

# Meta-Analysis of the Performance of $^{18}\text{F}$ -FDG PET in Primary Tumor Detection in Unknown Primary Tumors

Roberto C. Delgado-Bolton, MD<sup>1</sup>; Cristina Fernández-Pérez, MD<sup>2</sup>; Alfredo González-Maté, MD<sup>1</sup>; and José L. Carreras, MD, PhD<sup>1</sup>

<sup>1</sup>Department of Nuclear Medicine, Hospital Clínico San Carlos, Madrid, Spain; and <sup>2</sup>Research Unit, Hospital Clínico San Carlos, Madrid, Spain

Detection of the primary tumor has a key role in the management of patients with unknown primary tumors (UPT). The aim of this study was to perform a meta-analysis of the literature to evaluate the accuracy of  $^{18}\text{F}$ -FDG PET in primary tumor detection in patients with UPT. **Methods:** Systematic methods were used to identify, select, and evaluate the methodologic quality of the studies as well as to summarize the overall findings of sensitivity, specificity, and detection capacity of the primary tumor. The search strategy consisted of identifying studies published between January 1994 and May 2001 indexed in MEDLINE and CANCERLIT. Studies identified by manually searching reference lists of retrieved studies or by reviewing abstracts from recent conference proceedings were also included. Inclusion criteria were studies that evaluated primary tumor detection with  $^{18}\text{F}$ -FDG PET in patients with UPT. Exclusion criteria were duplicated studies or those outdated by subsequent ones. The statistical analysis included 95% confidence intervals (CI) of sensitivity and specificity, both in the pooled data and in the types of studies found. Variation in accuracy between studies was analyzed calculating the natural logarithm of the odds ratio (ln OR) due to study characteristics. Funnel plots of sensitivity and specificity and the summary receiver-operating-characteristic (ROC) curve were also represented. **Results:** Fifteen studies met the inclusion criteria and were analyzed. Although sample sizes were small, compliance with the methodologic quality criteria was adequate. Heterogeneity analysis showed that differences in the study quality did not correlate with differences in study results. The 95% CI of sensitivity and specificity presented global homogeneity, estimating the sensitivity at 0.87 (95% CI, 0.81–0.92) and the specificity at 0.71 (95% CI, 0.64–0.78). The summary ROC curve showed a good relationship between sensitivity and specificity. The ln OR presented significant values in >75% of the studies. **Conclusion:**  $^{18}\text{F}$ -FDG PET could be useful in patients with UPT for the detection of the primary tumor.  $^{18}\text{F}$ -FDG PET has intermediate specificity and high sensitivity, indicating the existence of few false-negative results, an important feature in the management of oncologic patients that could suggest its utility in the initial stages of the management process.

**Key Words:** neoplasms; unknown primary;  $^{18}\text{F}$ -FDG PET; meta-analysis; carcinoma of unknown primary syndrome

**J Nucl Med 2003; 44:1301–1314**

**T**he incidence of unknown primary tumors (UPT) in oncologic patients is 0.5%–7% at the time of the initial diagnosis (1,2), and its prevalence is between 3% and 15% (3). The mean survival from the time of the initial diagnosis is <6 mo and the survival at 3 and 5 y is 11% and 6%, respectively (2). UPT presents metastatic dissemination patterns that are different from those observed in oncologic conditions with known primary tumors (2,4): (a) Short symptomatic prediagnostic intervals and clinically fast tumor growth have often been observed in UPT; (b) UPT becomes symptomatic at the time of metastatic dissemination, whereas their primary sites remain symptomatically silent; and (c) the most frequent primary sites in patients with UPT do not include several of the common primary tumors in the general population, such as breast and prostate cancer. Moreover, no specific metastatic location or combination of locations has been consistently associated with a specific primary tumor site (4). These aspects make it difficult to locate the primary tumor, which is one of the most important factors for establishing the most effective treatment.

The primary tumor is detected in <40% of the patients by conventional diagnostic procedures, frequently after having performed many examinations in all patients. These examinations are often nonconclusive and generate discomfort for the patient as well as a high economic cost (2,3).

In 1994, studies that evaluated the usefulness of PET with  $^{18}\text{F}$ -FDG in detecting the primary tumor in patients with UPT began to be published (5). The reason was the technique's capacity to detect different tumor types in all of the body noninvasively and in a single examination (6). Most studies ensure that  $^{18}\text{F}$ -FDG PET detects the primary tumor in around 40% of patients with negative results in the conventional diagnostic procedures. When  $^{18}\text{F}$ -FDG PET

Received Jun. 11, 2002; revision accepted Mar. 21, 2003.

For correspondence or reprints contact: Roberto C. Delgado-Bolton, MD, Servicio de Medicina Nuclear, Hospital Clínico San Carlos, c/o Prof. Martín Lagos s/n, Madrid, 28040 Spain.

E-mail: delgadobolton@eresmas.com

does not locate the primary tumor, it is not detected during follow-up in most cases either, due to high sensitivity and specificity of  $^{18}\text{F}$ -FDG PET for primary tumor detection (6–8).

However, the reduced number of patients in most of the studies (9–11), the inclusion of patients from specific populations in some cases (12–14), and the existence of conflicting or nonconclusive results (6,15,16) make it impossible to draw conclusions on the utility of  $^{18}\text{F}$ -FDG PET in this disease. Thus, a meta-analysis needs to be performed to increase statistical power and to estimate the overall accuracy, to resolve uncertainties when studies disagree, and to examine the variation in the accuracy due to different study characteristics (17–21). The aim of this study was to perform a meta-analysis to evaluate the accuracy of  $^{18}\text{F}$ -FDG PET in UPT in the detection of the primary tumor. Assessment of methodologic quality was performed to determine the influence of study quality on reported results. The homogeneity between the different types of studies detected was analyzed to integrate the results and increase the statistical power and estimation of accuracy.

## MATERIALS AND METHODS

### Study Identification

Two investigators performed a systematic search of the literature to identify relevant studies published between January 1994 and May 2001 in the MEDLINE and CANCERLIT databases (22). The search strategies developed in MEDLINE are presented in Table 1. One investigator also manually reviewed the reference lists of retrieved articles and abstracts from recent conference proceedings. We included studies published in any language (22,23) as well as published abstracts presented at congresses (24).

### Study Selection

Two investigators independently evaluated the titles, abstracts, and complete articles (if available) of >300 studies identified in the search for inclusion. Disagreements were resolved by discus-

sion with the participation of a third investigator. Reviewers were not blinded to the journal, author, institution, or date of publication. We included studies published in any language that (a) evaluated  $^{18}\text{F}$ -FDG PET in patients with UPT for primary tumor detection, (b) included at least 4 patients with UPT, and (c) reported primary data sufficient to allow calculation of both sensitivity and specificity for primary tumor detection. Those studies that fulfilled these inclusion criteria were examined to exclude duplicated studies or those that were out-dated by other more recent ones when the patients presented overlapped. Abstracts from MEDLINE or from congresses were only included when the aims, methods, and results of sensitivity and specificity of the study were clear. These studies were included in the sensitivity analysis but not in the analysis of the methodologic quality. In studies that included patients with different diseases, only those patients with UPT were included. Fifteen eligible studies were selected from >300 potentially relevant studies (Table 2).

### Study Quality

The selection criteria were designed to identify studies that fulfilled the minimum requirements. To assign a quality score to each study we performed a methodologic quality assessment. Second, the grade of the evidence and the contribution to the patients' management were evaluated. Reviewers were not blinded to the study title, results, authors, institution, or journal in which the study results were published.

To assess the methodologic quality we modified previously developed criteria by Huebner et al. (25) and Gould et al. (26). Quality assessment was only possible in the studies in which the complete published article could be accessed. One investigator evaluated the fulfillment of the methodologic quality criteria, assessing the exhaustiveness of the description and the adherence to the guidelines formulated. The criteria evaluated covered 7 dimensions: description of study design, description of the study population, indications leading to  $^{18}\text{F}$ -FDG PET use, technical and image interpretation issues, final confirmation, sensitivity and specificity data, and change in management information. Table 3 presents the items analyzed in each guideline. Adherence to each item in the 7 guidelines was considered adequate (A), partial (P),

**TABLE 1**  
Detailed Literature Search

Search strategy no.	Literature search strategies (combination of terms)	Results (no. of references)	
		Identified*	Complied with inclusion criteria†
1	Unknown primary tumor and PET	39	13
2	Unknown primary tumor and FDG	24	13
3	Occult primary tumor and PET	39	9
4	Occult primary tumor and FDG	22	9
5	Unknown primary and PET	80	16
6	Unknown primary and FDG	43	16
7	FDG and PET and metastatic tumors	5	1
8	Neoplasm and unknown primary and FDG and PET	9	3
9	Carcinoma of unknown primary and FDG and PET	1	1
10	Primary tumor and FDG PET and detection	28	9
Total no. of references		290	90

\*No. of references identified that were published between January 1994 and May 2001.

†No. of references that complied with inclusion criteria. Following stage was to apply exclusion criteria.

**TABLE 2**  
Studies Included in Meta-Analysis, Participant Characteristics, Types of Studies, and Results

Study	Year	Information analyzed	UPT patient characteristics				Study type				Results: primary tumor detection <sup>¶</sup>
			<i>n</i> <sup>*</sup> [NC]	M/F <sup>†</sup>	Mean age ± SD (y) [range] <sup>‡</sup>	UPT characteristics	Type I/II	Timing of <sup>18</sup> F-FDG PET <sup>§</sup>	Type A/B	Region studied <sup>  </sup>	
Rege et al. (5)	1994	Published article	4	3/1	51.2 ± 2.06 [49–53]	Extracranial head and neck cancers from UPT	Type II	MRI	Type B	PET: base of cerebellum to larynx	PET: 2/4 MRI: 0/4
Schipper et al. (11)	1996	MEDLINE abstract	16	NS	NS	Cervical lymph nodes from UPT	Type I	Negative CI	Type A	PET: WB	PET: 4/16
Kole et al. (16)	1998	Published article	29	20/9	56.4 ± 11.59 [37–77]	UPT	Type I	Negative CI	Type A	PET: WB	PET: 7/29
Shinohara et al. (10)	1999	MEDLINE abstract	4	NS	NS	UPT	Type I	Negative CI	Type A	PET: WB	PET: 2/4
Hanasono et al. (13)	1999	Published article	20	NS	NS	SCC metastatic to cervical lymph nodes from UPT	Type II	PET+CT+MRI: 2 Pt PET+CT: 7 Pt PET+MRI: 9 Pt Only PET: 2 Pt PET+CT: 8 Pt PET+MRI: 5 Pt DPB: 13 Pt	Type A	PET: WB	PET: 7/20 CT: 2/9 MRI: 4/11
Greven et al. (15)	1999	Published article	13	NS	NS	SCC metastatic to cervical lymph nodes from UPT	Type II		Type B	PET: midcranium down to midchest	PET: 1/13 CT: 0/13 MRI: 0/13 DPB: 0/13 PET: 9/20
Lassen et al. (8)	1999	Published article	20	9/11	52.7 ± 14.25 [18–75]	UPT	Type I	Negative CI	Type A	PET: WB	PET: 3/14
Safa et al. (14)	1999	MEDLINE abstract	14	NS	NS	SCC metastatic to cervical lymph nodes from UPT	Type I	Negative CI	Type A	PET: WB	PET: 3/14
Gupta et al. (12)	1999	Published article	31	NS	67 ± 14.00 [45–81]	Intracranial <sup>#</sup> metastatic disease	Type II	PET+CT: 22 Pt (double-blinded) CT or MRI	Type A	PET: WB	PET: 20/31 CT: 6/31
Aassar et al. (35)	1999	Published article	17	13/2**	58.6 ± 15.40 [39–87]	Cervical lymph nodes from UPT	Type II		Type B	PET: skull base to thoracic inlet	PET: 9/17 CT or MRI: 5/17
Lonneux and Reffad (7)	2000	Published article	24	15/9	59 ± 10.00 [35–79]	UPT	Type I	Negative CI	Type A	PET: WB	PET: 13/24
Trampal et al. (9)	2000	Congress abstract	9 [3]	NS	NS	UPT	Type I	Negative CI	Type A	PET: WB	PET: 4/9
Bohuslavizki et al. (6)	2000	Published article	52 [1]	33/19	59.5 ± 10.04 [38–82]	UPT	Type I	Negative CI	Type A	PET: WB	PET: 20/52
Jungehülsing et al. (36)	2000	Published article	27	22/5	60.1 ± 10.53 [36–74]	UPT	Type I	Negative CI	Type B	PET: head to diaphragm	PET: 7/27
Cheng et al. (34)	2000	Congress abstract	18	16/2	61 [31–88]	Cervical lymph nodes from UPT	Type II	CT or MRI	Type B	PET: head to chest	PET: 9/18 CT or MRI: 2/18

\**n* = number of patients included in meta-analysis; NC = no. of patients with <sup>18</sup>F-FDG PET results that were not confirmed and who were excluded from meta-analysis.

<sup>†</sup>Sex of patients included in meta-analysis.

<sup>‡</sup>Mean age ± SD and range of age (in years) of patients included in meta-analysis.

<sup>§</sup>Moment of management process in which <sup>18</sup>F-FDG PET study was performed. Type I = <sup>18</sup>F-FDG PET was performed when all diagnostic procedures carried out did not detect primary tumor; Type II = <sup>18</sup>F-FDG PET was compared with CT or MRI in double-blinded study that included UPT patients with negative results in other tests.

<sup>||</sup>Region included in <sup>18</sup>F-FDG PET study. Type A = whole-body <sup>18</sup>F-FDG PET study was carried out; Type B = <sup>18</sup>F-FDG PET study only included head-neck-thorax region (as shown in this column).

<sup>#</sup>Proportion of patients analyzed in study in whom primary tumor is correctly identified by <sup>18</sup>F-FDG PET, CT, MRI, or other.

<sup>¶</sup>Thirty-one patients with intracranial metastatic disease suspected with CT or MRI and confirmed in 22 of them (12).

\*\*Two patients in whom <sup>18</sup>F-FDG PET correctly detected lung cancer were included in meta-analysis, although they had not been included in original study as it focused on detection with <sup>18</sup>F-FDG PET of head and neck primary tumors from UPT (35). Age and sex of both patients were not specified in article.

NS = not specified; CI = conventional imaging; WB = whole-body <sup>18</sup>F-FDG PET; SCC = squamous cell carcinoma; Pt = patients; DPB = direct panendoscopy with biopsy.

**TABLE 3**  
Methodologic Quality Criteria Guidelines\*

Guideline	Criteria for adherence evaluation
Guideline 1: description of study design and patient selection criteria	
1. Study design	Type of study conducted must be clearly presented.
2. Patient selection criteria	Inclusion criteria, especially particular specifications, should be explained.
3. Exclusion of patients from study's final analysis	Exclusion of patients from study's final analysis should be explained, as this can affect study results.
Guideline 2: characteristics of patient population finally studied	
1. Mean age with range and sex	Ideally, age and sex of all patients included must be described, as this will allow comparisons of patient populations of different studies.
2. Comorbid conditions	Presence or absence should be stated, as this can affect interpretation.
3. Diabetes mellitus	Patient description should include presence or absence of diabetes mellitus. <sup>†</sup>
4. Localization of metastasis of UPT	Localization of metastasis of UPT should be clearly stated for every patient included in study, together with histopathologic findings.
5. Special institution characteristics	Institution characteristics should be described or stated in published article, as it can be useful for defining study population.
Guideline 3: patient indications leading to use of <sup>18</sup> F-FDG PET	
1. Reasons for use of <sup>18</sup> F-FDG PET	Results of previous diagnostic procedures and clinical situations used to refer patient for further evaluation with <sup>18</sup> F-FDG PET must be clearly explained.
2. Reasons for use of <sup>18</sup> F-FDG PET correlated with specific <sup>18</sup> F-FDG PET findings	Reasons for patient referral to <sup>18</sup> F-FDG PET and results of <sup>18</sup> F-FDG PET scan should be stated for every patient included in study analysis.
3. Extent of metastatic disease in every patient	If patients present several metastatic lesions from UPT, localization and extent of every lesion should be clearly stated.
Guideline 4: details of technologies used during study and image-interpretation issues	
1. Imaging techniques used in study and resolution	All imaging techniques used, together with their resolutions, should be stated.
2. Patient preparation	Patient preparation should follow Society Nuclear Medicine guidelines (27).
3. Minimum 4-h fasting before <sup>18</sup> F-FDG injection	Minimum 4-h fasting before <sup>18</sup> F-FDG administration should be carried out in every patient and clearly stated.
4. Blood glucose level checked before <sup>18</sup> F-FDG injection	Blood glucose level should be checked before <sup>18</sup> F-FDG administration in every patient and clearly stated.
5. Attenuation correction (AC)	If AC is performed, it is scored A. If AC is not performed, but this is correctly stated, it is scored P. N score is given when this issue is not addressed.
6. Explanation of special characteristics of interpreters	No. of interpreters and their experience in reading <sup>18</sup> F-FDG PET scans should be clearly stated.
7. Definition of positive and negative <sup>18</sup> F-FDG PET findings	Characteristics that defined imaging findings as positive or negative should be clearly explained, as these aspects represent study's threshold.
8. Additional scans	Additional imaging procedures performed in patients should be stated.
Guideline 5: final diagnostic confirmation	
1. Final confirmation	Diagnostic procedures used as reference standards must be described.
2. Association between specific <sup>18</sup> F-FDG PET finding and final confirmation technique used	If different confirmation methods are used, correspondence between each specific imaging finding and specific confirmation method used should be clearly described.
3. Histopathologic confirmation	Although histopathologic confirmation is considered the gold standard, technique used (surgery, biopsy, fine-needle aspiration cytology) must be stated, as diagnostic accuracy can vary between techniques.
4. Nonhistopathologic confirmation	When histopathologic confirmation is not possible, other techniques used for confirmation should be explained and amount of time of follow-up stated.
5. Patients lost to clinical follow-up confirmation	Patients lost to clinical follow-up confirmation should be described together with clear statement of their inclusion (or not) in study's final analysis.
Guideline 6: sensitivity and specificity data	
1. Description of TP, FP, TN, and FN	True-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) values should be quantitatively described for all imaging methods analyzed in study and clearly presented.
2. Qualitative explanation of FP and FN	Errors made by imaging modality (FP, FN) need to be explained.
3. Specific region studied by <sup>18</sup> F-FDG PET	When study performs whole-body <sup>18</sup> F-FDG PET scans and this is clearly stated, score is A. When <sup>18</sup> F-FDG PET scan includes only 1 region of body and it is clearly stated, score is P. When region studied by <sup>18</sup> F-FDG PET is not clearly stated, score is N.
4. Location and no. of primary tumors and other metastatic lesions detected by <sup>18</sup> F-FDG PET	All lesions detected by <sup>18</sup> F-FDG PET must be described in every patient. Localization of primary tumors and metastatic lesions detected by PET must be clearly presented.
5. Confidence intervals	Confidence intervals of values reported should be included.
6. Equivocal <sup>18</sup> F-FDG PET findings	When equivocal <sup>18</sup> F-FDG PET findings exist, it should be clear how findings were defined in final analysis of study.
7. Data reported in patients and in corresponding lesions	Studies should report data both for patients and for lesions detected.

**TABLE 3** (Continued)

Guideline	Criteria for adherence evaluation
Guideline 7: change-in-management information	
1. <sup>18</sup> F-FDG PET-directed change in management	Management changes that are consequence of <sup>18</sup> F-FDG PET should be clearly stated. When study does not analyze change-in-management information, an N/A score is assigned (to most or all items in guideline 7).
2. Diagnostic tool used to make initial treatment decision	Diagnostic tool used and its results that were used to make initial treatment decision must be clearly described, allowing comparisons with results of <sup>18</sup> F-FDG PET scan.
3. Diagnostic tool used to make final treatment decision	Diagnostic tool used and its results that were used to make final treatment decision must be clearly described.
4. Medical treatment made both initially and after <sup>18</sup> F-FDG PET	Changes in medical treatment must be stated, allowing evaluation of management changes induced by <sup>18</sup> F-FDG PET.
5. Correct or incorrect <sup>18</sup> F-FDG PET-directed management change	Clear statement of correct or incorrect <sup>18</sup> F-FDG PET-directed management change must be presented.
6. Upstaging and downstaging	Extent of disease diagnosed before and after <sup>18</sup> F-FDG PET scan must be presented.
7. <sup>18</sup> F-FDG PET in management algorithm	Management data presented in study must be discussed and conclusion on benefits of <sup>18</sup> F-FDG PET and moment of management algorithm in which to perform <sup>18</sup> F-FDG PET scan must be clearly presented.

\*Methodologic quality criteria guidelines formulated by Huebner et al. (25) and adapted for review of studies of patients with UPT in which <sup>18</sup>F-FDG PET is performed to detect primary tumor.

<sup>†</sup>Patients with diabetes mellitus (DM) can present high levels of blood glucose, especially when DM is poorly controlled with treatment. During hyperglycemic states, uptake of <sup>18</sup>F-FDG by tumor can be reduced due to competitive effect between excess plasma glucose and <sup>18</sup>F-FDG. As a consequence, <sup>18</sup>F-FDG PET image quality is reduced and sensitivity decreases. Studies should describe in all patients: presence or absence of DM and blood glucose level during <sup>18</sup>F-FDG administration.

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

not addressed (N), or not applicable (N/A) depending on compliance with the established guidelines and the amount of information presented. An A score was assigned when an item was described exhaustively and complied with the methodologic guidelines. A P score signified that an item was not described sufficiently or that it only partially complied with the methodologic guidelines. When an item in our guidelines was not described at all in the article or did not comply with the guidelines, an N score was awarded. An N/A score was assigned to all items included in those guidelines that were not applicable to the study's analysis. The guidelines and scoring system described are designed to assess the methodologic rigor and scientific quality of the articles included. The guidelines published by the Society of Nuclear Medicine (27) were used as a reference for the assessment of the technical quality of <sup>18</sup>F-FDG PET. Fulfillment of the methodologic quality criteria for each article was considered high, acceptable, or low, when the percentage of A scores of adherence for each article was >70%, 50%–70%, or <50%, respectively. Quality levels for each article were statistically analyzed to assess the existence of correlation with differences in study results.

Next, 2 investigators independently assessed the validity, grade of the evidence, generalizability, and contribution to the patients' management in all studies included in the meta-analysis (28,29). Methodologic aspects such as the selection of the study population and the application of the reference tests were analyzed (30), and their combined assessment made it possible to assign a grade of evidence and generalizability (28). Four grades of evidence have been defined: A, B, C, and D, with grade A and B being defined as high-quality evidence and a wide or moderate spectrum of generalizability; grade C as weak evidence in studies with several methodologic defects, small sample sizes, or incomplete description; and grade D as nonconclusive studies with multiple methodologic defects (28). Finally, the contribution of <sup>18</sup>F-FDG PET to

the patient management process according to the model described by Fryback and Thornbury was assessed (29). This model consists in a hierarchic model of efficacy with 6 levels of efficacy: technical efficacy (level 1), diagnostic accuracy efficacy (level 2), diagnostic thinking efficacy (level 3), therapeutic efficacy (level 4), patient outcome efficacy (level 5), and societal efficacy (level 6). Reaching a higher level in the hierarchy means that its efficacy is demonstrated at lower levels, but the reverse is not true. Disagreements between the 2 investigators were resolved by discussion.

### Data Abstraction

One investigator abstracted the following data from each eligible study: number, demographic characteristics, and inclusion criteria of the patients; study design; and UPT characteristics. Those patients in whom the <sup>18</sup>F-FDG PET result was not confirmed were not included in the meta-analysis. On the basis of their design, 2 types of studies were differentiated; in some, <sup>18</sup>F-FDG PET was performed when all diagnostic procedures performed did not detect the primary tumor (type I); whereas, in others (type II), <sup>18</sup>F-FDG PET was compared with CT or MRI in a double-blind study that included UPT patients who had presented negative results for primary tumor detection in all of the following diagnostic procedures (if performed in each particular patient): (a) careful clinical history and complete physical examination; (b) laboratory analysis; (c) endoscopic evaluations; (d) radiologic or isotopic procedures except <sup>18</sup>F-FDG PET, CT, and MRI; or even (e) surgical exploration, biopsy, or fine-needle aspiration cytology of suspicious lesions. On the basis of the region studied by <sup>18</sup>F-FDG PET, whole-body studies (type A) were differentiated from head-neck-thorax studies (type B).

To calculate sensitivity and specificity, true-positive (TP) was considered when <sup>18</sup>F-FDG PET suggested the location of the primary tumor and was subsequently confirmed, whereas false-

positive (FP) was considered when this location was not confirmed. The sites suggested by  $^{18}\text{F}$ -FDG PET were confirmed by histopathologic analysis of tissue obtained by biopsy or surgery, considered as the gold standard; however, imaging procedures or clinical follow-up were accepted if no histopathologic proof could be obtained. Even if other lesions were detected, when  $^{18}\text{F}$ -FDG PET did not suggest the location of the primary tumor, it was considered to be true-negative (TN) if the primary tumor remained unknown in the follow-up. It was considered false-negative (FN) if the primary tumor was identified subsequently to negative  $^{18}\text{F}$ -FDG PET.

### Statistical Analysis

To evaluate agreement between investigators for the assessment of the grade of the evidence and change-in-management information, the weighted  $\kappa$  index, considering a discordant ordinal weight, was calculated.

To perform a subgroup analysis we estimated the 95% confidence intervals (CI) of sensitivity and specificity in type I, type II, type A, and type B studies. Homogeneity in the 95% CI of sensitivity and specificity was analyzed to assess the possibility of obtaining global values (31).

The natural logarithm of the odds ratio (ln OR) was calculated for each study. The ln OR is a measurement of the performance of a diagnostic test, based on the positive correlation between the presence of disease and a positive result of the diagnostic test.

To characterize the performance of a diagnostic test based on the results of multiple studies, we developed summary receiver-operating-characteristic (ROC) curves as described by Moses et al. (32). The estimates of sensitivity and specificity from the included studies were combined to construct the summary ROC curve, which illustrates the trade-off between sensitivity and specificity as the threshold for defining a positive test is changed. The summary ROC curve is characterized by the point of maximum joint sensitivity and specificity. This point is defined by the intersection of the summary ROC curve with a diagonal line that runs from the top left corner to the bottom right corner of the diagram, along which sensitivity and specificity are equal. This point is the maximum attainable common value for sensitivity and specificity of a test and is a global measure of test accuracy. The maximum joint sensitivity and specificity of a perfect test is 1.0, and the maximum joint sensitivity and specificity of a test that has no diagnostic value is 0.5 (26,32).

To assess the variation of sensitivity and specificity due to study characteristics we used a multiple linear regression model (33). This model allows the combination of data from independent diagnostic test studies and provides a method of assessing the association between test accuracy and study characteristics. Variation in accuracy was measured by the ln OR depending on study characteristics, whether type I or type II, type A or type B, using the formula  $\ln \text{OR} = \alpha + \beta_1 (\text{type I or type II}) + \beta (\text{type A or type B})$ . The adjusted  $\beta$  values and their 95% CI were calculated.

The positive and negative likelihood ratios (LR) were calculated to assess changes in pretest probability induced by the diagnostic test result.

### Sensitivity Analysis

To determine if study quality affected diagnostic accuracy we compared high-quality studies (overall A adherence score > 70%) with acceptable or low-quality studies (overall A adherence score < 70%). Study quality and study results were correlated to assess the influence of study quality on the results (21,26). Study

quality rating was used to weight the individual study results when pooling the results (21).

To assess the presence of publication bias we created funnel plots of sensitivity and specificity plotted against sample size. A symmetric plot would provide reassuring evidence that no study had been left out, whereas an asymmetric plot would suggest the presence of publication bias (26,31).

## RESULTS

### Study Identification and Eligibility

Our search identified 290 potentially relevant studies in MEDLINE (Table 1). There were 90 eligible studies that met all inclusion criteria. We excluded 76 of the 90 studies because they were duplicated or were outdated by other more recent ones when the patients presented overlapped. Those studies in which only the abstract was available were included in the sensitivity analysis but not in the analysis of the methodologic quality. Search strategies similar to those described above were applied to CANCERLIT, identifying only 1 study that had not been detected in the MEDLINE search. This study was excluded as it was outdated by Lassen et al. (8). Our manual search of the reference lists of retrieved articles and the review of abstracts from recent conference proceedings identified several studies that had not been previously detected. Only 2 congress abstracts met all inclusion criteria. One of them was outdated by Kole et al. (16), whereas the other was included (34). Therefore, 15 studies were finally included in the meta-analysis.

### Study Description

The 15 studies selected for the meta-analysis are summarized in Table 2 (5–16,34–36). These studies analyzed 302 patients with UPT, 298 of whom were included in the meta-analysis, whereas 4 patients without confirmation of  $^{18}\text{F}$ -FDG PET were excluded. The complete article was available in 10 of the studies (5–8,12,13,15,16,35,36), which included 237 patients (79.53%). Only the abstract was available in the remaining 5 studies (9–11,14,34), which included the remaining 61 patients (20.47%). The mean age of the patients ranged from 51.25 to 67 y, and the proportion of males to females ranged from 45% to 89%. The average proportion of males per study was 71%, and the male-to-female ratio was 2.2. The localization of the metastatic lesions from UPT is described in Table 4. The histopathology of the metastatic lesions from UPT is shown in Figure 1. The localization of the primary tumors detected by  $^{18}\text{F}$ -FDG PET or by clinical follow-up is shown in Figure 2.

The 15 studies analyzed included UPT patients. Although 7 studies included patients that fulfill this definition (6–10,16,36), 3 studies only included those with cervical adenopathies (11,34,35), 3 studies only included those with cervical adenopathies of squamous cell carcinoma (13–15), 1 study included only those with extracranial metastases from UPT (5), and 1 study included patients with intracranial metastatic disease suspected with CT or MRI (12). With regard to the types of studies described in the methods, 9

**TABLE 4**  
Localization of Metastasis from UPT

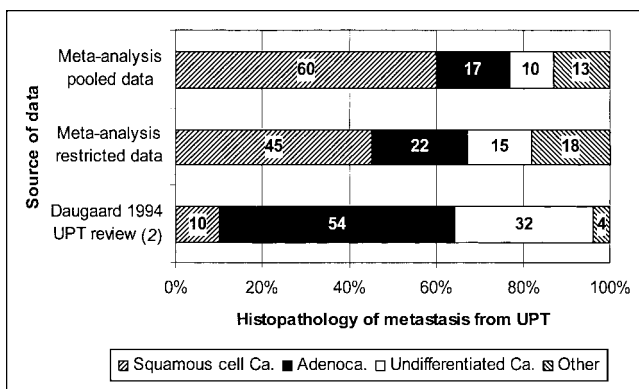
Localization	Patients	
	No.	%
Cervical and supraclavicular lymph nodes	199	66.78
Axillary lymph nodes	9	3.02
Other lymph nodes	6	2.01
Bone	11	3.69
Brain	42	14.09
Lung	6	2.01
Hepatic	4	1.34
Skin	5	1.68
Other localizations	11	3.69
Several metastases	5	1.68
Total number of patients described*	298	

\*Localization of metastasis of UPT was described in all patients included in meta-analysis.

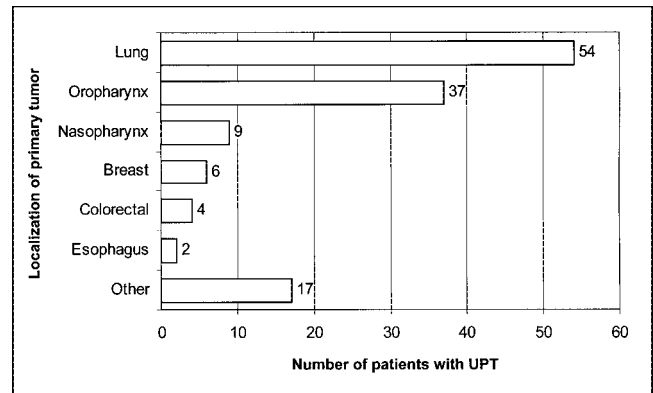
studies (195/298 [65.44%] patients) were type I (6–11,14,16,36), 6 studies (103/298 [34.56%] patients) were type II (5,12,13,15,34,35), 10 studies (219/298 [73.49%] patients) were type A (6–14,16), and 5 studies (79/298 [26.51%] patients) were type B (5,15,34–36). Attenuation correction was performed in 9 studies (5,7,12–15,34–36), in 3 studies attenuation correction was not performed (6,8,16), and in the remaining 3 studies the abstract did not describe this detail (9–11).  $^{18}\text{F}$ -FDG PET image interpretation was performed by using qualitative methods in 14 studies (5–16,35,36), and only 1 study used qualitative and semiquantitative methods (34).

### Study Quality

Table 5 presents the scores assigned to each quality item for the 9 complete articles (5,6,8,12,13,15,16,35,36) that



**FIGURE 1.** Histopathology of metastasis from UPT as described by Daugaard (2) compared with meta-analysis pooled-data and restricted data. Meta-analysis restricted data excluded studies that only included patients with cervical lymphadenopathies (11,34,35) and patients with cervical lymphadenopathies of squamous cell carcinoma (13–15). Squamous cell Ca. = squamous cell carcinoma; Adenoca. = adenocarcinoma; Undifferentiated Ca. = undifferentiated carcinoma.



**FIGURE 2.** Localization of primary tumors detected by  $^{18}\text{F}$ -FDG PET or by clinical follow-up in patients with UPT included in meta-analysis.

were included in the methodologic quality analysis. Percentages of A, P, and N scores for each article were calculated. The range of A adherence scores for each article was 51.61% (16/31) (13) to 81.58% (31/38) (36), with a mean of 68.44% and a SD of 10.73. The mean percentage for P items for each article was 13.86%, with a range of 5.26% (2/38) (36) to 28.95% (11/38) (12); SD was 7.10. The N adherence scores for each article had a mean of 17.7% and a SD of 8.62, with a range from 7.89% (3/38) (8) to 32.26% (10/31) (5,13). Six of the 9 articles analyzed (66.67%) received a percentage of A adherence score of >70% (6,8,15,16,35,36), considered as high quality. The 3 remaining articles (5,12,13) presented a percentage of A scores between 50% and 70%, considered as acceptable quality. Five of the 9 articles analyzed (55.55%) had more N items than P items (5,13,15,16,36). Heterogeneity analysis showed that differences in study quality did not correlate with differences in study results.

Apart from the summary of adherence scores for each article across all items, we determined the combined scores of all articles for each guideline (Table 5). Guidelines 1, 3, 5, and 7 received an overall A adherence score of >70%, indicating adequate fulfillment of these guidelines. The remaining guidelines presented between 50% and 70% of A items. Guidelines 2, 4, and 5 scored more N items than P items.

Table 6 presents the consensus scores of the 2 investigators in the assessment of study validity, grade of the evidence, and contribution to the patients' management for all studies included in the meta-analysis. Nonweighted interrater agreement between the 2 investigators for the assessment of these aspects was 0.80 (95% CI, 0.66–0.94), indicating good agreement. No study included a control group free of the disease analyzed among its patients. In the 6 type II studies,  $^{18}\text{F}$ -FDG PET is compared with other techniques (5,12,13,15,34,35), but there is no control group. All studies used appropriate and objective reference tests: histopathologic confirmation and clinical follow-up.  $^{18}\text{F}$ -FDG PET was performed and interpreted blinded to both reference tests in all studies. However, histopathologic confirmation is

**TABLE 5**  
Study Quality Assessment with Methodologic Quality Criteria Guidelines\*

Guideline	Study										Guideline (%)
	Rege et al. (5)	Kole et al. (16)	Hanasono et al. (13)	Greven et al. (15)	Lassen et al. (8)	Gupta et al. (12)	Aassar et al. (35)	Bohuslavizki et al. (6)	Jungehülsing et al. (36)		
Guideline 1: description of study design and patient selection criteria											
1. Study design	A	A	A	A	A	P	A	A	A	} A = 70.37 P = 25.93 N = 3.85	
2. Patient selection criteria	P	A	P	P	A	P	P	A	A		
3. Exclusion of patients from study's final analysis	A	A	A	A	A	N	P	A	A		
Guideline 2: characteristics of patient population finally studied											
1. Mean age with range and sex	A	A	N	N	A	P	A	A	A	} A = 55.55 P = 4.44 N = 40.00	
2. Comorbid conditions	N	N	N	N	N	N	N	N	N		
3. Diabetes mellitus	N	N	N	N	N	N	N	P	N		
4. Localization of metastasis of UPT	A	A	A	A	A	A	A	A	A		
5. Special institution characteristics	A	A	A	A	A	A	A	A	A		
Guideline 3: patient indications leading to use of <sup>18</sup> F-FDG PET											
1. Reasons for use of <sup>18</sup> F-FDG PET	A	A	A	A	A	A	A	A	A	} A = 100.00	
2. Reasons for use of <sup>18</sup> F-FDG PET correlated with specific <sup>18</sup> F-FDG PET findings	A	A	A	A	A	A	A	A	A		
3. Extent of metastatic disease in every patient	A	A	A	A	A	A	A	A	A		
Guideline 4: details of technologies used during study and image-interpretation issues											
1. Imaging techniques used in study and resolution	A	A	P	A	P	A	A	A	A	} A = 62.50 P = 16.67 N = 20.83	
2. Patient preparation	N	A	N	A	A	A	A	A	A		
3. Minimum 4-h fasting before <sup>18</sup> F-FDG injection	N	A	N	A	A	A	A	A	A		
4. Blood glucose level checked before <sup>18</sup> F-FDG injection	N	N	N	N	P	N	N	N	A		
5. Attenuation correction	A	P	A	A	P	A	A	P	A		
6. Explanation of special characteristics of interpreters	A	N	A	A	A	P	A	N	N		
7. Definition of positive and negative <sup>18</sup> F-FDG PET findings	A	A	A	A	A	P	P	P	N		
8. Additional scans	A	A	A	A	P	A	A	P	A		
Guideline 5: final diagnostic confirmation											
1. Final confirmation	A	A	A	A	A	A	A	A	A	} A = 82.22 P = 6.67 N = 11.11	
2. Association between specific <sup>18</sup> F-FDG PET finding and final confirmation technique used	A	A	P	A	A	A	A	A	A		
3. Histopathologic confirmation	A	A	A	A	A	A	A	A	A		
4. Nonhistopathologic confirmation	N	A	A	A	P	A	A	A	A		
5. Patients lost to clinical follow-up confirmation	N	N	N	A	A	N	P	A	A		
Guideline 6: sensitivity and specificity data											
1. Description of TP, FP, TN, and FN	P	P	A	A	A	P	A	A	P	} A = 57.14 P = 22.22 N = 20.63	
2. Qualitative explanation of FP and FN	N	A	N	A	P	P	A	P	A		
3. Specific region studied by <sup>18</sup> F-FDG PET	P	A	A	P	A	A	P	A	P		
4. Location and no. of primary tumors and other metastatic lesions detected by <sup>18</sup> F-FDG PET	A	A	P	A	A	A	A	A	A		
5. Confidence intervals	N	N	N	N	N	N	N	N	N		
6. Equivocal <sup>18</sup> F-FDG PET findings	N	A	N	A	A	P	A	A	A		
7. Data reported in patients and in corresponding lesions	A	A	P	A	A	A	A	A	A		
Guideline 7: change-in-management information											
1. <sup>18</sup> F-FDG PET-directed change in management	N/A	A	N/A	N/A	A	P	N/A	A	A	} A = 77.14 P = 17.14 N = 5.71	
2. Diagnostic tool used to make initial treatment decision	N/A	P	N/A	N/A	A	P	N/A	P	A		
3. Diagnostic tool used to make final treatment decision	N/A	A	N/A	N/A	A	P	N/A	A	A		
4. Medical treatment made both initially and after <sup>18</sup> F-FDG PET	N/A	P	N/A	N/A	A	N	N/A	N	A		
5. Correct or incorrect <sup>18</sup> F-FDG PET-directed management change	N/A	A	N/A	N/A	A	A	N/A	A	A		
6. Upstaging and downstaging	N/A	A	N/A	N/A	A	A	N/A	A	A		
7. <sup>18</sup> F-FDG PET in management algorithm	N/A	A	N/A	N/A	A	A	N/A	A	A		
Percentage of scores for each article											
A (%)	58.06	73.68	51.61	77.42	76.31	55.26	70.97	71.05	81.58		
P (%)	9.68	10.53	16.13	6.45	15.79	28.95	16.13	15.79	5.26		
N (%)	32.26	15.79	32.26	16.13	7.89	15.79	12.90	13.16	13.16		

\*Methodologic quality criteria guidelines formulated by Huebner et al. (25) and adapted for review of studies of patients with UPT in which <sup>18</sup>F-FDG PET is performed to detect primary tumor.

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

**TABLE 6**  
Assessment of Study Validity, Grade of Evidence, and Level of Efficacy

Study	Year	Study validity criteria*						Grade of evidence <sup>†</sup> of Kent and Larson (28)	Level of efficacy <sup>‡</sup> of Fryback and Thornbury (29)
		Control group clearly identified and ≥1 free of disease	Adequate reference test	Interpretation of diagnostic test blinded to reference test results	Interpretation of reference test blinded to diagnostic test results	Reference test	Other diagnostic procedures directly compared with <sup>18</sup> F-FDG PET		
Rege et al. (5)	1994	No	Yes	Yes	No	Histopathologic confirmation or follow-up	MRI	C	2
Schipper et al. (11)	1996	No	Yes	Yes	No		—	C	2
Kole et al. (16)	1998	No	Yes	Yes	No		—	C	5
Shinohara et al. (10)	1999	No	Yes	Yes	No		—	C	2
Hanasono et al. (13)	1999	No	Yes	Yes	No		CT or MRI	C	2
Greven et al. (15)	1999	No	Yes	Yes	No		CT, MRI, or DPB	C	2
Lassen et al. (8)	1999	No	Yes	Yes	No		—	C	4
Safa et al. (14)	1999	No	Yes	Yes	No		—	C	2
Gupta et al. (12)	1999	No	Yes	Yes	No		CT or MRI	C	4
Aassar et al. (35)	1999	No	Yes	Yes	No		CT or MRI	C	2
Lonneux and Reffad (7)	2000	No	Yes	Yes	No		—	C	4
Trampal et al. (9)	2000	No	Yes	Yes	No		—	C	4
Bohuslavizky et al. (6)	2000	No	Yes	Yes	No		—	C	4
Jungehülsing et al. (36)	2000	No	Yes	Yes	No		—	C	4
Cheng et al. (34)	2000	No	Yes	Yes	No		CT or MRI	C	2

\*Study validity criteria were adapted from Kent and Larson (28).

<sup>†</sup>Four grades of evidence were described by Kent and Larson (28): A and B grades = high-quality evidence; C grade = weak evidence; D grade = noncontributory.

<sup>‡</sup>Efficacy is assessed using hierarchic model of efficacy described by Fryback and Thornbury (29), where contribution to patients' management is analyzed: 1 = technical efficacy; 2 = diagnostic accuracy efficacy; 3 = diagnostic thinking efficacy; 4 = therapeutic efficacy; 5 = patient outcome efficacy; 6 = societal efficacy.

Assessment scores presented here were obtained by discussion of 2 investigators. DPB = direct panendoscopy with biopsy.

a nonindependent reference test because it is based in part on imaging results with <sup>18</sup>F-FDG PET. Clinical follow-up information is an independent reference test that was correctly applied in all studies, although follow-up times were too short. Only 1 study reported a follow-up of >12 mo (12), the minimum required to consider the reference test as unbiased. Other studies described the mean or range of follow-up times (6,15,16,35). Therefore, application of reference tests was considered partially inadequate. Eight studies (53.33%) collected data prospectively (5,8,11,12,14–16,36), whereas 4 studies are retrospective (6,7,13,35); in 3 studies, this information is not described (9,10,34). All studies included fewer than 35 patients, except for 1 study (6). Assessment of the validity and quality of the research methods classified all studies in grade of evidence C. Grade C is considered weak evidence and includes studies with several flaws in research methods, small sample sizes, or incomplete reporting; these studies present a narrow spectrum of generalizability (28). With regard to the assessment of the contribution to the patients' management according to the efficacy model described by Fryback and Thornbury (29), 8 studies (53%) reached level 2 (diagnostic accuracy efficacy) (5,10,11,13–15,34,35), 6 studies (40%) reached

level 4 (therapeutic efficacy) (6–9,12,36), and only 1 study reached level 5 (patient outcome efficacy) (16).

### Diagnostic Accuracy

Of 298 patients with UPT, <sup>18</sup>F-FDG PET detected the primary tumor in 43% of the patients (95% CI, 0.35–0.49), with a range of 7.69% (15) to 64.52% (12). Sensitivity of <sup>18</sup>F-FDG PET was 0.87 (95% CI, 0.81–0.92), with a range of 0.50 (15) to 1.00 (5,7,9,11,34,35). Specificity was 0.71 (95% CI, 0.64–0.78), with a range of 0.45 (7,15) to 1.00 (5,10,12,16,36). <sup>18</sup>F-FDG PET's sensitivity, specificity, diagnostic accuracy, and proportion of primary tumor detection for each of the studies are presented in Table 7.

Our subgroup analysis is presented in Figure 3, where 95% CI and estimated values of sensitivity and specificity in type I, type II, type A, and type B studies are shown. The 95% CI presented a Q test of heterogeneity with *P* = 0.65 for both sensitivity and specificity of the different types of studies. This indicated that there was global homogeneity between the types of studies, because the differences were not significant, and study results could be pooled to obtain a global estimation of sensitivity and specificity. Sensitivity of type I studies was 0.87 (95% CI, 0.78–0.93); type II,

**TABLE 7**  
Reported Sensitivity, Specificity, Accuracy, and Proportion of Primary Tumor Detection

Study	Year	Study type		n*	Results							
		Type I/II	Type A/B		TP	FP	TN	FN	S	Sp	Acc	% PD†
Rege et al. (5)	1994	II	B	4	2	0	2	0	1.00	1.00	1.00	50.00
Schipper et al. (11)	1996	I	A	16	4	3	9	0	1.00	0.75	0.81	25.00
Kole et al. (16)	1998	I	A	29	7	0	19	3	0.70	1.00	0.90	24.14
Shinohara et al. (10)	1999	I	A	4	2	0	1	1	0.67	1.00	0.75	50.00
Hanasono et al. (13)	1999	II	A	20	7	4	6	3	0.70	0.60	0.65	35.00
Greven et al. (15)	1999	II	B	13	1	6	5	1	0.50	0.45	0.46	7.69
Lassen et al. (8)	1999	I	A	20	9	4	5	2	0.82	0.56	0.70	45.00
Safa et al. (14)	1999	I	A	14	3	1	9	1	0.75	0.90	0.86	21.43
Gupta et al. (12)	1999	II	A	31	20	0	9	2	0.91	1.00	0.93	64.52
Aassar et al. (35)	1999	II	B	17	9	3	5	0	1.00	0.62	0.82	52.94
Lonneux and Reffad (7)	2000	I	A	24	13	6	5	0	1.00	0.45	0.75	54.17
Trampal et al. (9)	2000	I	A	9	4	1	4	0	1.00	0.80	0.89	44.44
Bohuslavizki et al. (6)	2000	I	A	52	20	6	22	4	0.83	0.79	0.81	38.46
Jungehülsing et al. (36)	2000	I	B	27	7	0	18	2	0.78	1.00	0.93	25.93
Cheng et al. (34)	2000	II	B	18	9	4	5	0	1.00	0.56	0.78	50.00

\*No. of patients included in meta-analysis (after excluding patients without confirmation of  $^{18}\text{F}$ -FDG PET study).

†Proportion of UPT patients in which  $^{18}\text{F}$ -FDG PET correctly detects primary tumor, presented in percentage.

Type I =  $^{18}\text{F}$ -FDG PET was performed when all diagnostic procedures carried out did not detect primary tumor; Type II =  $^{18}\text{F}$ -FDG PET was compared with CT or MRI after negative results in conventional diagnostic procedures; Type A = whole-body  $^{18}\text{F}$ -FDG PET study was carried out; Type B =  $^{18}\text{F}$ -FDG PET study only included head-neck-thorax region; TP = true-positive; FP = false-positive; TN = true-negative; FN = false-negative; S = sensitivity; Sp = specificity; Acc = diagnostic accuracy.

0.89 (95% CI, 0.77–0.95); type A, 0.87 (95% CI, 0.80–0.92); and type B, 0.92 (95% CI, 0.62–1.00). Estimated values of sensitivity were similar between the different types of studies and there was little dispersion in all 95% CI. Specificity of type I studies was 0.77 (95% CI, 0.69–0.84); type II, 0.58 (95% CI, 0.44–0.71); type A, 0.73 (95% CI, 0.65–0.80); and type B, 0.57 (95% CI, 0.34–0.77). Estimated values of specificity were lower in type II and type B studies. These also presented a greater dispersion of the 95% CI in comparison with type I and type A studies.

Figure 4 shows the ln OR and its 95% CI for each of the studies as well as in the pooled data. The ln OR in the pooled data was 2.50 (95% CI, 1.97–3.03), indicating that  $^{18}\text{F}$ -FDG PET produced statistically significant changes because the 95% CI of the ln OR did not include the value 0 (ln 1 = 0). If the 95% CI of the ln OR had included ln OR = 0, then OR = 1 (the intermediate step is  $e^0 = 1$ ), it would have indicated that  $^{18}\text{F}$ -FDG PET was not associated with statistically significant changes. Ten studies (66.67%) also presented 95% CI of the ln OR that did not include the value 0 (5–7,9,11,12,16,34–36); whereas, in the remaining 5 studies, the 95% CI of the ln OR included the value 0, indicating that in these studies the diagnostic test did not produce statistically significant changes (8,10,13–15).

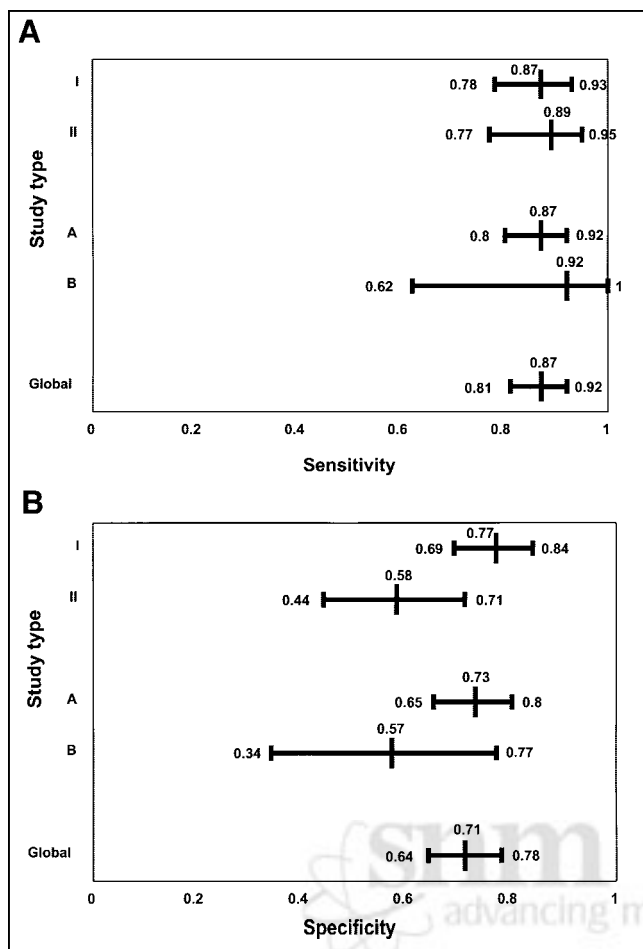
The variation in  $^{18}\text{F}$ -FDG PET's accuracy due to study characteristics was assessed using a multiple linear regression model (33), as described previously. The ln OR was calculated depending on study characteristics (type I or type II and type A or type B) to evaluate the variation in accuracy. The formula used was  $\ln \text{OR} = \alpha + \beta_1$  (type I or type

II) +  $\beta$  (type A or type B). The partial regression coefficients ( $\beta$ ) in the multiple regression model were  $\beta_1 = 0.72$  for type I or type II and  $\beta = 0.16$  for type A or type B; after replacing the  $\beta$  in the formula we obtained:  $\ln \text{OR} = \alpha + 0.72$  (type I or type II) +  $0.16$  (type A or type B). The coding used for the types of studies was for type I = 1, type II = 0 and for type A = 1, type B = 0. After substituting the coding values depending on the type of study, the following adjusted  $\beta$  values and their 95% CI were obtained:  $-0.73$  to  $+2.17$  with  $P = 0.30$  for timing of  $^{18}\text{F}$ -FDG PET (type I or type II) and  $-1.62$  to  $+1.93$  with  $P = 0.85$  for the region studied by  $^{18}\text{F}$ -FDG PET (type A or type B). This suggests that the variation in  $^{18}\text{F}$ -FDG PET's accuracy due to study characteristics is not statistically significant in either characteristic.

The positive LR was 3.048 (95% CI, 2.39–3.88), indicating that a positive result of  $^{18}\text{F}$ -FDG PET induced small changes in the pretest probability. However, the negative LR was 0.174 (95% CI, 0.11–0.27), indicating that when  $^{18}\text{F}$ -FDG PET was negative, it induced moderate changes in the pretest probability.

Figure 5 shows the summary ROC curve that suggested a correct trade-off between sensitivity and specificity. The threshold used in most studies favored sensitivity against specificity, as most of the studies are situated at the top of the diagram.

Figure 6 shows the funnel plots of sensitivity and specificity. The funnel plot of sensitivity did not suggest the existence of publication bias. On the other hand, the funnel plot of specificity shows an asymmetric distribution of the



**FIGURE 3.** Estimated values and 95% CI of sensitivity (A) and specificity (B) in different types of studies (type I, type II, type A, and type B) and in pooled data (global). Global = pooled data; type I =  $^{18}\text{F}$ -FDG PET was performed when all diagnostic procedures performed did not detect primary tumor; type II =  $^{18}\text{F}$ -FDG PET was compared with CT or MRI in double-blind study after negative results for primary tumor detection with other diagnostic procedures; type A = whole-body  $^{18}\text{F}$ -FDG PET study was performed; type B =  $^{18}\text{F}$ -FDG PET study only included head-neck-thorax region.

studies, suggesting the presence of publication bias. Several studies presented specificities close to 1.00 because they reported very few or no FP results. Apart from publication bias, incorrect application of the reference tests (confirmation bias) and inclusion of patients with high pretest probability of disease (inclusion bias) could be the cause of the asymmetric appearance described.

## DISCUSSION

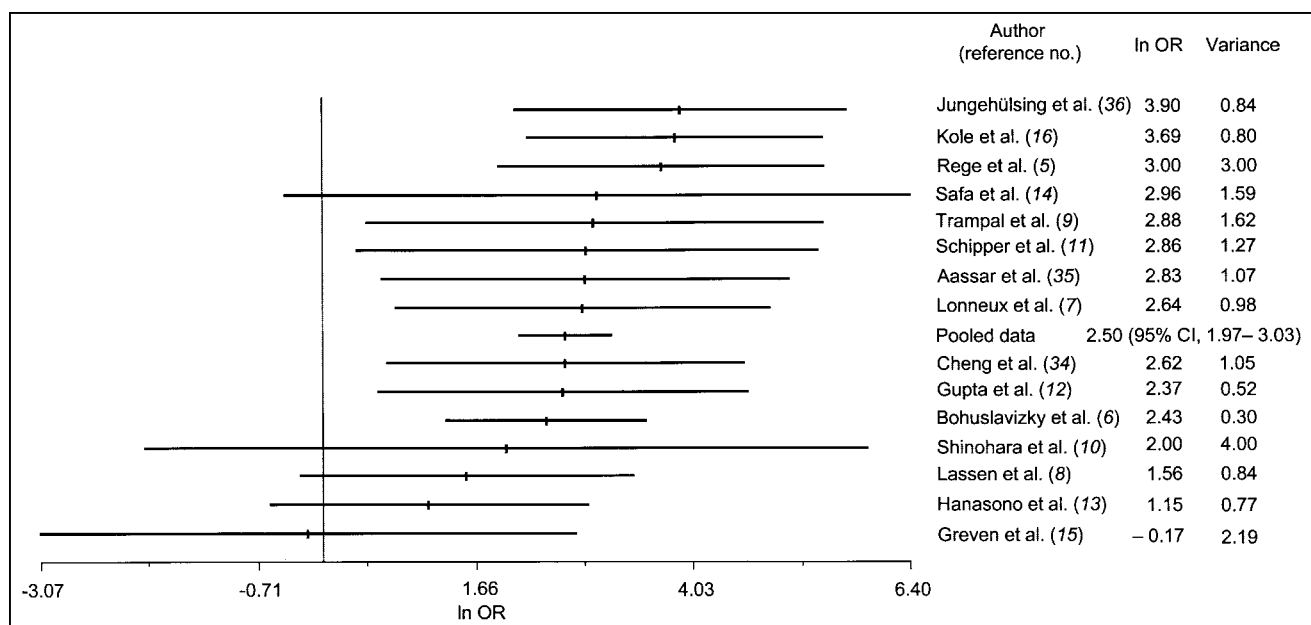
The results of the literature review and meta-analysis suggest that  $^{18}\text{F}$ -FDG PET could be useful in patients with UPT for the detection of the primary tumor.  $^{18}\text{F}$ -FDG PET presents intermediate specificity and high sensitivity, indicating the existence of few false-negative results. This is important in the management of oncologic patients and

suggests more benefits could be obtained if  $^{18}\text{F}$ -FDG PET was performed in the initial stages of the management process.

The relatively recent application of  $^{18}\text{F}$ -FDG PET to UPT patients accounts for the small number of studies available and the reason why some have been presented in congresses but not yet published (9,34). These have been included in the meta-analysis to avoid publication bias due to the greater probability that studies with positive or statistically significant results would be published than studies that do not have these results (24,37–39) and the delay of up to 3 y observed between congress presentations and complete publication of studies (24). On the other hand, studies published in any language are included to prevent the so-called Tower of Babel bias described by Grégoire et al. (23) and, thus, increase accuracy and decrease systematic errors (40). This bias refers to the fact that investigators working in a language other than English could be sending studies with positive results to international journals. When negative or nonsignificant results are found, the authors could be less confident about having it published in an international journal written in English and would, thus, only send it to a national journal in their language. By only including studies published in English, studies with negative results could have been left out.

Although the funnel plot of sensitivity did not suggest evidence of publication bias, the funnel plot of specificity suggested the presence of bias. As mentioned above, publication bias, inclusion bias, or confirmation bias could cause the asymmetric distribution. Inclusion bias could be the consequence of the inclusion of patients with a high pretest probability of disease and of the selection of those patients who present negative findings in all other diagnostic procedures. Confirmation bias of the negative results could be related to the incorrect application of the reference tests; they are considered TN after a relatively short clinical follow-up and the use of diagnostic tools that do not present 100% diagnostic accuracy. To avoid these biases,  $^{18}\text{F}$ -FDG PET could be performed at the onset of the diagnostic algorithm, and the follow-up times of the negative results could be lengthened, because the performance of invasive procedures in patients with negative results is not justified.

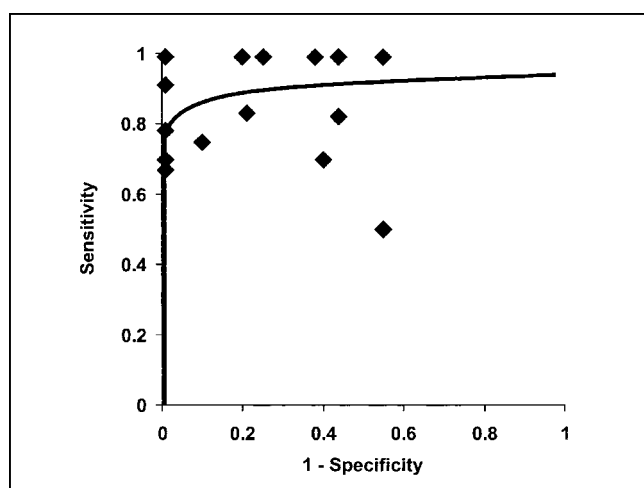
Sensitivity analysis revealed that differences in the methodologic quality did not correlate with differences in study results. However, application of the methodologic quality guidelines was only possible in 9 studies (5,6,8,12,13,15,16,35,36) because the complete article was not available in the rest. Thus, the results obtained must be interpreted cautiously, as the nonincluded studies could present different results. Analysis of the results of the methodologic quality guidelines suggest that the studies analyzed presented enough information overall and satisfied most of the requirements established. However, some changes that would improve the quality of the information described could be performed. It must be mentioned that the criteria adherence scores are not an indication of the validity of a



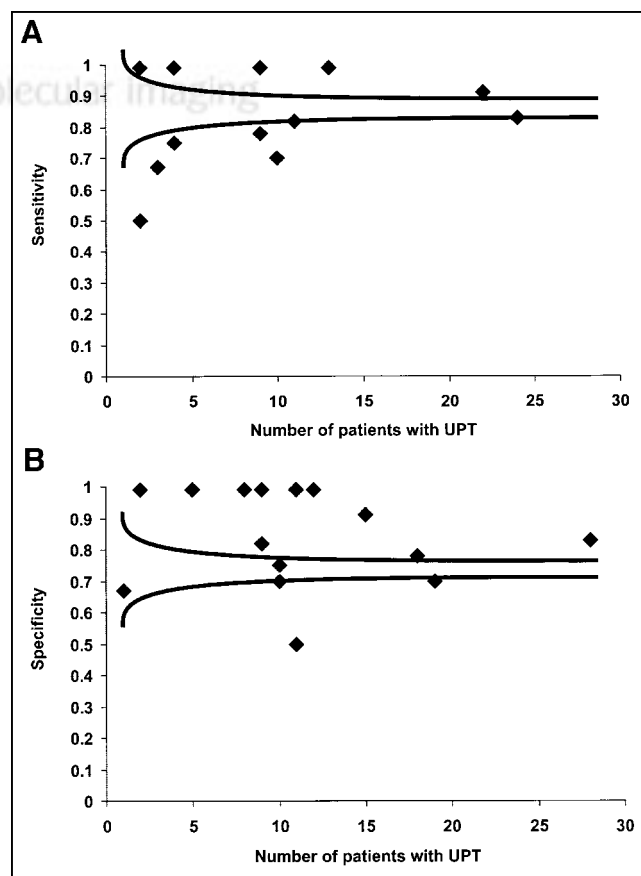
**FIGURE 4.** ln OR in studies included in meta-analysis and in pooled data. Calculated values of ln OR and variance in all studies are shown as well as ln OR and 95% CI in pooled data.

study. What has been evaluated is the amount of information supplied in the article and its compliance with established guidelines or requirements. The analysis of the data can show features that have not been described and that could have a repercussion on the interpretation and results of  $^{18}\text{F}$ -FDG PET—for example, the presence of comorbid conditions, only described in 1 article (12), or the measurement of glycemia before the administration of  $^{18}\text{F}$ -FDG, only described in 2 articles (6,36).

The 15 studies included in the meta-analysis presented methodologic defects in the grade of evidence analysis, concerning reference test application, sample sizes, or incomplete reporting, and, thus, were classified as weak evi-



**FIGURE 5.** Summary ROC curve for primary tumor detection with  $^{18}\text{F}$ -FDG PET in patients with UPT based on 15 studies included in meta-analysis.



**FIGURE 6.** Funnel plots of sensitivity and specificity. (A) Funnel plot of sensitivity shows no publication bias. (B) Funnel plot of specificity suggests presence of bias.

dence. Study populations were small and, in some cases, selected by specific characteristics. Several studies focused on cervical lymph nodes from UPT (11,13–15,34,35) and some focused only on squamous cell carcinoma metastases (13–15), the most frequent histopathology in this location (2). This could justify the observation that the most frequent histopathology in our meta-analysis was squamous cell carcinoma (59.73%), whereas adenocarcinoma has been described as the most frequent (45%–61%) in a published review on UPT (2). However, in the studies on UPT non-selected by specific characteristics (5–10,16,36), the most frequent is also squamous cell carcinoma (45.22%), so that there could have been a reference bias of the patients even in these studies (Fig. 1).

The application of the reference tests is an important methodologic aspect. The availability of 2 reference tests—one dependent on the image, the histopathologic confirmation, and another independent one, the clinical follow-up—makes it possible to evaluate the presence of verification bias due to the incorporation of the imaging information into the final diagnosis (30). The clinical follow-up is frequently performed with times that are too short and with diagnostic procedures whose diagnostic accuracy is <100%, so that it never reaches the certainty level of the histopathologic confirmation (30). However, this is not justified in patients in whom no alterations are observed in  $^{18}\text{F}$ -FDG PET.

The contribution of an imaging procedure to the management of a patient is difficult to measure because many variables and effects must be considered. The Fryback and Thornbury model assigns each of these variables an efficacy level that would indicate the contributions of this study (29). The efficacy level 5, patient outcome efficacy, is reached by only 1 study (16), which analyzes the survival of the patients on the basis of the contribution of the image, an important aspect in the validation of expensive procedures such as  $^{18}\text{F}$ -FDG PET. The study suggests that benefits are obtained in survival, but in a limited number of patients. Six studies (6–9,12,36) reach the level of therapeutic efficacy, because they describe the changes in the treatments applied as a consequence of the imaging results.

The homogeneity observed in the 95% CI of sensitivity and specificity of the different types of studies makes it possible to combine all information and obtain a combined effect (31). The 95% CI of sensitivity show a similar estimation and little dispersion of the 95% CI of the different types of studies. However, the 95% CI of the specificity show a lower estimation and greater dispersion of the 95% CI in type II and type B studies. The greater dispersion of the 95% CI could be due to the fact that the number of patients from type II and type B studies is lower than that from type I and type A, as is described in the Results. Three type II and type B studies (15,34,35) and 1 type II and type A study (13) present many FP results that correspond to a high proportion of patients from these types of studies and could explain the lower estimate of specificity. The authors

of these studies justify these results as inflammatory alterations (35), elevations of the standardized uptake value slightly above the cutoff level, which is reinterpreted as normal after elevating the cutoff (34), or the small size of the primary tumor associated with the elevated background activity and benign uptakes (15).

The summary ROC curve shows a good tradeoff between sensitivity and specificity, although the sensitivity values predominate above the specificity ones. The  $\ln \text{OR}$  places most of the studies in values that indicate that the contribution of  $^{18}\text{F}$ -FDG PET is significant. The positive LR suggests small changes, whereas the negative LR suggests moderate changes.

## CONCLUSION

The results obtained indicate that  $^{18}\text{F}$ -FDG PET could be useful in patients with UPT for the detection of the primary tumor.  $^{18}\text{F}$ -FDG PET has intermediate specificity and high sensitivity, indicating the existence of few false-negative results, an important feature in the management of oncologic patients that could suggest its utility in the initial stages of the management process. However, more data are needed to determine the clinical utility of  $^{18}\text{F}$ -FDG PET in assessing patients with UPT. Evaluation of the role of  $^{18}\text{F}$ -FDG PET in UPT patients' management has yet to be properly assessed with methodologically rigorous studies. In these studies, the incremental value of  $^{18}\text{F}$ -FDG PET over other diagnostic tests must be demonstrated. If evidence of favorable changes in management is finally reported, as the preliminary data presented in this study suggest, a cost-effectiveness study could be performed. Thus, analysis of costs reduction, because of avoiding unnecessary procedures, and improvement of accuracy in primary tumor detection when  $^{18}\text{F}$ -FDG PET is used instead of other procedures will be properly evaluated. This way the applicability of  $^{18}\text{F}$ -FDG PET in this clinical situation could be assessed and rational recommendations could be made for the use of  $^{18}\text{F}$ -FDG PET in UPT patients presented. On the other hand, most of the studies analyzed were done using instrumentation that, at this time, may not be considered state of the art. Finally, future studies will have to assess whether ROC curves will be significantly improved by the introduction of combined CT/PET systems and the introduction of new software fusion approaches using techniques such as mutual information theory. If the diagnostic performance improves significantly in the future because of technical advances, new studies will have to assess the role of  $^{18}\text{F}$ -FDG PET in UPT.

## ACKNOWLEDGMENT

This study was partially financed by the Spanish Health Technology Assessment Agency (Agencia de Evaluación de Tecnologías Sanitarias) of the Instituto de Salud Carlos III of Madrid, Spain, with file number 00/10028 and granted in

September 2000. The final report was presented in November 2001.

## REFERENCES

- Abbruzzese JL, Abbruzzese MC, Hess KR, Raber MN, Lenzi R, Frost P. Unknown primary carcinoma: natural history and prognostic factors in 657 consecutive patients. *J Clin Oncol*. 1994;12:1272–1280.
- Daugaard G. Unknown primary tumours. *Cancer Treatment Rev*. 1994;20:119–147.
- Abbruzzese JL, Abbruzzese MC, Lenzi R, Hess KR, Raber MN. Analysis of a diagnostic strategy for patients with suspected tumors of unknown origin. *J Clin Oncol*. 1995;13:2094–2103.
- Leonard RJ, Nystrom JS. Diagnostic evaluation of patients with carcinoma of unknown primary tumor site. *Semin Oncol*. 1993;20:244–250.
- Rege S, Maass A, Chaiken L, et al. Use of positron emission tomography with fluorodeoxyglucose in patients with extracranial head and neck cancers. *Cancer*. 1994;73:3047–3058.
- Bohuslavizki KH, Klutmann S, Kröger S, et al. FDG PET detection of unknown primary tumors. *J Nucl Med*. 2000;41:816–822.
- Lonneux M, Reffad AM. Metastases from unknown primary tumor: PET-FDG as initial diagnostic procedure? *Clin Positron Imaging*. 2000;3:137–141.
- Lassen U, Daugaard G, Eigved A, Damgaard K, Friberg L. <sup>18</sup>F-FDG whole body positron emission tomography (PET) in patients with unknown primary tumours (UPT). *Eur J Cancer*. 1999;35:1076–1082.
- Trampal C, Sörensen J, Engler H, Långström B. <sup>18</sup>F-FDG whole body positron emission tomography (PET) in the detection of unknown primary tumors [abstract]. *Clin Positron Imaging*. 2000;3:160.
- Shinohara M, Tomita M, Ohira S, Nishimura A. Evaluation of <sup>18</sup>F-FDG positron emission tomography (PET) in the detection of unknown primary tumors. *Hokkaido Igaku Zasshi*. 1999;74:249–256.
- Schipper JH, Schrader M, Arweiler D, Muller S, Sciuc J. Positron emission tomography for primary tumor detection in lymph node metastases with unknown primary tumor. *HNO*. 1996;44:254–257.
- Gupta NC, Nicholson P, Bloomfield SM. FDG-PET in the staging work-up of patients with suspected intracranial metastatic tumors. *Ann Surg*. 1999;230:202–206.
- Hanasono MM, Kunda LD, Segall GM, Ku GH, Terris DJ. Uses and limitations of FDG positron emission tomography in patients with head and neck cancer. *Laryngoscope*. 1999;109:880–885.
- Safa AA, Tran LM, Rege S, et al. The role of positron emission tomography in occult primary head and neck cancers. *Cancer J Sci Am*. 1999;5:214–218.
- Greven KM, Keyes JW Jr, Williams DW III, McGuiert WF, Joyce WT III. Occult primary tumors of the head and neck: lack of benefit from positron emission tomography imaging with 2-[F-18]fluoro-2-deoxy-D-glucose. *Cancer*. 1999;86:114–118.
- Kole AC, Nieweg OE, Pruim J, et al. Detection of unknown primary tumors using positron emission tomography. *Cancer*. 1998;82:1160–1166.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-analysis. *J Clin Epidemiol*. 1995;48:167–171.
- Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450–455.
- Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667–676.
- Thacker SB. Meta-analysis: a quantitative approach to research integration. *JAMA*. 1988;259:1685–1689.
- L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med*. 1987;107:224–233.
- Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127:380–387.
- Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol*. 1995;48:159–163.
- Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts: a meta-analysis. *JAMA*. 1994;272:158–162.
- Huebner RH, Park KC, Schepherd JE, et al. A meta-analysis of the literature for whole-body FDG PET detection of recurrent colorectal cancer. *J Nucl Med*. 2000;41:1177–1189.
- Gould MK, Maclean CC, Kushner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA*. 2001;285:914–924.
- Schelbert HR, Hoh CK, Royal HD, et al. Procedure guideline for tumor imaging using fluorine-18-FDG. *J Nucl Med*. 1998;39:1302–1305.
- Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol*. 1992;27:245–254.
- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11:88–94.
- Thornbury JR, Kido DK, Mushlin AI, Phelps CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Invest Radiol*. 1991;26:829–835.
- Lau J, Ioannidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med*. 1997;127:820–826.
- Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;12:1293–1316.
- Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol*. 1995;141:680–689.
- Cheng KC, Yeung HWD, Macapinlac H, Larson SM. Detection of occult primary tumors in patients presenting with cervical lymph node metastases by <sup>18</sup>F-FDG PET [abstract]. *Eur J Nucl Med*. 2000;27:902.
- Aassar OS, Fischbein NJ, Caputo GR, et al. Metastatic head and neck cancer: role and usefulness of FDG PET in locating occult primary tumors. *Radiology*. 1999;210:177–181.
- Jungehülsing M, Scheidhauer K, Damm M, et al. 2[<sup>18</sup>F]-fluoro-2-deoxy-D-glucose positron emission tomography is a sensitive tool for the detection of occult primary cancer (carcinoma of unknown primary syndrome) with head and neck lymph node manifestation. *Otolaryngol Head Neck Surg*. 2000;123:294–301.
- Cook DJ, Guyatt GH, Ryan G, et al. Should unpublished data be included in meta-analyses? current convictions and controversies. *JAMA*. 1993;269:2749–2753.
- Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med*. 1987;6:11–29.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867–872.
- Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet*. 1996;347:363–366.