

Test Performance of Positron Emission Tomography and Computed Tomography for Mediastinal Staging in Patients with Non–Small-Cell Lung Cancer

A Meta-Analysis

Michael K. Gould, MD, MS; Ware G. Kuschner, MD; Chara E. Rydzak, BA; Courtney C. Maclean, BA; Anita N. Demas, MD; Hidenobu Shigemitsu, MD; Jo Kay Chan, BS; and Douglas K. Owens, MD, MS

Purpose: To compare the diagnostic accuracy of computed tomography (CT) and positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) for mediastinal staging in patients with non–small-cell lung cancer and to determine whether test results are conditionally dependent (the sensitivity and specificity of FDG-PET depend on the presence or absence of enlarged mediastinal lymph nodes on CT).

Data Sources: Computerized search of MEDLINE, EMBASE, BIOSIS, and CancerLit through March 2003 and reference lists of retrieved studies and review articles.

Study Selection: Studies in any language that examined FDG-PET for mediastinal staging in patients with known or suspected non–small-cell lung cancer, enrolled at least 10 participants (including at least 5 participants with mediastinal metastasis), and provided enough data to permit calculation of sensitivity and specificity for identifying lymph node involvement.

Data Extraction: One reviewer (of non–English-language studies) or 2 reviewers (of English-language studies) independently evaluated studies for inclusion, rated methodologic quality, and abstracted relevant data.

Data Synthesis: Thirty-nine studies met inclusion criteria. Methodologic quality varied, but few aspects of study quality affected diagnostic accuracy. The authors constructed summary receiver-

operating characteristic curves for CT and FDG-PET. Positron emission tomography with 18-fluorodeoxyglucose was more accurate than CT for identifying lymph node involvement ($P < 0.001$). For CT, median sensitivity and specificity were 61% (interquartile range, 50% to 71%) and 79% (interquartile range, 66% to 89%), respectively. For FDG-PET, median sensitivity and specificity were 85% (interquartile range, 67% to 91%) and 90% (interquartile range, 82% to 96%), respectively. Fourteen studies provided information about the conditional test performance of CT and FDG-PET. Positron emission tomography with 18-fluorodeoxyglucose was more sensitive but less specific when CT showed enlarged lymph nodes (median sensitivity, 100% [interquartile range, 90% to 100%]; median specificity, 78% [interquartile range, 68% to 100%]) than when CT showed no lymph node enlargement (median sensitivity, 82% [interquartile range, 65% to 100%]; median specificity, 93% [interquartile range, 92% to 100%]; $P = 0.002$).

Conclusions: Positron emission tomography with 18-fluorodeoxyglucose is more accurate than CT for mediastinal staging. Positron emission tomography with 18-fluorodeoxyglucose is more sensitive but less specific when CT shows enlarged mediastinal lymph nodes.

Ann Intern Med. 2003;139:879-892.

For author affiliations, see end of text.

See editorial comment on pp 950-951.

www.annals.org

Accurate mediastinal staging is crucial in managing patients with non–small-cell lung cancer. Regional lymph node status is an important determinant of prognosis, and decisions about treatment depend critically on tumor stage. Conventional methods for mediastinal staging include computed tomography (CT) and various biopsy procedures. However, CT has poor sensitivity and specificity for identifying mediastinal metastases (1–3), and biopsy procedures are inconvenient and potentially risky.

Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) is a promising but expensive functional imaging test that is rapidly gaining acceptance as a tool for lung cancer staging (4, 5). Positron emission tomography with 18-fluorodeoxyglucose identifies malignant cells in tumors and lymph nodes on the basis of their increased metabolic rate (6). In the past decade, several studies of PET imaging for mediastinal staging were published. These studies suggested that FDG-PET is more accurate than CT for identifying mediastinal metastases. However, most were small and potentially limited by other methodologic shortcomings. In addition, previous studies

have not systematically addressed the conditional test performance of FDG-PET and CT. Conditional test performance refers to the possibility that the sensitivity and specificity of 1 test might differ depending on the results of the other test (7). The results of FDG-PET and CT might be mutually dependent, despite the fact that they identify malignant lymph nodes by different mechanisms. In a preliminary analysis, we found that FDG-PET was more sensitive but less specific in patients with lymph node enlargement on CT (8). If confirmed, this finding has important implications for selecting and interpreting tests in mediastinal staging. For example, if FDG-PET is more sensitive when lymph node enlargement is present on CT, then a negative PET result would “rule out” disease more reliably (because its negative predictive value would be higher). Consequently, confirmatory mediastinal biopsy might not be necessary in some of these patients, especially when pretest probability is low.

We performed this meta-analysis to compare the accuracy of FDG-PET and CT for identifying mediastinal metastasis in patients with non–small-cell lung cancer. We

Context

Is computed tomography (CT) or positron emission tomography with 18-fluorodeoxyglucose (FDG-PET) better for mediastinal staging of non-small-cell lung cancer?

Contribution

This synthesis of 39 studies found that FDG-PET was more accurate than CT for identifying lymph node involvement. Positron emission tomography with 18-fluorodeoxyglucose was more sensitive but less specific when CT showed enlarged nodes than when CT showed no node enlargement.

Implications

Positron emission tomography with 18-fluorodeoxyglucose is more accurate than CT for mediastinal staging. Because FDG-PET has more true-positive and false-positive findings in patients with enlarged nodes, positive findings warrant biopsy confirmation. Interpretation of negative FDG-PET findings should rely heavily on pretest probability of metastasis regardless of CT findings.

—The Editors

also aimed to determine whether the results of FDG-PET and CT are conditionally dependent, that is, whether the sensitivity and specificity of FDG-PET depend on the presence or absence of lymph node enlargement on CT. Finally, we explored whether various aspects of study methods affected diagnostic accuracy.

METHODS

We used systematic review methods to identify potentially relevant studies, assess studies for eligibility, evaluate study quality, and derive summary estimates of diagnostic test performance (9–12). We previously used similar methods to evaluate the accuracy of FDG-PET imaging for diagnosis of pulmonary nodules and mass lesions (13). Additional details about our methods can be found in the Appendix (available at www.annals.org).

Study Identification

We attempted to identify all published studies that examined FDG-PET imaging for mediastinal staging in patients with known or suspected non-small-cell lung cancer. We sought studies that evaluated both FDG-PET and CT, but we did not attempt to identify studies that examined only CT for mediastinal staging. An investigator and a professional librarian searched MEDLINE, CancerLit, and EMBASE databases in August 2001 and repeated searches in June 2002 (Appendix Table 1, available at www.annals.org). We updated the literature search in MEDLINE, EMBASE, Current Contents, and BIOSIS through 27 March 2003 as part of a technology assessment performed for the U.S. Department of Veterans Affairs (Appendix Table 2, available at www.annals.org). We aug-

mented our computerized literature searches by manually reviewing the reference lists of identified studies and review articles. We included studies published in any language but did not include abstracts. For English-language studies, 2 investigators independently evaluated studies for inclusion, rated the methodologic quality of included studies, and abstracted relevant data. Disagreements were resolved by discussion. One reviewer performed these tasks for non-English-language studies. Reviewers were blinded to journal, author, institutional affiliation, and date of publication.

Study Eligibility

We included studies that examined FDG-PET imaging for mediastinal lymph node staging in patients with known or suspected non-small-cell lung cancer; enrolled at least 10 participants, including at least 5 participants with lymph node metastases; and provided enough data to permit calculation of sensitivity and specificity for identifying malignant lymph node involvement.

Study Quality

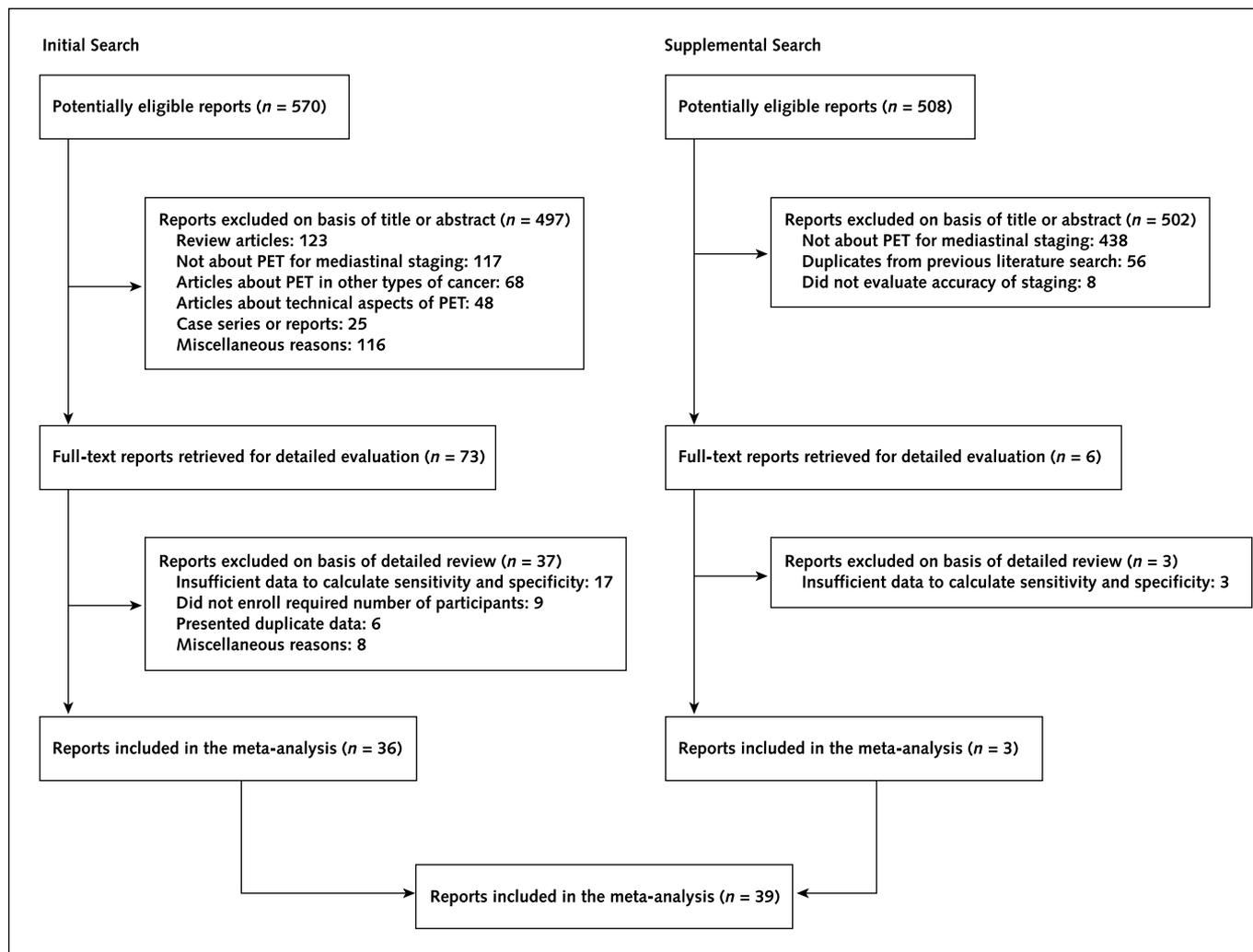
We adapted an existing instrument (11, 13) to examine 7 aspects of study quality: technical quality of the index tests, technical quality and application of the reference test, independence of test interpretation, description of the study population, cohort assembly, sample size, and unit of analysis (Appendix Table 3, available at www.annals.org).

Data Abstraction

We abstracted data about the demographic characteristics of participants, the prevalence of malignant lymph node involvement, and the sensitivity and specificity of CT and FDG-PET for identifying malignant lymph nodes. For studies that reported results by using the patient as the unit of analysis, we determined the ability of CT and FDG-PET to distinguish ipsilateral or contralateral mediastinal lymph node involvement (N2 or N3) from hilar, intrapulmonary, or no lymph node involvement (N0 or N1). This distinction is critical because involvement of N2 or N3 nodes usually indicates non-surgically treatable disease. When it was not possible to make this distinction, we determined test sensitivity and specificity for distinguishing N0 lymph node status from N1, N2, or N3 lymph node status. For studies in which the individual patient was not the unit of analysis, we determined the test sensitivity and specificity for identifying malignant lymph nodes or lymph node stations. Because observations are not independent when several lymph nodes from the same patient are analyzed separately, these studies may yield biased estimates of diagnostic test performance. Therefore, we analyzed data from these studies separately.

To determine whether the sensitivity and specificity of FDG-PET depended on the presence or absence of enlarged nodes on CT, we recorded the results of FDG-PET, CT, and the reference test or tests for each patient. This enabled us to derive separate estimates for the sensitivity

Figure 1. Reports evaluated for inclusion in the meta-analysis.



The initial search took place from 1966 through 1 June 2002, and the supplemental search took place from 1998 through 27 March 2003. PET = positron emission tomography.

and specificity of FDG-PET in patients with and without lymph node enlargement on CT.

Data Synthesis and Statistical Analysis

For each study, we constructed 2×2 contingency tables in which all participants were classified as having positive (N2 or N3) or negative (N0 or N1) results and as having or not having mediastinal lymph node involvement as determined by the reference test or tests. We calculated the true-positive rate (true-positive rate = sensitivity), the false-positive rate (false-positive rate = $1 - \text{specificity}$), and the log odds ratio (log odds true-positive rate - log odds false-positive rate) for CT and FDG-PET. The log odds ratio is a measure of diagnostic test performance that accounts for the correlation between the true-positive rate and the false-positive rate. We calculated exact 95% CIs for the true-positive rate and the false-positive rate on the basis of the binomial distribution (14).

To derive summary estimates of diagnostic test perfor-

mance, we constructed summary receiver-operating characteristic (ROC) curves by using the method of Moses (12, 13, 15, 16), which confirmed that the curves were symmetrical and could be described by a single parameter, the summary log odds ratio. Because this method requires the use of a correction factor when the reported sensitivity or specificity is 100%, we calculated the summary diagnostic odds ratios by using a fixed-effects model (17), or a random-effects model when there was evidence of heterogeneity (18), and reported results derived from these models. Because the summary log odds ratio is difficult to interpret clinically, we express our results in terms of the maximum joint sensitivity and specificity (12), a transformation of the summary log odds ratio that is a global measure of diagnostic accuracy, similar to the area under the ROC curve. The maximum joint sensitivity and specificity is the point on the summary ROC curve at which sensitivity and specificity are equal. It varies from 0.5 for a test that pro-

vides no diagnostic information to 1.0 for a test that is perfect.

We used meta-regression to make all statistical comparisons (19), with 1 exception. To compare the sensitivity and specificity of FDG-PET in patients with and without lymph node enlargement, we used discriminant function analysis (20) and a nonparametric permutation test (21). We considered a 2-sided *P* value less than 0.05 to be significant for all statistical tests.

Sensitivity Analysis

In prespecified analyses, we examined the effect of year of publication, language, and individual aspects of study quality on the diagnostic accuracy of FDG-PET. We used meta-regression to compare groups of studies. To check for publication bias, we created inverted funnel plots of individual study log odds ratios plotted against sample size (22).

Role of the Funding Source

The funding source had no role in the collection, analysis, or interpretation of the data or in the decision to publish the manuscript.

RESULTS

Our initial search identified 570 studies, including 73 studies that were potentially relevant to mediastinal staging with FDG-PET (Figure 1). Of these, we excluded 29 studies because the number of participants was not sufficient (23–31), data necessary to permit calculation of sensitivity and specificity were not provided (25, 28, 32–46), or duplicate data were reported (32, 47–51). Four additional studies were excluded because they evaluated FDG imaging with a modified gamma camera (52–55). We also excluded 3 non-English-language papers that were review articles or case reports (56–58). Another study of mediastinal staging was excluded because almost one third of the participants had small-cell lung cancer or mesothelioma (59).

The supplemental search conducted for the Department of Veterans Affairs identified 508 studies, including 70 studies that were potentially relevant to mediastinal staging with FDG-PET (Figure 1). Of these, 6 studies that had not been identified previously were judged to be potentially eligible for inclusion and underwent detailed review. Three of these studies were excluded because they did not present enough data to permit calculation of sensitivity and specificity (60–62).

Study Description

Thirty-nine studies met the inclusion criteria (63–101). Of these, 28 studies reported results by using the patient as the unit of analysis, 6 studies reported results by using lymph nodes or lymph node stations as the unit of analysis, and 5 studies reported results in both ways (Appendix Table 4, available at www.annals.org). The median number of participants per study was 51 (range, 18 to 237). The mean age of participants was 56 to 69 years, and the median proportion of male participants was 64% (range, 48%

to 99%). In studies that reported results by using the patient as the unit of analysis, the median prevalence of malignant lymph nodes was 32% (range, 5% to 64%). In studies that reported results by using lymph nodes or lymph node stations as the unit of analysis, the median prevalence of malignancy was lower (median, 16% [range, 7% to 37%]; *P* = 0.001). Ten studies provided usable data for FDG-PET but not for CT (73, 85, 89, 91–93, 96, 98, 99, 101).

Study Quality

Because our criteria for assessing quality were stringent, no study met all of them. Seventeen studies (44%) met at least 70% of the 22 criteria on our study quality checklist (63–67, 70, 71, 73–75, 77, 79, 90, 92, 94, 97, 98). Five studies met fewer than 50% of the criteria (82, 85, 93, 95, 100). Appendix Table 4 (available at www.annals.org) shows selected aspects of methodologic quality for each study. In general, studies followed guidelines published by the Society of Nuclear Medicine for performing FDG-PET imaging. However, only 11 studies (28%) indicated that participants with hyperglycemia were excluded. Most studies adequately described the technical aspects of CT, although only 32% specified the use of spiral CT or an acquisition time of 2 seconds or less. While more than 90% of the studies required positive results from biopsy specimens to confirm mediastinal metastasis, only 47% required thoracotomy with systematic sampling of both normal- and abnormal-appearing lymph nodes at all accessible mediastinal stations to verify the absence of mediastinal involvement. In 56% of the studies, readers of FDG-PET and CT were blinded to the final diagnosis, and imaging tests were interpreted independently in less than half of the studies. Ninety percent of the studies enrolled a clinically relevant cohort of participants with known or suspected non-small-cell lung cancer, but participants were enrolled prospectively in only 51% of the studies. Characteristics of participants were completely described in just over half of the studies. Almost 45% of the studies enrolled at least 35 participants with lymph node metastases or 35 participants without lymph node metastases.

Diagnostic Accuracy of CT and FDG-PET: Patient-Level Data

In studies that reported results by using the individual patient as the unit of analysis, FDG-PET was more accurate than CT for identifying mediastinal metastasis (*P* < 0.001). The median sensitivity and specificity of CT were 61% (interquartile range, 50% to 71%) and 79% (interquartile range, 66% to 89%), respectively (Figure 2, Table). The median sensitivity and specificity of FDG-PET were 85% (interquartile range, 67% to 91%) and 90% (interquartile range, 82% to 96%), respectively (Figure 3, Table).

The maximum joint sensitivity and specificity of CT was 70% (95% CI, 67% to 73%), indicating that diagnostic accuracy was only fair. Sensitivity was 59% (CI, 52% to 66%) at the point on the summary ROC curve that cor-

responded to the median specificity of 79% (Figure 4). Corresponding likelihood ratios for positive and negative CT results were 2.8 and 0.5, respectively.

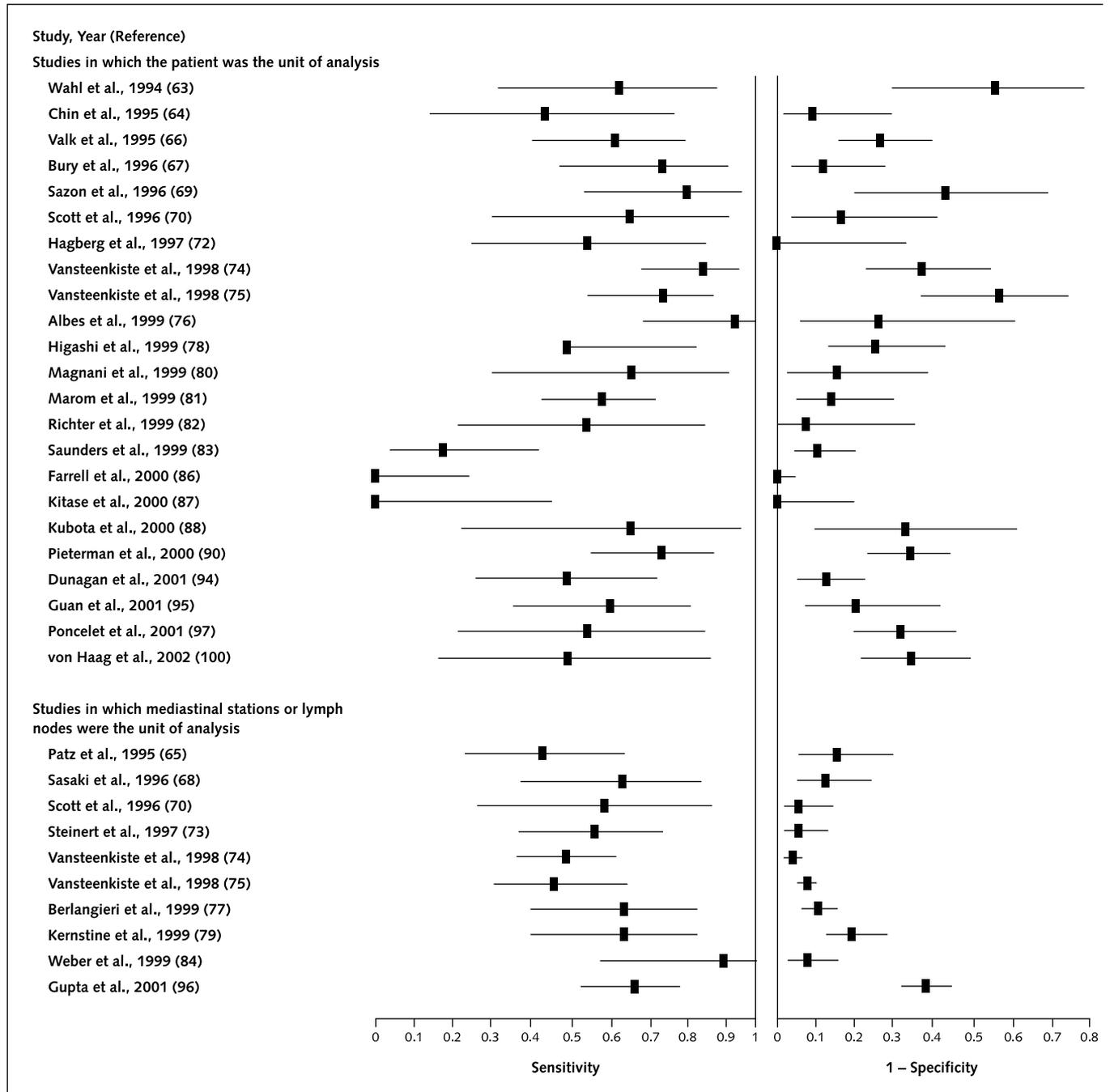
The maximum joint sensitivity and specificity of FDG-PET was 86% (CI, 83% to 88%), indicating that diagnostic accuracy was very good. Sensitivity was 81% (CI, 74% to 86%) at the point on the summary ROC curve that corresponded to the median specificity of 90% (Figure 4). Corre-

sponding likelihood ratios for positive and negative FDG-PET results were 8.1 and 0.2, respectively.

Diagnostic Accuracy of CT and FDG-PET: Lymph Node- or Lymph Node Station-Level Data

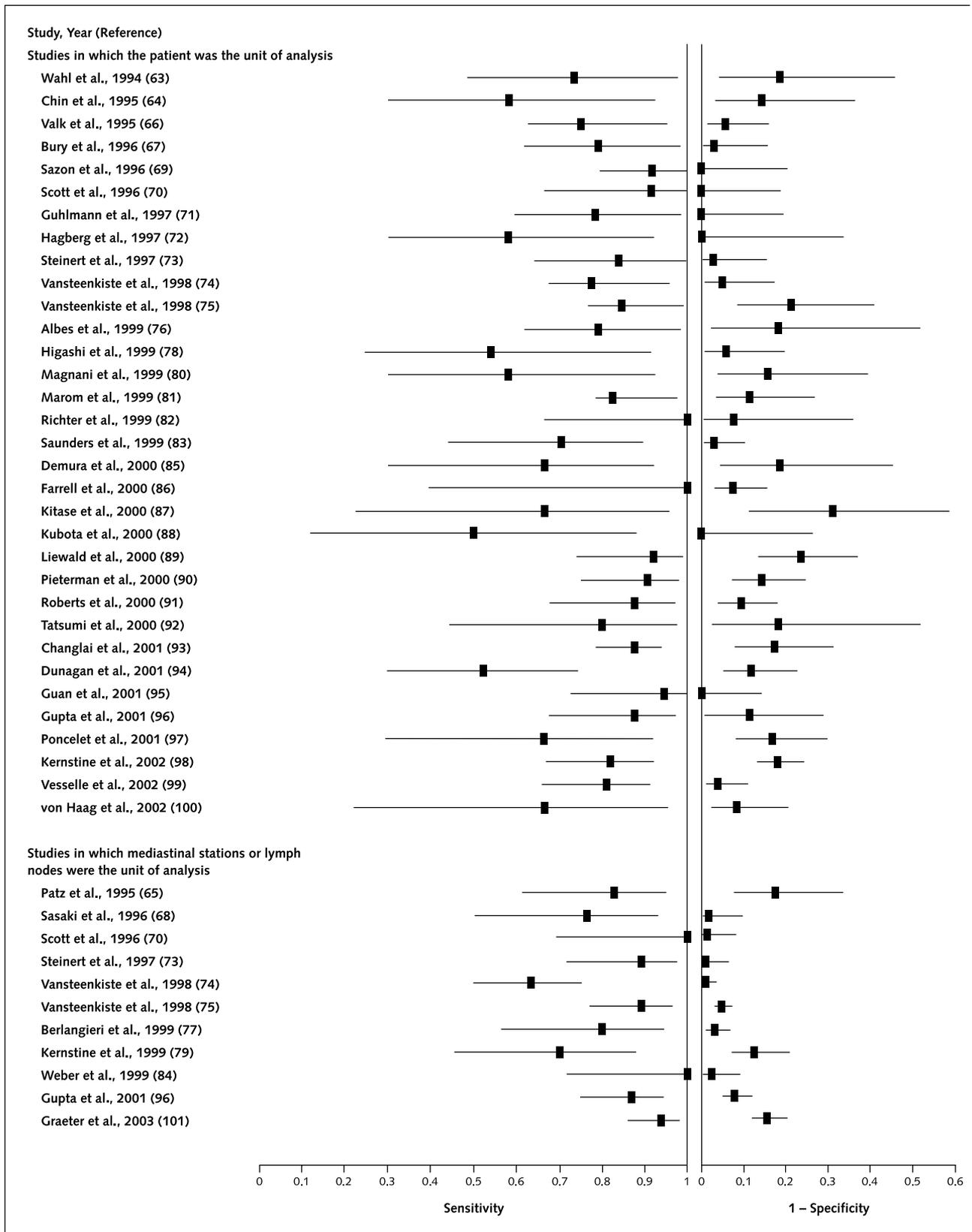
In studies that reported results by using lymph nodes or lymph node stations as the unit of analysis, the median sensitivity and specificity of CT were 62% (interquartile range,

Figure 2. Individual study estimates of sensitivity and 1 – specificity of computed tomography for identifying mediastinal metastasis.



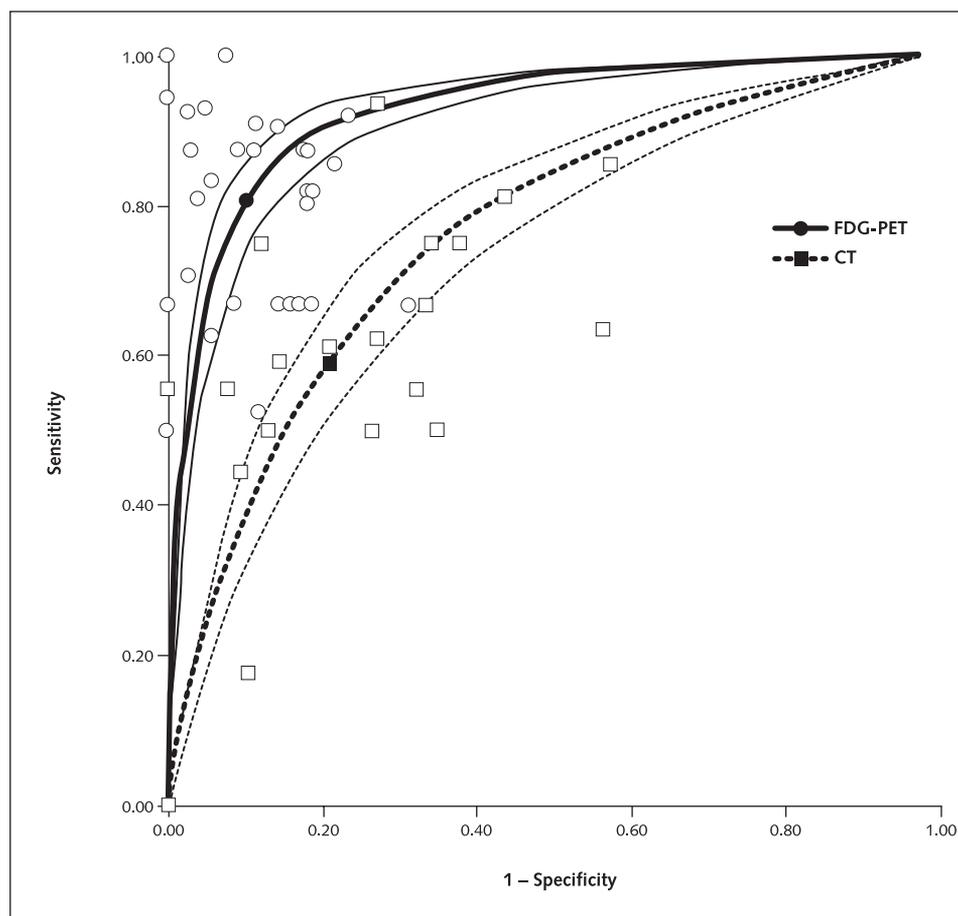
Error bars represent 95% CIs. Three studies reported results by using both the patient and lymph nodes or lymph node stations as the units of analysis; these 3 studies are listed twice (70, 74, 75).

Figure 3. Individual study estimates of sensitivity and 1 – specificity of positron emission tomography with 18-fluorodeoxyglucose for identifying mediastinal metastasis.



Error bars represent 95% CIs. Five studies reported results by using both the patient and lymph nodes or lymph node stations as the units of analysis; these 5 studies are listed twice (70, 73–75, 96).

Figure 4. Summary receiver-operating characteristic curves and 95% CIs for mediastinal staging with positron emission tomography with 18-fluorodeoxyglucose (FDG-PET) and computed tomography (CT).



Individual study estimates of sensitivity and 1 – specificity are shown for FDG-PET (*open circles*) and CT (*open squares*). The approximate points on the curves where FDG-PET and CT operate in current practice are indicated (*solid circle* and *solid square*, respectively).

ative, respectively (Figure 6). Post-test probabilities based on unconditional estimates of sensitivity and specificity would have been 69% and 6%, respectively (Appendix Figure, available at www.annals.org).

DISCUSSION

In this meta-analysis, we found that FDG-PET is more accurate than CT for mediastinal staging in patients with non-small-cell lung cancer. We estimate that, in current practice, the sensitivity and specificity of CT are approximately 59% and 79%, respectively. In comparison, the sensitivity and specificity of FDG-PET are approximately 81% and 90%, respectively. While FDG-PET is an important advance in noninvasive staging, it is not perfect. False-positive FDG-PET results have been reported in patients with postobstructive pneumonia and granulomatous disease. In addition, the spatial resolution of the current generation of PET scanners is approximately 7 mm. While it is possible for PET imaging to detect smaller lesions that are intensely hypermetabolic, the high false-negative rate (approximately 25%) in patients without lymph node en-

largement on CT confirms that nodal size has an important effect on diagnostic accuracy. Another consequence of the limited spatial resolution of PET imaging is that the distinction between hilar (N1) and mediastinal (N2) lymph nodes, which has important implications for both treatment selection and prognosis, is sometimes difficult to make. Advances in PET technology, including the refinement of hybrid PET–CT scanners, may help to overcome some of these limitations in the future (102).

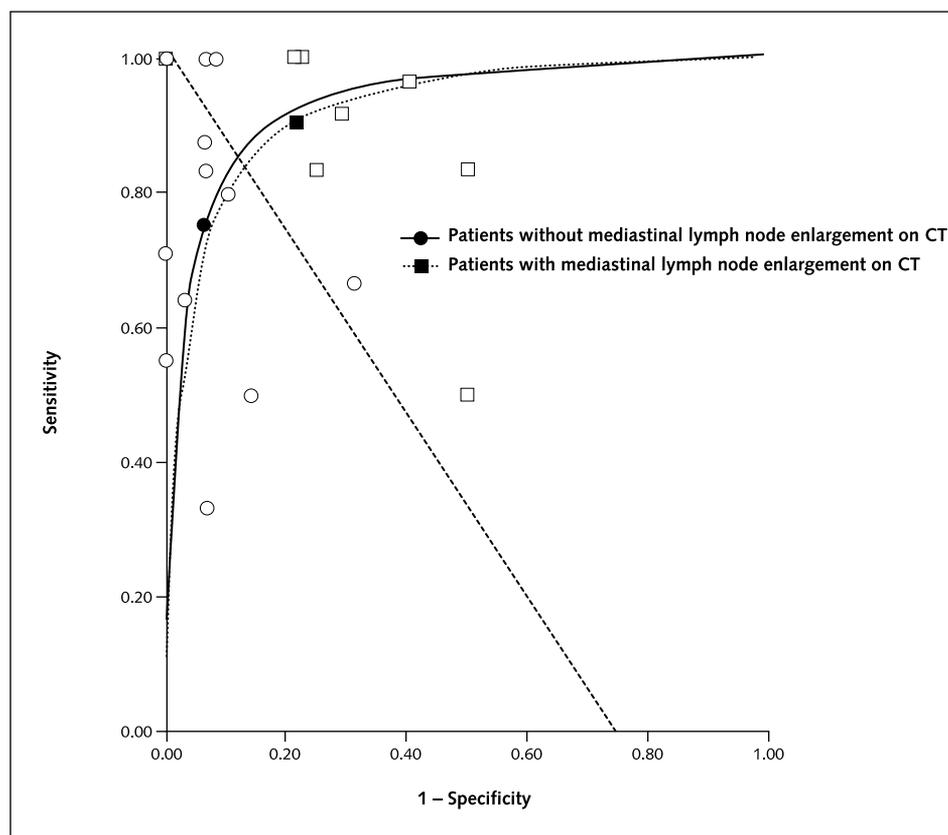
Although we identified considerable variability in study methods, only 3 factors affected PET performance. First, we found that diagnostic accuracy was better when studies reported that FDG-PET imaging was performed on patients in the fasting state. Second, we found that accuracy was better in studies published before 1999. Nevertheless, our estimates of sensitivity and specificity are similar to estimates previously reported in a meta-analysis by Dwamena and colleagues (103) who examined studies published before January 1998. As we stated previously in a meta-analysis of studies of FDG-PET for pulmonary nodule diagnosis, the lower accuracy observed in more re-

cent studies could be due to enrollment of less highly selected patient samples or dissemination of PET technology to centers with less experience (13). Finally, we found that studies in which lymph nodes or lymph node stations were used as the unit of analysis tended to overestimate diagnostic accuracy, especially specificity. In these studies, observations are not statistically independent, that is, if a given patient has 1 positive lymph node, that patient is more likely to have other positive lymph nodes. In addition, it is important to note that the clinically relevant unit of analysis is the patient, not the lymph node. In general, treatment decisions depend on the presence or absence of lymph node involvement rather than the number of involved nodes. Because of these considerations, we recommend that future studies of tests for mediastinal staging report results by using the individual patient as the unit of the analysis.

To our knowledge, this is the first study to systematically evaluate the conditional test performance of FDG-PET and CT, although others have briefly addressed this

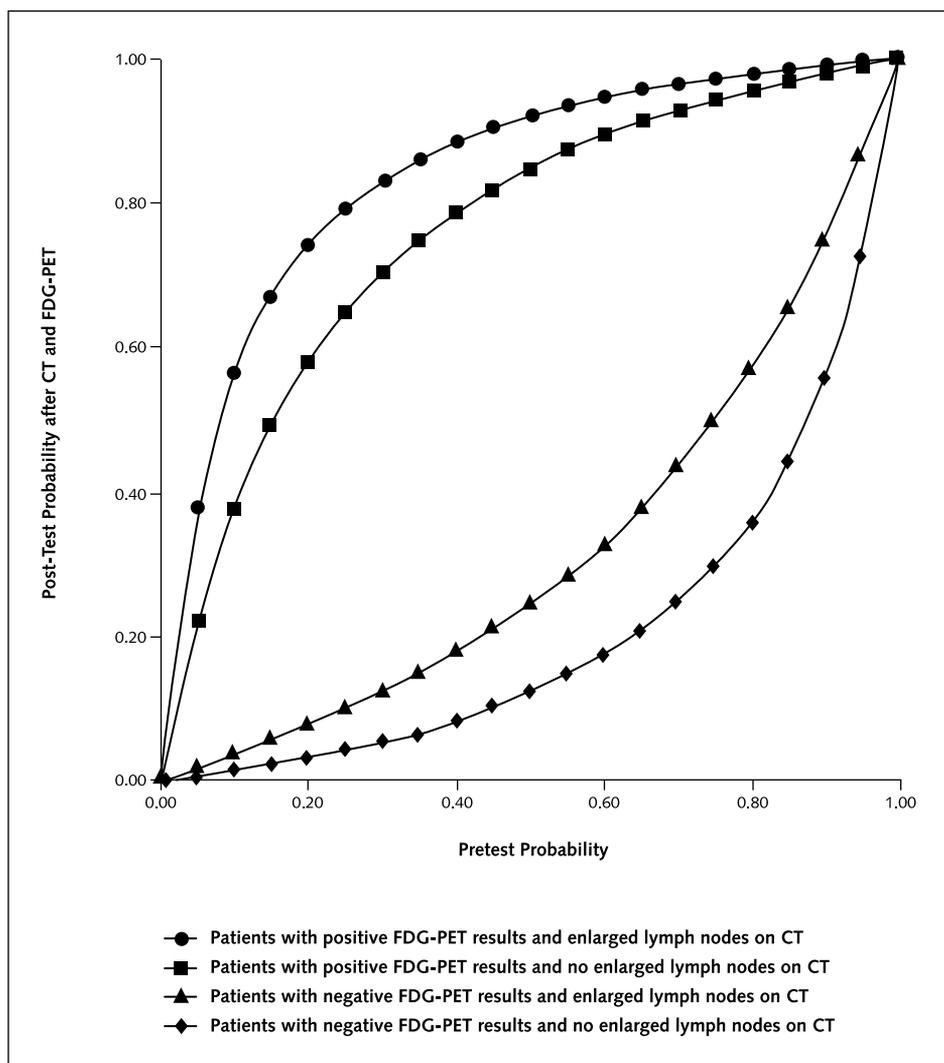
issue using less rigorous methods (104, 105). Summary ROC curves for FDG-PET in patients with and without enlarged mediastinal nodes were almost identical, suggesting that PET and CT results might be independent (Figure 5). However, FDG-PET appeared to operate at different points on the curves, depending on the presence or absence of lymph node enlargement. In patients with enlarged lymph nodes, FDG-PET operated near a point where sensitivity and specificity were 91% and 78%, respectively. In contrast, in patients without lymph node enlargement, FDG-PET operated near a point where sensitivity and specificity were 75% and 93%, respectively. In patients with enlarged lymph nodes, FDG-PET is more likely to reveal both true-positive findings that are due to metastasis and false-positive findings that are due to infection or inflammation, respectively. The increase in true-positive findings leads to higher estimated sensitivity, and the increase in false-positive findings results in lower estimated specificity. Conversely, FDG-PET is more likely to yield both true-negative and false-negative findings in pa-

Figure 5. Summary receiver-operating characteristic curves for mediastinal staging with positron emission tomography with 18-fluorodeoxyglucose in patients with and without mediastinal lymph node enlargement on computed tomography (CT).



Individual study estimates of sensitivity and 1 – specificity are shown for positron emission tomography with 18-fluorodeoxyglucose in patients with enlarged lymph nodes (*open squares*) and without enlarged lymph nodes (*open circles*). The 2 receiver-operating characteristic curves are nearly identical. However, in patients with enlarged lymph nodes on CT, studies tend to cluster on a portion of the curve at which sensitivity is favored over specificity. In patients without lymph node enlargement, studies tend to cluster on a portion of the curve at which specificity is favored over sensitivity. The approximate points on the curves where positron emission tomography with 18-fluorodeoxyglucose operates in current practice in patients with and without lymph node enlargement are indicated (*solid square* and *solid circle*, respectively). The discriminant function that separates the 2 groups of patients is shown (*dashed line*) ($P = 0.002$ by nonparametric permutation test).

Figure 6. Post-test probabilities of mediastinal metastasis after computed tomography (CT) and positron emission tomography with 18-fluorodeoxyglucose (FDG-PET).



Post-test probabilities are shown as a function of pretest probability in patients with positive FDG-PET results and enlarged lymph nodes on CT (circles), patients with positive FDG-PET results and no enlarged lymph nodes on CT (squares), patients with negative FDG-PET results and enlarged lymph nodes on CT (triangles), and patients with negative FDG-PET results and no enlarged lymph nodes on CT (diamonds).

tients without lymph node enlargement because of the test's inherent limitations in its ability to detect small hypermetabolic lesions of any origin. The increase in true-negative findings leads to higher estimated specificity, and the increase in false-negative findings results in decreased sensitivity.

Because the negative consequences of false-positive staging evaluations are so serious (missed opportunities for surgical cure), we believe that a positive FDG-PET result does not “rule in” disease with enough certainty unless pretest probability is very high (>85% to 90%) and that confirmatory mediastinal biopsy should be performed before excluding surgery as a treatment option. In fact, when CT shows lymph nodes that are accessible by transbronchial needle aspiration biopsy or endoscopic ultrasound-guided biopsy, these tests should be considered before PET

imaging, especially if bronchoscopy is already planned for primary tumor diagnosis.

When FDG-PET results are negative, the decision to perform biopsy or surgery should be guided by the pretest probability of mediastinal metastasis, the presence or absence of lymph node enlargement, the risk for surgical complications, and patient preferences. While our results can help to inform clinical decision making, additional studies are needed to determine the threshold post-test probability below which surgery without previous mediastinal biopsy can be performed.

This study has several limitations. First, we did not attempt to identify all published studies of CT for mediastinal staging, only studies of FDG-PET that also reported results for CT. Thus, our estimates of diagnostic accuracy for CT may be biased. However, our estimates are

similar to those reported in recent studies of CT for lymph node staging (1–3), as well as those reported in the meta-analysis by Dwamena and colleagues (103). We present the additional finding that CT provides little diagnostic information once the results of FDG-PET are known, confirming an observation made by Pieterman and colleagues (90) in a study conducted at a single center in the Netherlands. Second, it is possible that we did not identify all studies of FDG-PET for mediastinal staging, particularly unpublished studies. However, we conducted an exhaustive search for studies published in any language, and an inverted funnel plot did not support the hypothesis that several small “negative” studies were not identified. Finally, our estimates of diagnostic accuracy do not capture all of the potential benefits of staging with whole-body PET, which identifies unsuspected distant metastasis in approximately 10% of patients with otherwise resectable non-small-cell lung cancer (81, 90).

We conclude that FDG-PET is more accurate than CT for mediastinal staging in patients with potentially resectable non-small-cell lung cancer and that the sensitivity and specificity of FDG-PET depend on the presence or absence of enlarged mediastinal lymph nodes on CT. Positive findings on PET imaging should be confirmed by biopsy before curative surgery is excluded as a treatment option. Negative findings on FDG-PET should be interpreted in light of the patient’s pretest probability of mediastinal metastasis and whether CT reveals enlarged mediastinal nodes.

From Veterans Affairs Palo Alto Health Care System, Palo Alto, and Stanford University School of Medicine, Stanford, California.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Acknowledgments: The authors thank Jean-Dominique Delcroix, PhD, Eran Geller, MD, MS, Eric Hsiao, MD, and Annette Langer-Gould, MD, for reviewing non-English-language studies; James Fletcher, MD, Ann Leung, MD, and George Segall, MD, for helping to develop criteria for the technical quality of CT and FDG-PET; Christopher Stave, MLS, and Elaine Alligood, MLS, for assisting with literature searches; and Dena Bravata MD, MS, Lincoln Moses, PhD, and Trevor Hastie, PhD, for providing statistical advice.

Grant Support: Drs. Gould and Owens received Research Career Development Awards from the Veterans Affairs Health Services Research and Development Service. This study was also supported by Veterans Affairs Cooperative Study 27, “18-F-Fluorodeoxyglucose (FDG) Positron Emission Tomography (PET) Imaging for Management of Patients with Solitary Pulmonary Nodules.”

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Michael K. Gould, MD, MS, Pulmonary Section (111P), Veterans Affairs Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304; e-mail, gould@stanford.edu.

Current author addresses, author contributions, and additional Appendix Tables are available at www.annals.org.

References

1. Webb WR, Gatsonis C, Zerhouni EA, Heelan RT, Glazer GM, Francis IR, et al. CT and MR imaging in staging non-small cell bronchogenic carcinoma: report of the Radiologic Diagnostic Oncology Group. *Radiology*. 1991;178:705-13. [PMID: 1847239]
2. McLoud TC, Bourgouin PM, Greenberg RW, Kosiuk JP, Templeton PA, Shepard JA, et al. Bronchogenic carcinoma: analysis of staging in the mediastinum with CT by correlative lymph node mapping and sampling. *Radiology*. 1992;182:319-23. [PMID: 1732943]
3. Primack SL, Lee KS, Logan PM, Miller RR, Müller NL. Bronchogenic carcinoma: utility of CT in the evaluation of patients with suspected lesions. *Radiology*. 1994;193:795-800. [PMID: 7972827]
4. Cook GJ, Maisey MN. The current status of clinical PET imaging. *Clin Radiol*. 1996;51:603-13. [PMID: 8810687]
5. Lowe VJ, Naunheim KS. Current role of positron emission tomography in thoracic oncology. *Thorax*. 1998;53:703-12. [PMID: 9828860]
6. Patz EF Jr, Goodman PC. Positron emission tomography imaging of the thorax. *Radiol Clin North Am*. 1994;32:811-23. [PMID: 7980769]
7. Owens DK, Sox HC. Medical decision making: probabilistic medical reasoning. In: Shortliffe EH, Perreault LE, Wiederhold G, Fagan LM, eds. *Medical Informatics: Computer Applications in Health Care and Biomedicine*. New York: Springer-Verlag; 2000.
8. Gould MK, Owens DK. Positron emission tomography for mediastinal staging in non-small cell lung cancer: a meta-analysis [Abstract]. *Am J Respir Crit Care Med*. 1998;157:A256.
9. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127:380-7. [PMID: 9273830]
10. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med*. 1997;127:531-7. [PMID: 9313021]
11. Kent DL, Haynor DR, Larson EB, Deyo RA. Diagnosis of lumbar spinal stenosis in adults: a metaanalysis of the accuracy of CT, MR, and myelography. *AJR Am J Roentgenol*. 1992;158:1135-44. [PMID: 1533084]
12. 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;12:1293-316. [PMID: 8210827]
13. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA*. 2001;285:914-24. [PMID: 11180735]
14. Fleiss J. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: J Wiley; 1981.
15. Owens DK, Holodniy M, McDonald TW, Scott J, Sonnad S. A meta-analytic evaluation of the polymerase chain reaction for the diagnosis of HIV infection in infants. *JAMA*. 1996;275:1342-8. Erratum in: *JAMA* 1996;276:1302. [PMID: 8614121]
16. Owens DK, Holodniy M, Garber AM, Scott J, Sonnad S, Moses L, et al. Polymerase chain reaction for the diagnosis of HIV infection in adults. A meta-analysis with recommendations for clinical practice and study design. *Ann Intern Med*. 1996;124:803-15. [PMID: 8610949]
17. Robins J, Breslow N, Greenland S. Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*. 1986;42:311-23. [PMID: 3741973]
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-88. [PMID: 3802833]
19. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med*. 1997;127:820-6. [PMID: 9382404]
20. Dillon WR, Goldstein M. *Multivariate Analysis. Methods and Applications*. New York: J Wiley; 1984.
21. Fisher L, Van Belle G. *Biostatistics. A Methodology for the Health Sciences*. New York: J Wiley; 1993.
22. Begg C. Publication bias. In: Cooper H, Hedges L, eds. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994:399-409.
23. Matsuzawa T, Fujiwara T, Abe Y, Itoh M, Fukuda H, Yamaguchi K, et al.

- [Positron emission tomography of lung cancer using 18F-2-fluoro-2-deoxy-D-glucose and L-[11C-methyl]-methionine]. *Kokyu To Junkan*. 1987;35:15-20. [PMID: 3031787]
24. Gupta NC, Frank AR, Dewan NA, Redepinning LS, Rothberg ML, Mailliard JA, et al. Solitary pulmonary nodules: detection of malignancy with PET with 2-[F-18]-fluoro-2-deoxy-D-glucose. *Radiology*. 1992;184:441-4. [PMID: 1620844]
 25. Lewis P, Griffin S, Marsden P, Gee T, Nunan T, Malsey M, et al. Whole-body 18F-fluorodeoxyglucose positron emission tomography in preoperative evaluation of lung cancer. *Lancet*. 1994;344:1265-6. [PMID: 7967988]
 26. Scott WJ, Schwabe JL, Gupta NC, Dewan NA, Reeb SD, Sugimoto JT. Positron emission tomography of lung tumors and mediastinal lymph nodes using [18F]fluorodeoxyglucose. The Members of the PET-Lung Tumor Study Group. *Ann Thorac Surg*. 1994;58:698-703. [PMID: 7944691]
 27. Scott WJ, Gobar LS, Hauser LG, Sunderland JJ, Dewan NA, Sugimoto JT. Detection of scalene lymph node metastases from lung cancer. *Positron emission tomography*. *Chest*. 1995;107:1174-6. [PMID: 7705136]
 28. Gupta NC, Maloof J, Gunel E. Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET. *J Nucl Med*. 1996;37:943-8. [PMID: 8683316]
 29. Higashi K, Nishikawa T, Seki H, Oguchi M, Nambu Y, Ueda Y, et al. Comparison of fluorine-18-FDG PET and thallium-201 SPECT in evaluation of lung cancer. *J Nucl Med*. 1998;39:9-15. [PMID: 9443730]
 30. Lonneux M, Delval D, Bausart R, Moens R, Willockx R, Van Mael P, et al. Can dual-headed 18F-FDG SPET imaging reliably supersede PET in clinical oncology? A comparative study in lung and gastrointestinal tract cancer. *Nucl Med Commun*. 1998;19:1047-54. [PMID: 9861621]
 31. Nettelbladt OS, Sundin AE, Valind SO, Gustafsson GR, Lamberg K, Långström B, et al. Combined fluorine-18-FDG and carbon-11-methionine PET for diagnosis of tumors in lung and mediastinum. *J Nucl Med*. 1998;39:640-7. [PMID: 9544671]
 32. Bury T, Dowlati A, Paulus P, Hustinx R, Radermecker M, Rigo P. Staging of non-small-cell lung cancer by whole-body fluorine-18 deoxyglucose positron emission tomography. *Eur J Nucl Med*. 1996;23:204-6. [PMID: 8925857]
 33. Bury T, Dowlati A, Paulus P, Corhay JL, Hustinx R, Ghaye B, et al. Whole-body 18FDG positron emission tomography in the staging of non-small cell lung cancer. *Eur Respir J*. 1997;10:2529-34. [PMID: 9426090]
 34. Ferlin G, Rubello D, Chierichetti F, Zanco P, Bergamin R, Trento P, et al. The role of fluorine-18-deoxyglucose (FDG) positron emission tomography (PET) whole body scan (WBS) in the staging and follow-up of cancer patients: our first experience. *Tumori*. 1997;83:679-84. [PMID: 9267488]
 35. Wang H, Maurea S, Mainolfi C, Fiore F, Gravina A, Panico MR, et al. Tc-99m MIBI scintigraphy in patients with lung cancer. Comparison with CT and fluorine-18 FDG PET imaging. *Clin Nucl Med*. 1997;22:243-9. [PMID: 9099482]
 36. Kutlu CA, Pastorino U, Maisey M, Goldstraw P. Selective use of PET scan in the preoperative staging of NSCLC. *Lung Cancer*. 1998;21:177-84. [PMID: 9857995]
 37. Shreve PD, Steventon RS, Deters EC, Kison PV, Gross MD, Wahl RL. Oncologic diagnosis with 2-[fluorine-18]fluoro-2-deoxy-D-glucose imaging: dual-head coincidence gamma camera versus positron emission tomographic scanner. *Radiology*. 1998;207:431-7. [PMID: 9577492]
 38. Graeber GM, Gupta NC, Murray GF. Positron emission tomographic imaging with fluorodeoxyglucose is efficacious in evaluating malignant pulmonary disease. *J Thorac Cardiovasc Surg*. 1999;117:719-27. [PMID: 10096967]
 39. Gupta NC, Bishop HA, Rogers JS, Tamim WZ, Reesman SD. Treatment outcome of lung cancer patients as optimized by preoperative whole-body positron emission tomography fluorodeoxyglucose imaging. *Clinical Lung Cancer*. 2000;2:146-50.
 40. Hara T, Inagaki K, Kosaka N, Morita T. Sensitive detection of mediastinal lymph node metastasis of lung cancer with 11C-choline PET. *J Nucl Med*. 2000;41:1507-13. [PMID: 10994730]
 41. Weng E, Tran L, Rege S, Safa A, Sadeghi A, Juillard G, et al. Accuracy and clinical impact of mediastinal lymph node staging with FDG-PET imaging in potentially resectable lung cancer. *Am J Clin Oncol*. 2000;23:47-52. [PMID: 10683077]
 42. Imdahl A, Jenkner S, Brink I, Nitzsche E, Stoelben E, Moser E, et al. Validation of FDG positron emission tomography for differentiation of unknown pulmonary lesions. *Eur J Cardiothorac Surg*. 2001;20:324-9. [PMID: 11463551]
 43. Kalf V, Hicks RJ, MacManus MP, Binns DS, McKenzie AF, Ware RE, et al. Clinical impact of (18F) fluorodeoxyglucose positron emission tomography in patients with non-small-cell lung cancer: a prospective study. *J Clin Oncol*. 2001;19:1111-8. [PMID: 11134203]
 44. Rubinstein R, Breuer R, Chisin R. [Contribution of PET using FDG in the diagnosis of lung cancer—first results]. *Harefuah*. 2001;140:100-3, 191. [PMID: 11242909]
 45. Albes JM, Dohmen BM, Schott U, Schülen E, Wehrmann M, Ziemer G. Value of positron emission tomography for lung cancer staging. *Eur J Surg Oncol*. 2002;28:55-62. [PMID: 11869015]
 46. Ryu JS, Choi NC, Fischman AJ, Lynch TJ, Mathisen DJ. FDG-PET in staging and restaging non-small cell lung cancer after neoadjuvant chemoradiotherapy: correlation with histopathology. *Lung Cancer*. 2002;35:179-87. [PMID: 11804691]
 47. Bury T, Corhay JL, Paulus P, Weber T, D'Harcour JB, Limet R, et al. [Positron emission tomography in the evaluation of intrathoracic lymphatic extension of non-small cell bronchial cancer. A preliminary study of 30 patients]. *Rev Mal Respir*. 1996;13:281-6. [PMID: 8765921]
 48. Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Verschakelen JA, Nackaerts KL, et al. Mediastinal lymph node staging with FDG-PET scan in patients with potentially operable non-small cell lung cancer: a prospective analysis of 50 cases. *Leuven Lung Cancer Group*. *Chest*. 1997;112:1480-6. [PMID: 9404742]
 49. Tatsumi M, Yutani K, Watanabe Y, Miyoshi S, Tomiyama N, Johkoh T, et al. Feasibility of fluorodeoxyglucose dual-head gamma camera coincidence imaging in the evaluation of lung cancer: comparison with FDG PET. *J Nucl Med*. 1999;40:566-73. [PMID: 10210214]
 50. Gupta NC, Graeber GM, Rogers JS 2nd, Bishop HA. Comparative efficacy of positron emission tomography with FDG and computed tomographic scanning in preoperative staging of non-small cell lung cancer. *Ann Surg*. 1999;229:286-91. [PMID: 10024112]
 51. Gupta NC, Graeber GM, Bishop HA. Comparative efficacy of positron emission tomography with fluorodeoxyglucose in evaluation of small (<1 cm), intermediate (1 to 3 cm), and large (>3 cm) lymph node lesions. *Chest*. 2000;117:773-8. [PMID: 10713005]
 52. Trampert L, Holle LH, Berberich R, Alexander C, Ukena D, Ruth T, et al. [18FDG in the primary staging of lung tumors. Results with a gamma camera and a 511 keV collimator]. *Nuklearmedizin*. 1995;34:79-86. [PMID: 7630746]
 53. Stokkel MP, Bakker PF, Heine R, Schlösser NJ, Lammers JW, Van Rijk PP. Staging of lymph nodes with FDG dual-headed PET in patients with non-small-cell lung cancer. *Nucl Med Commun*. 1999;20:1001-7. [PMID: 10572909]
 54. Grahek D, Tofighi M, Montravers F, Kerrou K, Tamgac F, Breau JL, et al. [18FDG CDET in staging of lung cancer]. *Medecine Nucleaire*. 2000;24:99-106.
 55. Roman MR, Rossleigh MA, Angelides S, Walker BM, Dixon J. Staging and managing lung tumors using F-18 FDG coincidence detection. *Clin Nucl Med*. 2001;26:383-8. [PMID: 11317014]
 56. Torres García AJ, Carreras Delgado JL. [Positron-emission tomography in the evaluation of lung cancer] [Editorial]. *Arch Bronconeumol*. 1997;33:553-5. [PMID: 9508470]
 57. Sugio K, Sasaki M, Yamazaki K, Kase S, Sugimachi K. [Positron emission tomography for detection of lymph node metastases in lung cancer]. *Nippon Geka Gakkai Zasshi*. 1999;100:718-23. [PMID: 10629837]
 58. Adams S, Nickel E, Hör G. [Problems of differential diagnosis in the diagnosis of mediastinal and pulmonary metastases with F-a8-FDG PET]. *Nuklearmedizin*. 2000;39:N83-4. [PMID: 10984898]
 59. Pieterman RM, Que TH, Elsinga PH, Pruijm J, van Putten JW, Willemsen AT, et al. Comparison of (11)C-choline and (18)F-FDG PET in primary diagnosis and staging of patients with thoracic cancer. *J Nucl Med*. 2002;43:167-72. [PMID: 11850480]
 60. Eschmann SM, Friedel G, Paulsen F, Budach W, Harer-Mouline C, Dohmen BM, et al. FDG PET for staging of advanced non-small cell lung cancer prior to neoadjuvant radio-chemotherapy. *Eur J Nucl Med Mol Imaging*.

- 2002;29:804-8. [PMID: 12029555]
61. Fritscher-Ravens A, Bohuslavizki KH, Brandt L, Bobrowski C, Lund C, Knöfel WT, et al. Mediastinal lymph node involvement in potentially resectable lung cancer: comparison of CT, positron emission tomography, and endoscopic ultrasonography with and without fine-needle aspiration. *Chest*. 2003;123:442-51. [PMID: 12576364]
62. Kiernan PD, Sheridan MJ, Lamberti J, Diccio B, Wigton R, Hetrick V, et al. Mediastinal staging of non-small cell lung carcinoma using computed and positron-emission tomography. *South Med J*. 2002;95:1168-72. [PMID: 12425503]
63. Wahl RL, Quint LE, Greenough RL, Meyer CR, White RI, Orringer MB. Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation. *Radiology*. 1994;191:371-7. [PMID: 8153308]
64. Chin R Jr, Ward R, Keyes JW, Choplin RH, Reed JC, Wallenhaupt S, et al. Mediastinal staging of non-small-cell lung cancer with positron emission tomography. *Am J Respir Crit Care Med*. 1995;152:2090-6. [PMID: 8520780]
65. Patz EF Jr, Lowe VJ, Goodman PC, Herndon J. Thoracic nodal staging with PET imaging with 18FDG in patients with bronchogenic carcinoma. *Chest*. 1995;108:1617-21. [PMID: 7497771]
66. Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB, et al. Staging non-small cell lung cancer by whole-body positron emission tomographic imaging. *Ann Thorac Surg*. 1995;60:1573-81; discussion 1581-2. [PMID: 8787446]
67. Bury T, Paulus P, Dowlati A, Corhay JL, Weber T, Ghaye B, et al. Staging of the mediastinum: value of positron emission tomography imaging in non-small cell lung cancer. *Eur Respir J*. 1996;9:2560-4. [PMID: 8980969]
68. Sasaki M, Ichiya Y, Kuwabara Y, Akashi Y, Yoshida T, Fukumura T, et al. The usefulness of FDG positron emission tomography for the detection of mediastinal lymph node metastases in patients with non-small cell lung cancer: a comparative study with X-ray computed tomography. *Eur J Nucl Med*. 1996;23:741-7. [PMID: 8662111]
69. Sazon DA, Santiago SM, Soo Hoo GW, Khonsary A, Brown C, Mandelkern M, et al. Fluorodeoxyglucose-positron emission tomography in the detection and staging of lung cancer. *Am J Respir Crit Care Med*. 1996;153:417-21. [PMID: 8542152]
70. Scott WJ, Gobar LS, Terry JD, Dewan NA, Sunderland JJ. Mediastinal lymph node staging of non-small-cell lung cancer: a prospective comparison of computed tomography and positron emission tomography. *J Thorac Cardiovasc Surg*. 1996;111:642-8. [PMID: 8601980]
71. Guhlmann A, Storck M, Kotzerke J, Moog F, Sunder-Plassmann L, Reske SN. Lymph node staging in non-small cell lung cancer: evaluation by [18F]FDG positron emission tomography (PET). *Thorax*. 1997;52:438-41. [PMID: 9176535]
72. Hagberg RC, Segall GM, Stark P, Burdon TA, Pompili MF. Characterization of pulmonary nodules and mediastinal staging of bronchogenic carcinoma with F-18 fluorodeoxyglucose positron emission tomography. *Eur J Cardiothorac Surg*. 1997;12:92-7. [PMID: 9262087]
73. Steinert HC, Hauser M, Allemann F, Engel H, Berthold T, von Schulthess GK, et al. Non-small cell lung cancer: nodal staging with FDG PET versus CT with correlative lymph node mapping and sampling. *Radiology*. 1997;202:441-6. [PMID: 9015071]
74. Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Bogaert J, Maes A, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol*. 1998;16:2142-9. [PMID: 9626214]
75. Vansteenkiste JF, Stroobants SG, Dupont PJ, De Leyn PR, De Wever WF, Verbeke EK, et al. FDG-PET scan in potentially operable non-small cell lung cancer: do anatomometabolic PET-CT fusion images improve the localisation of regional lymph node metastases? The Leuven Lung Cancer Group. *Eur J Nucl Med*. 1998;25:1495-501. [PMID: 9799345]
76. Albes JM, Lietzenmayer R, Schott U, Schülen E, Wehrmann M, Ziemer G. Improvement of non-small-cell lung cancer staging by means of positron emission tomography. *Thorac Cardiovasc Surg*. 1999;47:42-7. [PMID: 10218620]
77. Berlangieri SU, Scott AM, Knight SR, Fitt GJ, Hennessy OF, Tochon-Danguy HJ, et al. F-18 fluorodeoxyglucose positron emission tomography in the non-invasive staging of non-small cell lung cancer. *Eur J Cardiothorac Surg*. 1999;16 Suppl 1:S25-30. [PMID: 10536941]
78. Higashi K, Oguchi M, Tamamura H, Wang XM, Yamamoto I, Ueda Y, et al. [Comparison of TI SPECT and FDG PET in the diagnosis of lymph node metastases from lung cancer]. *Japanese Journal of Clinical Radiology*. 1999;44:191-7.
79. Kernstine KH, Stanford W, Mullan BF, Rossi NP, Thompson BH, Bushnell DL, et al. PET, CT, and MRI with Combidex for mediastinal staging in non-small cell lung carcinoma. *Ann Thorac Surg*. 1999;68:1022-8. [PMID: 10510001]
80. Magnani P, Carretta A, Rizzo G, Fazio F, Vanzulli A, Lucignani G, et al. FDG/PET and spiral CT image fusion for mediastinal lymph node assessment of non-small cell lung cancer patients. *J Cardiovasc Surg (Torino)*. 1999;40:741-8. [PMID: 10597015]
81. Marom EM, McAdams HP, Erasmus JJ, Goodman PC, Culhane DK, Coleman RE, et al. Staging non-small cell lung cancer with whole-body PET. *Radiology*. 1999;212:803-9. [PMID: 10478250]
82. Richter JA, Torre W, Gamez C, Aramendia JM, Crespo A, Nicolas A, Brugarolas A. [Value of Pet-18FDG in lung cancer] *Med Clin (Barc)*. 1999;113:567-71. [PMID: 10605681]
83. Saunders CA, Dussek JE, O'Doherty MJ, Maisey MN. Evaluation of fluorine-18-fluorodeoxyglucose whole body positron emission tomography imaging in the staging of lung cancer. *Ann Thorac Surg*. 1999;67:790-7. [PMID: 10215230]
84. Weber WA, Nerve J, Sklarek J, Ziegler SI, Bartenstein P, King B, et al. Imaging of lung cancer with fluorine-18 fluorodeoxyglucose: comparison of a dual-head gamma camera in coincidence mode with a full-ring positron emission tomography system. *Eur J Nucl Med*. 1999;26:388-95. [PMID: 10199945]
85. Demura Y, Mizuno S, Wakabayashi M, Totani Y, Okamura S, Ameshima S, et al. [Evaluation of fluorine-18-fluorodeoxyglucose whole body positron emission tomography imaging in the clinical diagnosis of lung cancer]. *Nihon Koryuiki Gakkai Zasshi*. 2000;38:676-81. [PMID: 11109804]
86. Farrell MA, McAdams HP, Herndon JE, Patz EF Jr. Non-small cell lung cancer: FDG PET for nodal staging in patients with stage I disease. *Radiology*. 2000;215:886-90. [PMID: 10831716]
87. Kitase M, Hara M, Patz Jr EF, Katoh K, Satoh Y, Satake M, et al. [FDG-PET in patient with clinical T1N0 lung cancer; Determination of nodal status]. *Japanese Journal of Clinical Radiology*. 2000;45:209-14.
88. Kubota K, Imran MB, Ono S, Akaizawa T, Gotoh R, Fukuda H, et al. [Diagnostic value of whole-body positron emission tomography using fluorine-18 fluorodeoxyglucose for lung and other cancer]. *Japanese Journal of Clinical Radiology*. 2000;45:199-208.
89. Liewald F, Grosse S, Storck M, Guhlmann A, Halter G, Reske S, et al. How useful is positron emission tomography for lymphnode staging in non-small-cell lung cancer? *Thorac Cardiovasc Surg*. 2000;48:93-6. [PMID: 11028710]
90. Pieterman RM, van Putten JW, Meuzelaar JJ, Mooyaart EL, Vaalburg W, Koëter GH, et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med*. 2000;343:254-61. [PMID: 10911007]
91. Roberts PF, Follette DM, von Haag D, Park JA, Valk PE, Pounds TR, et al. Factors associated with false-positive staging of lung cancer by positron emission tomography. *Ann Thorac Surg*. 2000;70:1154-9; discussion 1159-60. [PMID: 11081861]
92. Tatsumi M, Yutani K, Nishimura T. Evaluation of lung cancer by 99mTc-tetrofosmin SPECT: comparison with [18F]FDG-PET. *J Comput Assist Tomogr*. 2000;24:574-80. [PMID: 10966189]
93. Changlai SP, Tsai SC, Chou MC, Ho YJ, Kao CH. Whole body 18F-2-deoxyglucose positron emission tomography to restage non-small cell lung cancer. *Oncol Rep*. 2001;8:337-9. [PMID: 11182051]
94. Dunagan D, Chin R Jr, McCain T, Case L, Harkness B, Oaks T, et al. Staging by positron emission tomography predicts survival in patients with non-small cell lung cancer. *Chest*. 2001;119:333-9. [PMID: 11171706]
95. Guan Y, He S, Dong J. [Value of 18F-fluorodeoxyglucose positron emission tomography imaging in staging of non-small cell lung cancer]. *Zhonghua Yi Xue Za Zhi*. 2001;81:1180-3. [PMID: 11769705]
96. Gupta NC, Tamim WJ, Graeber GG, Bishop HA, Hobbs GR. Mediastinal lymph node sampling following positron emission tomography with fluorodeoxyglucose imaging in lung cancer staging. *Chest*. 2001;120:521-7. [PMID: 11769705]

11502653]

97. Poncelet AJ, Lonneux M, Coche E, Weynand B, Noirhomme P. PET-FDG scan enhances but does not replace preoperative surgical staging in non-small cell lung carcinoma. *Eur J Cardiothorac Surg*. 2001;20:468-74; discussion 474-5. [PMID: 11509265]

98. Kernstine KH, McLaughlin KA, Menda Y, Rossi NP, Kahn DJ, Bushnell DL, et al. Can FDG-PET reduce the need for mediastinoscopy in potentially resectable nonsmall cell lung cancer? *Ann Thorac Surg*. 2002;73:394-401; discussion 401-2. [PMID: 11845848]

99. Vesselle H, Pugsley JM, Vallières E, Wood DE. The impact of fluorodeoxyglucose F 18 positron-emission tomography on the surgical staging of non-small cell lung cancer. *J Thorac Cardiovasc Surg*. 2002;124:511-9. [PMID: 12202868]

100. von Haag DW, Follette DM, Roberts PF, Shelton D, Segel LD, Taylor TM. Advantages of positron emission tomography over computed tomography in mediastinal staging of non-small cell lung cancer. *J Surg Res*. 2002;103:160-4. [PMID: 11922730]

101. Graeter TP, Hellwig D, Hoffmann K, Ukena D, Kirsch CM, Schäfers HJ. Mediastinal lymph node staging in suspected lung cancer: comparison of positron emission tomography with F-18-fluorodeoxyglucose and mediastinoscopy. *Ann Thorac Surg*. 2003;75:231-5; discussion 235-6. [PMID: 12537221]

102. Lardinois D, Weder W, Hany TF, Kamel EM, Korom S, Seifert B, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med*. 2003;348:2500-7. [PMID:

12815135]

103. Dwamena BA, Sonnad SS, Angobaldo JO, Wahl RL. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s—meta-analytic comparison of PET and CT. *Radiology*. 1999;213:530-6. [PMID: 10551237]

104. Scott WJ, Shepherd J, Gambhir SS. Cost-effectiveness of FDG-PET for staging non-small cell lung cancer: a decision analysis. *Ann Thorac Surg*. 1998;66:1876-83; discussion 1883-5. [PMID: 9930463]

105. Dietlein M, Weber K, Gandjour A, Moka D, Theissen P, Lauterbach KW, et al. Cost-effectiveness of FDG-PET for the management of potentially operable non-small cell lung cancer: priority for a PET-based strategy after nodal-negative CT results. *Eur J Nucl Med*. 2000;27:1598-609. [PMID: 11105815]

106. Schelbert HR, Hoh CK, Royal HD, Brown M, Dahlbom MN, Dehdashti F, et al. Procedure guideline for tumor imaging using fluorine-18-FDG. Society of Nuclear Medicine. *J Nucl Med*. 1998;39:1302-5. [PMID: 9669415]

107. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. In: Kelsey JL, Marmot MG, Stolley PD, Vessey MP, eds. *Monographs in Epidemiology and Biostatistics*. 2nd ed. v 26. New York: Oxford Univ Pr; 1996.

108. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667-76. [PMID: 8135452]

109. Efron B. Non-parametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*. 1981;68:589-99.

Current Author Addresses: Drs. Gould and Kushner: Veterans Affairs Palo Alto Health Care System (111P), 3801 Miranda Avenue, Palo Alto, CA 94304.

Ms. Rydzak: 201 Rawson Road #3, Brookline, MA 02445.

Ms. Maclean: 2614 Cedar Creek Drive, Durham, NC 27705.

Dr. Demas: 2330 Post Street, Suite 460, San Francisco, CA 94115.

Dr. Shigemitsu: Veterans Affairs Medical Center (111), 1030 Jefferson Avenue, Memphis, TN 38104.

Ms. Chan and Dr. Owens: Center for Primary Care and Outcomes Research, 117 Encina Commons, Stanford, CA 94305-6019.

Author Contributions: Conception and design: M.K. Gould, D.K. Owens.

Analysis and interpretation of the data: M.K. Gould, W.G. Kushner, C.E. Rydzak, C.C. Maclean, H. Shigemitsu, D.K. Owens.

Drafting of the article: M.K. Gould, D.K. Owens.

Critical revision of the article for important intellectual content: M.K. Gould, W.G. Kushner, C.E. Rydzak, C.C. Maclean, A.N. Demas, H. Shigemitsu, D.K. Owens.

Final approval of the article: M.K. Gould, W.G. Kushner, C.E. Rydzak, C.C. Maclean, A.N. Demas, H. Shigemitsu, D.K. Owens.

Statistical expertise: M.K. Gould, D.K. Owens.

Obtaining of funding: M.K. Gould, D.K. Owens.

Administrative, technical, or logistic support: C.E. Rydzak, A.N. Demas, J.K. Chan.

Collection and assembly of data: M.K. Gould, W.G. Kushner, C.E. Rydzak, C.C. Maclean, A.N. Demas, J.K. Chan.

APPENDIX

Literature Searches

We performed the initial literature search in August 2001 and repeated searches of computerized databases in June 2002. These searches identified 570 potentially relevant studies, including 7 studies that did not appear in MEDLINE, CancerLit, or EMBASE (Figure 1). We excluded 497 studies after scanning their titles and abstracts: 123 review articles, meta-analyses, and cost-effectiveness analyses; 117 studies that examined FDG-PET for applications in thoracic oncology rather than mediastinal lymph node staging; 68 studies that examined FDG-PET for oncologic applications outside of lung cancer; 48 studies that focused on technical aspects of PET imaging; 25 case reports; and 116 other studies for miscellaneous reasons.

We subsequently evaluated 73 full reports for inclusion. Of these, we excluded 29 studies that did not enroll the required number of participants (23–31), that did not provide enough data to permit calculation of sensitivity and specificity (25, 28, 32–46), or that reported duplicate data (32, 47–51). Four additional studies were excluded because they evaluated FDG imaging with a modified gamma camera (52–55). We also excluded 3 non-English-language papers that were review articles or case reports (56–58). Another study of mediastinal staging was excluded because almost one third of the participants had small-cell lung cancer or mesothelioma (59).

The κ coefficient for inter-rater agreement for study eligibility was 0.64, indicating very good agreement for studies identified during the initial search.

More recently, 1 author participated in a technology assessment of FDG-PET imaging for the Department of Veterans Af-

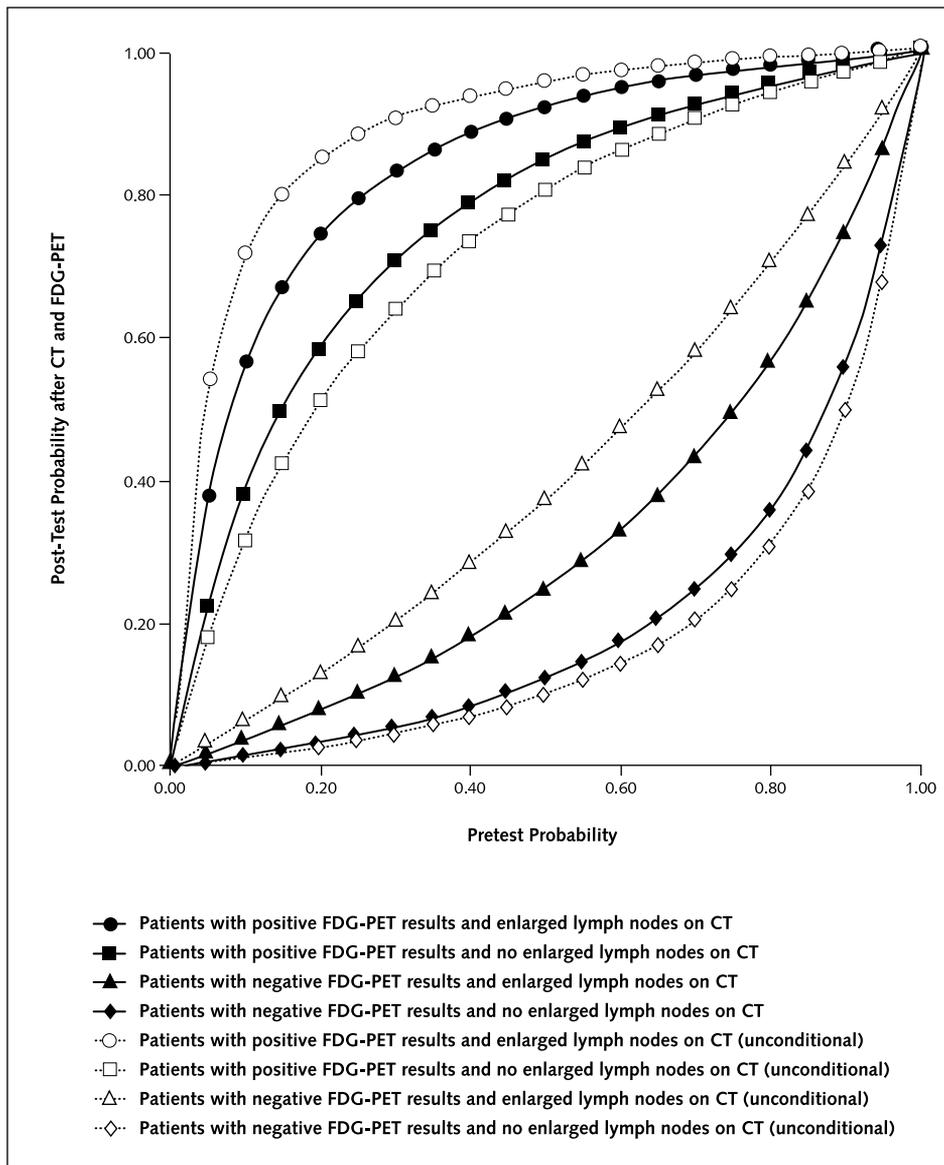
fairs. The technology assessment focused on the role of FDG-PET in managing patients with solitary pulmonary nodules, colon cancer, and non-small-cell lung cancer. A professional librarian searched MEDLINE, EMBASE, Current Contents, CancerLit (which is now defunct and rolled into MEDLINE), and BIOSIS by using Dialog, covering 1998 through January 2003. A full range of descriptors, text, keywords, and synonyms was used: *tomography, emission computed, positron emission tomography, gamma camera, PET, staging, solitary nodules, coin lesions, and colorectal cancer*. The search retrieved 340 unique reports. A second search using various approaches was done on 27 March 2003 to minimize the possibility of poor retrieval due to indexing flaws. In another series of searches, we expanded retrieval by using the “related articles” link available only from PubMed and retrieved articles related to citations that were deemed highly relevant but had not been retrieved with the first search. The original search strategy was then edited and expanded by adding additional terms for PET and eliminating selected terms for study type and quality but keeping terms for predictive value of tests, sensitivity and specificity, and neoplasm staging. This second search retrieved an additional 181 citations along with many duplicates. In total, after eliminating duplications, 508 full records were retrieved for the Veterans Affairs’ review.

Of these, 70 studies were judged to be potentially relevant to mediastinal staging with FDG-PET. Many were previously identified in the original literature review. One author evaluated 6 unique reports that were not identified previously. Three reports were excluded because they did not present enough data to permit calculation of sensitivity and specificity (60–62). In 1 of these studies, we could not calculate sensitivity and specificity because the final diagnosis was not reported for 6 participants with “indeterminate” results on FDG-PET imaging. Three studies from the supplemental search were included in the meta-analysis (91, 99, 101).

Study Quality

To assess study quality, we adapted criteria developed by Kent and colleagues (11), who evaluated imaging tests for the diagnosis of lumbar spinal stenosis. The revised criteria include 22 items that cover 8 dimensions of study quality: technical quality of FDG-PET, technical quality of CT, technical quality and application of the reference test or tests, independence of test interpretation, description of the study sample, cohort assembly, sample size, and unit of data analysis (Appendix Table 3). To develop criteria for the technical quality of FDG-PET, we consulted 2 nuclear medicine physicians experienced in FDG-PET imaging and referred to guidelines published by the Society of Nuclear Medicine (106). We used a 2-part definition to determine whether studies met criteria for adequacy of the reference test. To verify mediastinal lymph node involvement, we required confirmation by any type of biopsy. To verify the absence of mediastinal lymph node involvement, we required thoracotomy with systematic sampling of both normal- and abnormal-appearing lymph nodes at all accessible mediastinal stations. The median κ coefficient for inter-rater reliability was 0.67, indicating very good agreement.

Appendix Figure. Post-test probabilities of mediastinal metastasis after computed tomography (CT) and positron emission tomography with 18-fluorodeoxyglucose (FDG-PET).



Post-test probabilities are shown as a function of pretest probability in patients with positive FDG-PET results and enlarged lymph nodes on CT (*solid circles*), patients with positive FDG-PET results and no enlarged lymph nodes on CT (*solid squares*), patients with negative FDG-PET results and enlarged lymph nodes on CT (*solid triangles*), and patients with negative FDG-PET results and no enlarged lymph nodes on CT (*solid diamonds*). When unconditional estimates of FDG-PET performance are used to make the calculations, post-test probabilities are overestimated when FDG-PET results are positive and CT shows enlarged lymph nodes (*open circles*), underestimated when FDG-PET results are positive and CT shows no enlarged lymph nodes (*open squares*), overestimated when FDG-PET results are negative and CT shows enlarged lymph nodes (*open triangles*), and underestimated when FDG-PET results are negative and CT shows no enlarged lymph nodes (*open diamonds*).

Data Synthesis and Statistical Analysis

For each study, we constructed 2×2 contingency tables in which all participants were classified as having positive (N2 or N3) or negative (N0 or N1) results and as having or not having mediastinal lymph node involvement, as determined by the reference test or tests. We calculated the true-positive rate (true-positive rate = sensitivity), the false-positive rate (false-positive rate = $1 - \text{specificity}$), and the log odds ratio (log odds true-positive rate - log odds false-positive rate) for CT and FDG-

PET. The log odds ratio is a measure of diagnostic test performance that accounts for the positive correlation between the true-positive rate and the false-positive rate. We added 0.5 to each cell in any 2×2 table that contained 1 or more zero values; otherwise, it would not be possible to compute the log odds ratio for studies that reported perfect sensitivity or specificity. We calculated exact 95% CIs for the true-positive rate and the false-positive rate on the basis of the binomial distribution (14).

We used several methods for constructing summary ROC

curves. These methods have been described previously (13, 15, 16). The ROC curves illustrate the tradeoff between sensitivity and specificity, as the threshold that defines a positive test result varies from most stringent to least stringent. Our methods for constructing summary ROC curves depend on the assumption that individual study estimates of sensitivity and specificity represent unique points on a common ROC curve.

We used the method of Moses and colleagues (12) to test the hypothesis that the ROC curve is symmetrical and therefore can be described by a single parameter, the summary log odds ratio. In this method, the true-positive rate and false-positive rate are logistically transformed and simple linear regression is performed by using the log odds ratio as the dependent variable and an implied function of the test threshold (log odds true-positive rate + log odds false-positive rate) as the independent variable. As this function increases, the test threshold becomes less stringent. The slope of the regression equation indicates the degree to which the summary ROC curve is not symmetrical, while the intercept is a measure of diagnostic accuracy. A limitation of this method is that the logistic transformation requires the use of a correction factor when the reported sensitivity or specificity is 100%.

When the slope of the regression equation is not statistically significantly different from 0, the resulting ROC curve is symmetrical and can be described by the intercept, which is the summary log odds ratio. When this condition was met, we used a fixed-effects model for combining odds ratios when there was no evidence of heterogeneity (17) and a random-effects model when there was statistical evidence of heterogeneity (18), because these methods do not require the use of a correction factor (107). Nevertheless, all methods produced similar results.

For a global measure of test performance, we expressed our results in terms of the maximum joint sensitivity and specificity (12). The maximum joint sensitivity and specificity is the point on the summary ROC curve at which sensitivity and specificity are equal. It varies from 0.5 for a diagnostic test that has no diagnostic value to 1.0 for a diagnostic test that is perfect. Of note, the maximum joint sensitivity and specificity does not necessarily define the optimal operating point on the summary ROC curve but rather is a global measure of test performance, similar to the area under the curve. We calculated the maximum joint sensitivity and specificity by using the formula $Q^* = (1 + e^{-A/2})^{-1}$, where Q^* is the maximum joint sensitivity and specificity and A is the summary log odds ratio (12).

To estimate the approximate sensitivity and specificity of FDG-PET and CT in current clinical practice, we selected a point on the summary ROC curve that corresponded to the median specificity in the individual studies (108). Other approaches are possible but not necessarily better. We selected the point that corresponded to the median specificity because the data were not normally distributed; specificity was less variable than sensitivity for FDG-PET; and if we had selected the point on the curve that corresponded to the median sensitivity, the estimated sensitivity and specificity of FDG-PET in patients with enlarged lymph nodes on CT would have been 100% and 0%, respectively. It is beyond the scope of this analysis to identify the

Appendix Table 1. Initial Search Strategy for Computerized Databases*

S1 explode lung neoplasms/
S2 explode carcinoma, non-small-cell lung/
S3 1 or 2
S4 explode tomography, emission-computed/
S5 positron emission tomography.mp.
S6 pet\$.mp.
S7 animal not (human and animal).mp.
S8 6 not 7
S9 explode deoxyglucose/ or deoxyglucose.mp.
S10 deoxy-glucose.mp.
S11 fluorodeoxyglucose.mp.
S12 18fluorodeoxyglucose.mp.
S13 fludeoxyglucose.mp.
S14 fdg.mp.
S15 18fdg.mp.
S16 f-18-fdg.mp.
S17 fluoro-2-deoxy-d-glucose.mp.
S18 2fluoro-2deoxyglucose.mp.
S19 fluoro-d-glucose.mp.
S20 4 or 5 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19
S21 3 and 20

optimal operating point on the summary ROC curve for FDG-PET. This question can be addressed by using decision analysis.

To compare the sensitivity and specificity of FDG-PET in patients with and without lymph node enlargement, we performed discriminant function analysis (20). This technique is useful for comparing groups on the basis of 1 or more attributes. Because sensitivity and specificity were not normally distributed and their covariances were not equal in patients with and without lymph node enlargement, we used a permutation test to obtain a P value rather than relying on normal theory (21). We computed the discriminant analysis statistic and its permutation distribution under the null hypothesis of no group difference, that is, we randomly assigned individual study estimates of sensitivity and specificity to the 2 groups and recomputed the statistic 500 times. One of 500 permutations resulted in an F statistic that was more extreme than the one observed ($F = 15.03$), resulting in a nonparametric P value of 0.002.

To estimate posterior (post-test) probabilities and their 95% CIs, we used the bootstrap (109) and resampled data from the individual studies 999 times. For each of the 999 bootstrap samples, we calculated the summary log odds ratio by using the Mantel-Haenszel method, the median specificity, the sensitivity at the point on the ROC curve that corresponded to the median specificity, likelihood ratios for positive and negative test results, and post-test probabilities for different values of pretest probability. To estimate 95% CIs for post-test probabilities, we assumed that pretest probabilities were not uncertain but incorporated the uncertainty in our estimates of sensitivity and specificity.

We used Microsoft Excel 2000 (Microsoft, Inc., Redmond, Washington) to estimate summary diagnostic odds ratios by using fixed- and random-effects models and to perform statistical tests for heterogeneity. We used JMP, version 3.2.6 (SAS Institute, Cary, North Carolina), to perform all regression analyses and SPSS for Windows, version 11.5 (SPSS, Inc., Chicago, Illi-

Appendix Table 2. Supplementary Search Strategy Employed in Veterans Affairs Technology Assessment

Set	Items, n	Description
S1	72 496	POSITRON()EMISSION(2N)TOMOGRAPHY?/TI,DE OR TOMOGRAPHY, EMISSION-COMPUTED!
S2	10 040	PET/TI AND HUMAN/DE
S3	1874	COMPUTER(1N)ASSIST?(1N)EMISSION(1N)TOMOGRAPH?/TI,DE OR WHOLE BODY TOMOGRAPHY!
S4	74 203	S1 OR S2 OR S3
S5	3993	FDG()PET/TI
S6	75 236	S4 OR S5
S7	9691	DEOXYGLUCOSE!
S8	38 421	DEOXYGLUCOSE OR DEOXY()GLUCOSE OR FLUORODEOXYGLUCOSE OR 18FLUORODEOXYGLUCOSE OR FLUDEOXYGLUCOSE OR FDG/TI
S9	1782	18FDG OR F(18)()DG OR FLUORO()2()DEOXY()D()GLUCOSE OR 2FLUORO()2DEOXYGLUCOSE OR FLUORO()D()GLUCOSE
S10	17 371	S6 AND (S7 OR S8 OR S9)
S11	153 976	LUNG(1N)(NEOPLAS? OR CANCER? OR CARCINOMA? OR TUMOR? OR TUMOUR?)/TI,DE
S12	14 272	CARCINOMA, NON-SMALL-CELL LUNG! OR CARCINOMA, SMALL CELL!
S13	29 401	SMALL(1N)CELL(1N)CARCINOMA?/TI,DE OR NON(1N)SMALL(1N)CELL(1N)CARCINOMA/TI,DE
S14	25 795	(S12 OR S13) AND (LUNG? OR BRONCH? OR PULMONARY?)/TI,DE
S15	5194	SOLITARY()PULMONARY()NODULE?/TI,DE OR LUNG(1N)NODULE?/TI,DE OR COIN()LESION?/TI,DE OR PULMONARY()NODULE?/TI,DE
S16	6470	MEDIASTIN?(1N)(STAG? OR NEOPLAS? OR TUMOR? OR TUMOUR? OR CANCER? OR CARCINOMA?)/TI,DE
S17	77 796	NEOPLAS?(1N)STAG? OR CANCER(1N)STAG?/TI,DE
S18	1914	S16 AND (LUNG? OR BRONCHO? OR PULMON?)/TI,DE
S19	68 827	COLORECTAL NEOPLASMS! OR COLORECTAL CANCER!
S20	79 722	COLORECTAL?()NEOPLAS? OR TUMOR? OR TUMOUR? OR CANCER? OR CARCINOMA?)/TI,DE
S21	3014	S10 AND (S11 OR S12 OR S14 OR S15 OR S17 OR S18 OR S19 OR S20)
S22	2689	S21/ENG
S23	1606	RD (unique items)
S24	178	S23 AND (EFFICAC? OR EFFECTIV? OR ASSESS? OR META()ANALY? OR METAANALY? OR CONSENSUS? OR POSITION()PAPER?)/TI,DE
S25	167	S23 AND (REVIEW? OR REPORT? OR EVIDENCE? OR BIBLIOGRAPHY? OR GUIDELINE? OR SYSTEMATIC?)/TI,DE
S26	38	S23 AND (DT=REVIEW? OR QUANTITATIVE? OR QUALITATIVE? OR SURVEY?)/TI,DE
S27	292	S23 AND (SENSITIVITY A?D SPECIFICITY/TI,DE) OR (PREDICTIVE()VALUE)TESTS/TI,DE)
S28	29	S23 AND (RECOMMENDATION? OR PROTOCOL? OR CLINICAL()PATH OR CRITICAL()PATH)/TI,DE
S29	2689	S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28
S30	2122	S29 AND PY=1998:2003
S31	2087	S30/HUMAN
S32	2087	S31/ENG
S33	166 174	CLINICAL?()PROTOCOL? OR FEASIBILITY()STUD? OR REPRODUCIBILITY(1N)RESULT? OR RESEARCH()DESIGN
S34	775 763	DOUBLE()BLIND? OR RANDOM()ALLOCAT? OR CLINICAL()TRIAL? OR CONTROL()STUD??? ?
S35	614 005	CLINICAL?()STUD??? ?
S36	1 132 596	RANDOM?()CONTROL? OR COMPARATIV?()STUD??? ?
S37	1 330 134	CONTROL?(2N)TRIAL? OR EFFICACY OR EFFECTIVENESS
S38	706 082	GUIDELINE? OR CONSENSUS()DEVELOP? OR RECOMMENDATION? OR PROTOCOL? OR CLINICAL()PATH? OR POSITION()PAPER? OR CRITICAL()PATH?
S39	1 593 834	META()ANALY? OR META-META-ANALY? OR METAANALY? OR DT=META-ANALYSIS OR DT=REVIEW? OR DT=GUIDELINE? OR CONSENSUS?/TI
S40	244 195	OVERVIEW?/TI OR COCHRANE?/TI OR SURVEY?/TI OR EVIDENCE()BASE?/TI,AB,DE
S41	435 008	OVERVIEW?/TI OR COCHRANE?/TI OR SURVEY?/TI OR BIBLIOGRAPH?/TI OR SYSTEMATIC?/TI OR CRITICAL?/TI OR METHODOLOGIC?/TI
S42	315 571	QUANTITATIVE?/TI OR QUALITATIVE?/TI OR LITERATURE?/TI
S43	380 141	EVIDENCE?/TI OR EVIDENCE()BASE?/TI,AB,DE
S44	205 731	(S40 OR S41 OR S42) AND (REVIEW?/TI OR REPORT?/TI OR ASSESS?/TI,DE OR DT=REVIEW?)
S45	4 813 844	S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39
S46	564 422	S43 OR S44
S47	1032	S32 AND (S45 OR S46)
S48	629	RD (unique items)

nois), to perform discriminant function analysis. We used Microsoft Excel and Crystal Ball 2000 standard edition simulation software (Decionering, Inc., Denver, Colorado) to generate bootstrap confidence intervals for post-test probabilities.

Sensitivity Analysis

To determine whether particular study characteristics affected diagnostic accuracy, we used meta-regression. We performed multiple linear regression analysis by introducing additional covariates into the simple linear regression model described

above, one at a time. Specifically, we tested the effect of each item in our study quality instrument, as well as language and year of publication. We also examined the effect of overall study quality by entering a variable that indicated whether the study satisfied at least 70% of our criteria for methodologic quality. In other analyses, we examined the effect of excluding studies that reported the sensitivity and specificity of FDG-PET for distinguishing no lymph node involvement (N0) from any lymph node involvement (N1, N2, or N3) and studies that enrolled fewer than 10 participants with and without mediastinal metastasis.

RESULTS

Evidence of Statistical Heterogeneity

We used a chi-square test to identify heterogeneity in log odds ratios for studies that analyzed results by using the patient as the unit of analysis (18). There was no evidence of statistical heterogeneity in studies of CT ($P > 0.2$) or in studies of FDG-PET that reported results for patients with ($P > 0.2$) and without ($P > 0.2$) lymph node enlargement separately. Therefore, we used a fixed-effects model to derive estimates of diagnostic test performance for these data. Because we identified evidence of statistical heterogeneity in the log odds ratios of studies of FDG-PET ($P = 0.05$), we used a random-effects model to derive summary estimates of test performance for these studies.

We also inspected log odds ratios of individual studies to identify “outliers” that contributed to heterogeneity in the FDG-PET studies. Three studies were responsible for nearly 30% of the total heterogeneity (74, 87, 94). One high-quality study in 68 patients with a relatively high prevalence of mediastinal metastasis (41%) reported very high estimates of sensitivity (93%) and specificity (95%) (74). Another high-quality study of 81 participants with a lower prevalence of mediastinal metastasis (26%) reported poor sensitivity (52%) and intermediate specificity (88%) (94). A smaller study (22 participants) of average quality with a low prevalence of metastasis (27%) reported relatively low estimates for both sensitivity (67%) and specificity (69%) (87). Excluding these studies individually did not greatly affect statistical heterogeneity ($P = 0.08$ to 0.12), but excluding all 3 studies reduced the heterogeneity to a nonsignificant level ($P > 0.2$). Summary estimates of diagnostic accuracy were the same (maximum joint sensitivity and specificity, 86%) before and after excluding each of these studies individually and all 3 of them simultaneously.

We could not identify other sources of heterogeneity besides the year of publication and specifying whether participants underwent FDG-PET imaging in the fasting state. Of note, all 3 of the studies we mentioned earlier reported that participants underwent FDG-PET in the fasting state.

To examine diagnostic accuracy in studies that were least likely to be biased, we restricted the analysis to 9 studies of FDG-PET in which participants were enrolled prospectively, FDG-PET and CT readers were blinded to the results of the reference test or tests, and systematic sampling of both normal and abnormal-appearing lymph nodes at all accessible mediastinal stations was used to exclude mediastinal metastasis (64, 66, 67, 70, 73–76, 90). In this analysis, there was no evidence of statistical heterogeneity ($P = 0.18$), and diagnostic accuracy was slightly higher than our base-case estimate (maximum joint sensitivity and specificity, 89% [CI, 85% to 91%]), although the difference between these studies and those that did not meet all 3 criteria was not statistically significant ($P = 0.10$).

Bias in Estimates of Sensitivity and Specificity when the Unit of Analysis Is Not the Patient

To illustrate how not using the patient as the unit of analysis can result in biased estimates of sensitivity and specificity, we consider 10 hypothetical patients with potentially resectable

Appendix Table 3. Criteria for Assessing Study Quality*

Technical quality of FDG-PET Spatial resolution < 11 mm FDG uptake period ≥ 30 min FDG dose ≥ 10 mCi Acquisition time for emission scan specified Attenuation correction performed Participants with hyperglycemia excluded Participants studied in the fasting state Positive test results defined according to specific criteria
Technical quality of CT Section thickness ≤ 10 mm Intravenous contrast used or section thickness ≤ 5 mm to view aortopulmonary window Acquisition time ≤ 2 s or spiral mode used Positive test results defined according to specific criteria
Technical quality and application of the reference test or tests Biopsy to confirm mediastinal metastasis To confirm the absence of mediastinal metastasis, thoracotomy with systematic sampling of normal and abnormal lymph nodes at all accessible stations
Independence of test interpretation CT and FDG-PET readers blinded to the results of the reference test or tests CT readers blinded to FDG-PET results FDG-PET readers blinded to CT results
Clinical characteristics of the study sample described (age, sex, number of patients with non-small-cell lung cancer, and prevalence of mediastinal metastasis)
Cohort assembly Participants enrolled prospectively Clinically relevant cohort (patients with potentially resectable non-small-cell cancer)
Sample size of at least 35 participants with or without mediastinal metastasis
Individual patient used as unit of data analysis

* CT = computed tomography; FDG = 18-fluorodeoxyglucose; FDG-PET = positron emission tomography with 18-fluorodeoxyglucose.

non-small-cell lung cancer (Appendix Tables 5 and 6). Each patient undergoes FDG-PET imaging and has normal- and abnormal-appearing lymph nodes sampled from 5 accessible mediastinal stations during thoracotomy. At thoracotomy, 4 patients have evidence of mediastinal metastasis at 1 of 5 lymph node stations; FDG-PET results are true-positive at the involved lymph node station in all 4 of these patients. One patient has metastasis at all 5 lymph node stations; FDG-PET results are false-negative at all 5 stations. Five other patients have no evidence of lymph node metastasis; FDG-PET results are true-negative at all 5 lymph node stations in 4 of the patients and are false-positive at 1 lymph node station in 1 patient. When the lymph node station is the unit of analysis, the calculated prevalence of mediastinal metastasis is 18%, sensitivity is 44% (4 true-positive lymph node stations among 9 stations with mediastinal metastasis), and specificity is 98% (40 true-negative lymph node stations among 41 stations without mediastinal metastasis). In contrast, when the individual patient is the unit of analysis, the calculated prevalence of mediastinal metastasis is 50%, sensitivity is 80% (4 patients with true-positive findings among 5 patients with mediastinal metastasis), and specificity is 80% (4 patients with true-negative findings among 5 patients without mediastinal metastasis). Thus, in this example, not using the patient as the unit of analysis overestimates specificity and underestimates disease prevalence and sensitivity.

Appendix Table 4. Characteristics of Participants, Diagnostic Accuracy, and Aspects of Methodologic Quality in Studies of Computed Tomography and Positron Emission Tomography with 18-Fluorodeoxyglucose for Mediastinal Staging*

Author, Year (Reference)	Participants		Mean Age ± SD (Range)	Staging Evaluations	Prevalence of Mediastinal Metastasis
	n	%	y	n	
Studies in which the individual patient was the unit of analysis					
Wahl et al., 1994 (63)	23	74	63.8 (42–84)	27†	41
Chin et al., 1995 (64)	30	63	61 (47–76)	30	30
Valk et al., 1995 (66)‡	99	53	66 (46–87)	76§	32
Bury et al., 1996 (67)	50	NR	65 (44–75)	50	32
Sazon et al., 1996 (69)‡	107	99	62 ± 9	32	50
Scott et al., 1996 (70)	27	81	64 (40–85)	27	33
Guhlmann et al., 1997 (71)‡	46	89	56.7 (24–78)	32	47
Hagberg et al., 1997 (72)‡	49	92	63	18	50
Steinert et al., 1997 (73)‡	62	82	59.9 (39–75)	47	28
Vansteenkiste et al., 1998 (74)	68	NR	64 (40–83)	68	41
Vansteenkiste et al., 1998 (75)	56	NR	62 (33–75)	56	50
Albes et al., 1999 (76)	27	89	59	27	59
Higashi et al., 1999 (78)	42	NR	NR	42	19
Magnani et al., 1999 (80)	28	93	63 (50–75)	28	32
Marom et al., 1999 (81)‡	100	58	63 (25–83)	79	56
Richter et al., 1999 (82)	55	NR	NR	22	41
Saunders et al., 1999 (83)‡	97	66	63.3 (36–77)	81	20
Demura et al., 2000 (85)	65	NR	NR	25	36
Farrell et al., 2000 (86)	84	54	66 (47–82)	83¶	5
Kitase et al., 2000 (87)	22	59	64	22	27
Kubota et al., 2000 (88)	36	NR	NR	36	50
Liewald et al., 2000 (89)	80	85	69 (24–78)	80	31
Pieterman et al., 2000 (90)	102	86	63 (25–77)	102	31
Roberts et al., 2000 (91)	100	NR	NR	100	24
Tatsumi et al., 2000 (92)	21	57	61.7 ± 11.2	21	48
Changlai et al., 2001 (93)‡	156	48	67	127	64
Dunagan et al., 2001 (94)	152	NR	64 ± 9	81	26
Guan et al., 2001 (95)‡	82	61	Men: 56 Women: 58	42	43
Gupta et al., 2001 (96)‡	111	66	(35–84)	77	31
Poncelet et al., 2001 (97)	64	4:1 ratio	65.2 (45–83)	62	15
Kernstine et al., 2002 (98)	237	58	65 ± 11	237	19
Vesselle et al., 2002 (99)	142	NR	NR	118	36
von Haag et al., 2002 (100)	52	NR	NR	52	12
Studies in which lymph nodes, lymph node stations, or mediastinal sides were the unit of analysis					
Patz et al., 1995 (65)	42	62	57 (25–85)	62 stations	37
Sasaki et al., 1996 (68)	29	69	65 (46–82)	71 regions	24
Scott et al., 1996 (70)**	27	81	64 (40–85)	27 stations	13
Steinert et al., 1997 (73)‡**	62	82	59.9 (39–75)	112 stations	25
Vansteenkiste et al., 1998 (74)**	68	NR	64 (40–83)	690 stations	7
Vansteenkiste et al., 1998 (75)¶**	56	NR	62 (33–75)	493 stations	12
Berlangieri et al., 1999 (77)‡	50	74	64 (41–78)	201 stations	10
Kernstine et al., 1999 (79)	64	64	65 ± 9 (33–77)	122 sides	16
Weber et al., 1999 (84)‡	27	96	62 ± 9	88 stations	13
Gupta et al., 2001 (96)‡**	111	66	(35–84)	288 nodes	19
Graeter et al., 2003 (101)‡	102	84	62 ± 9	380 stations (82 patients)	21

* CT = computed tomography; FNR = false-negative rate; FPR = false-positive rate; NC = not able to calculate; NR = not reported; PET = positron emission tomography; TPR = true-positive rate; TNR = true-negative rate.

† 27 mediastinal sides were evaluated in 23 participants, including 4 participants with benign diagnoses who were presumed to have no lymph node involvement on 8 sides of the mediastinum.

‡ In these studies, summary data about clinical characteristics were provided for all participants, but separate data for the subgroup of participants who underwent mediastinal staging were not provided.

§ Two participants had bilateral tumors. Thus, 76 mediastinal sides were evaluated in 74 participants with 76 pulmonary tumors.

¶ Because this study and the study by Liewald et al. (89) probably had overlapping patient populations, this study was included only for calculating estimates of conditional test performance.

¶ Mediastinal status was not reported for 1 participant with evidence of distant metastases on positron emission tomography with 18-fluorodeoxyglucose.

** These studies reported results using both the patient and mediastinal stations or individual lymph nodes as the unit of analysis.

Appendix Table 4—Continued

Computed Tomography		Positron Emission Tomography		Prospective Enrollment?	Optimal Reference Tests Used?	PET and CT Readers Blinded to Results of Reference Tests?
Sensitivity (TPR/[TPR + FNR])	Specificity (TNR/[TNR + FPR])	Sensitivity (TPR/[TPR + FNR])	Specificity (TNR/[TNR + FPR])			
←----- % ----->						
63.6 (7/11)	43.8 (7/16)	81.8 (9/11)	81.3 (13/16)	Yes	No	Yes
44.4 (4/9)	90.5 (19/21)	66.7 (6/9)	85.7 (18/21)	Yes	Yes	Yes
62.5 (15/24)	73.1 (38/52)	83.3 (20/24)	94.2 (49/52)	Yes	No	Yes
75 (12/16)	87.9 (29/33)	87.5 (14/16)	97 (32/33)	Yes	Yes	Yes
81.3 (13/16)	56.3 (9/16)	100 (16/16)	100 (16/16)	No	No	Yes
66.7 (6/9)	83.3 (15/18)	100 (9/9)	100 (18/18)	Yes	Yes	Yes
46.7 (7/15)	82.4 (14/17)	86.7 (13/15)	100 (17/17)	No	Yes	Yes
55.6 (5/9)	100 (9/9)	66.7 (6/9)	100 (9/9)	No	No	Yes
NC	NC	92.3 (12/13)	97.1 (33/34)	Yes	Yes	Yes
75 (21/28)	62.5 (25/40)	92.9 (26/28)	95 (38/40)	Yes	Yes	Yes
85.7 (24/28)	42.9 (12/28)	85.7 (24/28)	78.6 (22/28)	Yes	Yes	Yes
93.8 (15/16)	72.7 (8/11)	87.5 (14/16)	81.8 (9/11)	Yes	Yes	Yes
50 (4/8)	73.5 (25/34)	62.5 (5/8)	94.1 (32/34)	No	Yes	No
66.7 (6/9)	84.2 (16/19)	66.7 (6/9)	84.2 (16/19)	Yes	Yes	No
59.1 (26/44)	85.7 (30/35)	90.9 (40/44)	88.6 (31/35)	Yes	No	No
55.6 (5/9)	92.3 (12/13)	100 (9/9)	92.3 (12/13)	No	No	No
17.6 (3/17)	89.6 (60/67)	70.6 (12/17)	97 (65/67)	No	No	Yes
NR	NR	66.7 (6/9)	81.3 (13/16)	No	No	No
0 (0/11)	100 (73/73)	100 (4/4)	92.4 (73/79)	No	No	No
0 (0/6)	100 (16/16)	66.7 (4/6)	68.8 (11/16)	No	Yes	No
66.7 (4/6)	66.7 (8/12)	50 (3/6)	100 (12/12)	Yes	No	Yes
NR	NR	92 (23/25)	76.4 (42/55)	No	Yes	Yes
75 (24/32)	65.7 (46/70)	90.6 (29/32)	85.7 (60/70)	Yes	Yes	Yes
NR	NR	87.5 (21/24)	90.7 (69/76)	No	Yes	Yes
NR	NR	80 (8/10)	81.8 (9/11)	Yes	No	Yes
NR	NR	87.7 (71/81)	82.6 (38/46)	No	No	Yes
50 (9/18)	87 (47/54)	52.4 (11/21)	88.3 (53/60)	No	No	Yes
61.1 (11/18)	79.2 (19/24)	94.4 (17/18)	100 (24/24)	No	No	No
NC	NC	87.5 (21/24)	88.7 (47/53)	No	Yes	No
55.6 (5/9)	67.9 (36/53)	66.7 (6/9)	83.0 (44/53)	Yes	Yes	No
NR	NR	81.8 (36/44)	81.9 (158/193)	No	No	No
NR	NR	81 (34/42)	96.1 (73/76)	Yes	No	No
50 (3/6)	65.2 (30/46)	66.7 (4/6)	91.3 (42/46)	No	No	No
43.5 (10/23)	84.6 (33/39)	82.6 (19/23)	82.1 (32/39)	Yes	Yes	Yes
64.7 (11/17)	87 (47/54)	76.5 (13/17)	98.1 (53/54)	No	Yes	No
60 (6/10)	93.8 (61/65)	100 (10/10)	98.5 (64/65)	Yes	Yes	Yes
57.1 (16/28)	94 (79/84)	89.3 (25/28)	98.8 (83/84)	Yes	Yes	Yes
46.8 (22/47)	95.8 (616/643)	89.4 (42/47)	98.9 (636/643)	Yes	Yes	Yes
		92.1 (399/433)	63.3 (38/60)			
65 (13/20)	89.5 (162/181)	80 (16/20)	96.7 (175/181)	Yes	Yes	No
65 (13/20)	80.4 (82/102)	70 (14/20)	87.3 (89/102)	Yes	Yes	No
90.9 (10/11)	92.2 (71/77)	100 (11/11)	97.4 (75/77)	Yes	Yes	Yes
67.9 (36/53)	61.3 (141/230)	86.8 (46/53)	92.2 (212/230)	No	Yes	No
NR	NR	93.8 (75/80)	84.3 (253/300)	No	Yes	No

Appendix Table 5. Hypothetical Data To Illustrate Bias in Estimates of Sensitivity and Specificity When the Patient Is Not the Unit of Analysis*

Patient	Lymph Node Station				
	1	2	3	4	5
1	TP	TN	TN	TN	TN
2	TP	TN	TN	TN	TN
3	TP	TN	TN	TN	TN
4	TP	TN	TN	TN	TN
5	FN	FN	FN	FN	FN
6	TN	TN	TN	TN	TN
7	TN	TN	TN	TN	TN
8	TN	TN	TN	TN	TN
9	TN	TN	TN	TN	TN
10	TN	TN	TN	TN	FP

* FN = false negative; FP = false positive; TN = true negative; TP = true positive. See Appendix Table 6 for calculations.

Additional Sensitivity Analyses

We obtained estimates of diagnostic accuracy that were very similar to our base-case estimate when we restricted the analysis to studies that enrolled at least 10 patients with mediastinal metastasis and 10 patients without mediastinal metastasis (maximum joint sensitivity and specificity, 87% [CI, 85% to 88%]). We also obtained similar estimates when we restricted the anal-

Appendix Table 6. Calculations Demonstrating Bias in Estimates of Sensitivity and Specificity When the Patient Is Not the Unit of Analysis*

Unit of Analysis	Value	Thoracotomy		
		Positive	Negative	Total
Lymph node station				
FDG-PET				
Positive		4	1	5
Negative		5	40	45
Total		9	41	50
Prevalence	0.18			
Sensitivity	0.44			
Specificity	0.98			
Patient				
FDG-PET				
Positive		4	1	5
Negative		1	4	5
Total		5	5	10
Prevalence	0.50			
Sensitivity	0.80			
Specificity	0.80			

* FDG-PET = positron emission tomography with 18-fluorodeoxyglucose.

ysis to studies that reported the sensitivity and specificity of FDG-PET for distinguishing N0 or N1 disease from N2 or N3 lymph node involvement (maximum joint sensitivity and specificity, 86% [CI, 84% to 88%]).