Special consideration should be given when interpreting the reported change-in-management percentage. In general, the studies in this meta-analysis reported this percentage on the basis of management decisions that were correctly made using FDG PET. These management studies are questionable because only 2 of them clearly addressed change-in-management decisions made by FDG PET that were found to be inappropriate (*3,31*). Because the accuracy of FDG PET is not perfect, there will inevitably be some incorrect management decisions. In the discussion of the management impact of FDG PET in most of the studies we analyzed, there was no mention of incorrect management decisions. To gain a complete understanding of the impact of FDG PET in the disease management of patients with recurrent CRC from these types of studies, incorrect management decisions should be included in all final evaluations. Also, another possibility that was not considered for this calculation in all but 2 studies (*3,31*) was when FDG PET incorrectly agreed with the initial medical management decision. For example, CT might identify a resectable lesion that is also seen on FDG PET, but the lesion may ultimately prove to be unresectable. Although this type of incorrect management decision is not included in the calculated percentage change in management, this situation represents an incorrect decision that was made on the basis of FDG PET findings.

CONCLUSION

With the results from this literature review and meta-analysis, future cost-effectiveness studies should be possible. The change-in-management results of these studies show the beneficial impact of FDG PET in the management of recurrent CRC in patients. This impact is shown by the ability of FDG PET to effectively direct patients with recurrent CRC to the most appropriate treatment. Furthermore, despite the limitations of the studies that reported sensitivity and specificity data, the values obtained from the meta-analysis can be used as starting estimates for FDG PET in a cost-effectiveness analysis. Through estimates, statistical methodology (sensitivity analysis) can be used to show the impact that varying these values over a wide range would have on the cost-effectiveness of FDG PET. Cost-effectiveness analysis using quantitative decision models is becoming a necessary tool for showing the applicability of new technology in the clinical setting, particularly in situations in which randomized clinical trials are not possible. With the information gathered in this study, clinically relevant management algorithms can be created to determine whether FDG PET can be used as a cost-effective management tool for patients with recurrent CRC.

Overall, we found that although the studies were acceptable for use in our cost-effectiveness analysis, improvements were needed in the presentation of results of FDG PET imaging studies. Ideally, articles should include all of the necessary items described in our formulated guidelines. The meta-analysis provided values for the sensitivity and specificity of FDG PET that were above 90%, except for the specificity of FDG PET in the whole body. In addition, the change-in-management percentage of 29% (95% confidence level, 25%–34%) calculated from this analysis shows the potential impact that FDG PET can have for the management of recurrent CRC in patients. Although the benefits have been indicated, a complete cost-effectiveness study should be performed to determine the value of using FDG PET for managing disease in these postsurgical patients with CRC.

ACKNOWLEDGMENTS

This study was partially supported by funding from Department of Energy contract DE-FC03-87ER60615, the Laubisch Foundation, and the Ahmanson Foundation.

REFERENCES

1. Landis SH, Murray T, Bolden S, Wingo P. Cancer statistics. *CA Cancer J Clin.* 1998;48:6–29.
2. Flanagan FL, Dehdashti F, Ogunbiyi OA, Kodner IJ, Siegel BA. Utility of FDG-PET for investigating unexplained plasma CEA elevation in patients with colorectal cancer. *Ann Surg.* 1998;227:319–323.
3. Valk PE, Abella-Columna E, Hasemann MK, et al. Whole body PET imaging with F-18 fluorodeoxyglucose in management of recurrent colorectal cancer. *Arch Surg.* 1999;134:503–511.
4. Ogunbiyi OA, Flanagan FL, Dehdashti F, et al. Detection of recurrent and metastatic colorectal cancer: comparison of positron emission tomography and computed tomography. *Ann Surg Oncol.* 1997;4:613–620.
5. Turk PS, Wanebo HJ. Resection of pelvic recurrences. In: Wanebo HJ, ed. *Colorectal Cancer.* St. Louis, MO: CV Mosby; 1993:443–463.
6. August DA, Ottow RT, Sugenberker PH. Clinical perspectives on human colorectal cancer metastases. *Cancer Metastasis Rev.* 1984;3:303–324.
7. Steele GJ, Bleday R, Mayer RJ, Lindblad A, Petrelli N, Weaver D. A prospective evaluation of hepatic resection for colorectal carcinoma metastases to the liver: gastrointestinal tumor study group protocol 6584. *J Clin Oncol.* 1991:1105–1112.
8. Saenz NC, Cady B, McDermott WVJ. Experience with colorectal carcinoma metastatic to the liver. *Surg Clin North Am.* 1989;69:359–368.
9. Hughes KS, Simon R, Songhorabody S. Resection of the liver for colorectal carcinoma metastases to the liver: a multi-institutional study of indications for resection. *Surgery.* 1988;103:278–288.
10. Nieweg OE. Potential applications of positron emission tomography in surgical oncology. *Eur J Surg Oncol.* 1994;20:415–424.
11. Strauss LG, Conti PS. The applications of PET in clinical oncology. *J Nucl Med.* 1991;32:623–650.
12. Wagner HNJ, Conti PS. Advances in medical imaging for cancer diagnosis and treatment. *Cancer.* 1991;67:1121–1128.
13. Conti PS, Lilien DL, Hawley K, Keppler J, Grafton ST, Bading JR. PET and 18-F-FDG in oncology: a clinical update. *Nucl Med Biol.* 1996;23:717–735.
14. Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddahi J. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma. *J Nucl Med.* 1996;37:1428–1436.
15. Gambhir SS, Shepherd JE, Shah BD, et al. Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol.* 1998;16:2113–2125.

16. Valk PE, Pounds TR, Tesar RD, Hopkins DM, Haseman MK. Cost-effectiveness of PET imaging in clinical oncology. *Nucl Med Biol*. 1996;23:737–743.
17. Abdel-Nabi H, Doerr RJ, Lamonica DM, et al. Staging of primary colorectal carcinomas with fluorine-18 fluorodeoxyglucose whole-body PET: correlation with histopathologic and CT findings. *Radiology*. 1998;206:755–760.
18. Scott W, Shepherd J, Gambhir S. Cost-effectiveness of FDG-PET for staging non-small cell lung cancer: a decision analysis. *Ann Thorac Surg*. 1998;66:1876–1883.
19. Thacker SB. A quantitative approach to research integration. *JAMA*. 1988;259:1685–1689.
20. Sacks HS, Berrier J, Reitman D. Meta-analysis of randomized controlled trials. *N Engl J Med*. 1987;316:450–455.
21. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med*. 1987;107:224–233.
22. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667–676.
23. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48:119–130.
24. Delbeke D, Vitola JV, Sandler MP, et al. Staging recurrent metastatic colorectal carcinoma with PET. *J Nucl Med*. 1997;38:1196–1201.
25. Keogan MT, Lowe VJ, Baker ME, McDermott VG, Lyerly HK, Coleman RE. Local recurrence of rectal cancer: evaluation of F-18 fluorodeoxyglucose PET imaging. *Abdom Imaging*. 1997;22:332–337.
26. Lai DT, Fulham M, Stephen MS, et al. The role of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucose in identifying operable colorectal cancer metastases to the liver. *Arch Surg*. 1996;131:703–707.
27. Vitola JV, Delbeke D, Sandler MP, et al. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. *Am J Surg*. 1996;171:21–26.
28. Schiepers C, Penninckx F, De Vadder N, et al. Contribution of PET in the diagnosis of recurrent colorectal cancer: comparison with conventional imaging. *Eur J Surg Oncol*. 1995;21:517–522.
29. Beets G, Pennickx F, Schiepers C, et al. Clinical value of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucose in recurrent colorectal cancer. *Br J Surg*. 1994;81:1666–1670.
30. Ruhlmann J, Schomburg A, Bender H, et al. Fluorodeoxyglucose whole-body positron emission tomography in colorectal cancer patients studied in routine daily practice. *Dis Colon Rectum*. 1997;40:1195–1204.
31. Flamen P, Stroobants S, Van Cutsem E, et al. Additional value of whole-body positron emission tomography with fluorine-18–2-fluoro-2-deoxy-D-glucose in recurrent colorectal cancer. *J Clin Oncol*. 1999;17:894–901.
32. Gleeson S. Blue Cross and Blue Shield association initiatives in technology assessment. In: Dawkins HV, Gelijns AC, eds. *Adopting New Medical Technology*. Washington, DC: National Academy Press; 1994:96–100.
33. Lairson PD. Kaiser Permanente's new technology committee: coverage decision making in a group model health maintenance organization. In: Dawkins HV, Gelijns AC, eds. *Adopting New Medical Technology*. Washington DC: National Academy Press; 1994:101–108.
34. Buto KA. Decision making in the health care finance administration. In: Dawkins HV, Gelijns AC, eds. *Adopting New Medical Technology*. Washington DC: National Academy Press; 1994:87–95.
35. Galandiuk S, Wieand H, Moertel C, et al. Patterns of recurrence after curative resection of carcinoma of the colon and rectum. *Surgery*. 1992;174:27–32.
36. Olson RM, Perencevich NP, Malcolm AW, Chaffey JT, Wilson RE. Patterns of recurrence following curative resection of adenocarcinoma of the colon and rectum. *Cancer*. 1980;45:2969–2974.
37. Pilipshen S, Martin H, Stuart H, Quan Q, Sternberg S, Enker W. Patterns of pelvic recurrences following definitive resections of rectal cancer. *Cancer*. 1984;53:1354–1362.
38. Papatheofanis FJ. Strategies for reporting evidence-based clinical studies involving positron imaging. *Clin Positron Imaging*. 1998;1:175–184.
39. Ota T, Ito K, Tadokoro M, et al. Evaluation of the performance of a whole-body positron imaging system with attenuation correction. *Radiat Med*. 1995;13:121–128.
40. Bengel FM, Ziegler SI, Avril N, Weber W, Laubenbacher C, Schwaiger M. Whole-body positron emission tomography in clinical oncology: comparison between attenuation-corrected and uncorrected images. *Eur J Nucl Med*. 1997;24:1091–1098.
41. Dahlbom M, Hoffman EJ, Hoh CK, et al. Whole-body positron emission tomography: part I. Methods and performance characteristics. *J Nucl Med*. 1992;33:1191–1199.
42. Littenberg B, Moses L. Estimating diagnostic accuracy from multiple conflicting reports. *Med Decis Making*. 1993;13:313–321.
43. Simel DL. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol*. 1991;44:763–70.
44. Freund J. *Mathematical Statistics*. 5th ed. Englewood Cliffs, NJ: Prentice Hall; 1992:477–480.