Importance  Positron emission tomography (PET) combined with fludeoxyglucose F 18 (FDG) is recommended for the noninvasive diagnosis of pulmonary nodules suspicious for lung cancer. In populations with endemic infectious lung disease, FDG-PET may not accurately identify malignant lesions.

Objectives  To estimate the diagnostic accuracy of FDG-PET for pulmonary nodules suspicious for lung cancer in regions where infectious lung disease is endemic and compare the test accuracy in regions where infectious lung disease is rare.

Data Sources and Study Selection  Databases of MEDLINE, EMBASE, and the Web of Science were searched from October 1, 2000, through April 28, 2014. Articles reporting information sufficient to calculate sensitivity and specificity of FDG-PET to diagnose lung cancer were included. Only studies that enrolled more than 10 participants with benign and malignant lesions were included. Database searches yielded 1923 articles, of which 257 were assessed for eligibility. Seventy studies were included in the analysis. Studies reported on a total of 8511 nodules; 5105 (60%) were malignant.

Data Extraction and Synthesis  Abstracts meeting eligibility criteria were collected by a research librarian and reviewed by 2 independent reviewers. Hierarchical summary receiver operating characteristic curves were constructed. A random-effects logistic regression model was used to summarize and assess the effect of endemic infectious lung disease on test performance.

Main Outcome and Measures  The sensitivity and specificity for FDG-PET test performance.

Results  Heterogeneity for sensitivity ($I^2 = 87\%$) and specificity ($I^2 = 82\%$) was observed across studies. The pooled (unadjusted) sensitivity was 89\% (95\% CI, 86\%-91\%) and specificity was 75\% (95\% CI, 71\%-79\%). There was a 16 percentage point–lower average adjusted specificity in regions with endemic infectious lung disease (61\% [95\% CI, 49\%-72\%]) compared with nonendemic regions (77\% [95\% CI, 73\%-80\%]). Lower specificity was observed when the analysis was limited to rigorously conducted and well-controlled studies. In general, sensitivity did not change appreciably by endemic infection status, even after adjusting for relevant factors.

Conclusions and Relevance  The accuracy of FDG-PET for diagnosing lung nodules was extremely heterogeneous. Use of FDG-PET combined with computed tomography was less specific in diagnosing malignancy in populations with endemic infectious lung disease compared with nonendemic regions. These data do not support the use of FDG-PET to diagnose lung cancer in endemic regions unless an institution achieves test performance accuracy similar to that found in nonendemic regions.

JAMA. 2014;312(12):1227-1236. doi:10.1001/jama.2014.11488
Clinicians rely heavily on radiographic imaging to identify and noninvasively diagnose lung nodules between 3 and 30 mm in diameter. The advent of lung cancer screening in high-risk populations using low-dose computed tomographic (CT) scans will increase the number of lung nodules detected, requiring clinical evaluation and diagnosis. Depending on the risk for cancer, diagnostic guidelines suggest or recommend fludeoxyglucose F 18 (FDG) combined with positron emission tomography (PET) as a noninvasive test to assess the risk of cancer or benign disease.

In previously published meta-analyses, FDG-PET was reported to be 90% to 94% accurate in the characterization of malignant or benign lung nodules, with a sensitivity of 94% to 96% and a specificity of 78% to 86%. Furthermore, combined FDG-PET and CT (FDG-PET/CT) scans have demonstrated a reduction in nontherapeutic resections (eg, resection for benign lesions or metastatic disease) by 17% to 20%. For these reasons, FDG-PET/CT is widely accepted for the clinical diagnosis and staging of lung cancer in patients with suspicious lung nodules.

Recent studies examining FDG-PET accuracy in diagnosing lung cancer in patients with lung lesions who reside in regions where fungal and other infectious lung diseases are endemic have shown mixed results. Histoplasmosis, coccidioidomycosis, and blastomycosis are the most prevalent fungal lung diseases in the United States and are common etiologies of lung granulomas. Histoplasmosis and blastomycosis are endemic across much of the Mississippi, Ohio, and Missouri river valleys through southern Ontario, Canada, whereas coccidioidomycosis is prevalent in the southwestern United States. Two international studies in areas with endemic tuberculosis found reduced FDG-PET/CT specificities of 25% and 21%.

We undertook a systematic review and meta-analysis of the literature published after the 2001 meta-analysis by Gould et al describing FDG-PET accuracy to diagnose lung cancer among patients being evaluated with lung nodules or masses. This updated meta-analysis investigates the accuracy of FDG-PET to diagnose lung lesions in regions with locally endemic infectious lung diseases.

Methods

Studies evaluating individuals for possible lung cancer using FDG-PET, FDG-PET/CT, or FDG-PET combined with another imaging modality were reviewed. We searched MEDLINE using the PubMed interface, EMBASE, and the Web of Science for studies published between October 1, 2000, and April 28, 2014 (eTable 1 in the Supplement). The guidelines for Preferred Reporting Items for Systematic Reviews and Meta-Analyses were followed. Reasons for study exclusion are detailed in Figure 1.

A study was classified as being from an endemic infectious region or population when the study reported presence of infectious lung diseases in the population from which participants were recruited or if granulomas arising from infectious lung diseases comprised at least 50% of reported benign etiologies. PET scan results were described by method of measuring FDG-PET avidity, the levels of risk or avidity, and the standardized uptake value threshold used to differentiate benign and cancerous diagnosis of disease. Details of study selection, data extraction, and synthesis are described in the eMethods in the Supplement.

Data Synthesis and Analysis

The sensitivity and specificity for FDG-PET test performance (pooled across studies) are displayed using forest plots. Study heterogeneity was assessed with the I2 statistic. Test performance in the presence of heterogeneity was summarized using hierarchical summary receiver operator curves (HSROC). Stability of test accuracy over time was assessed with diagnostic odds ratios and in the context of a random-effects model.

Publication bias was visually inspected using a funnel plot and quantitatively measured (eMethods in the Supplement).
Accuracy of FDG-PET to Diagnose Lung Cancer

Results

A total of 1923 articles were found; 16 articles were added from bibliography reviews along with an unpublished abstract. Forty-six articles were removed as duplicates and 1893 studies were screened. Upon initial abstract review, 1636 articles were excluded. An article could be excluded for multiple reasons, but the most common reason for exclusion during either portion of the review was inclusion of participants with 100% cancer prevalence (n = 1025).

Two hundred and fifty-seven studies received full review, of which 187 were excluded upon secondary review. The remaining 70 studies met all inclusion criteria and were used for the final analysis (eTable 4 in the Supplement). The total number of reported nodules being evaluated by FDG-PET among the 70 studies was 8511 and the median number of nodules per study was 83 (interquartile range, 56-140 nodules/study). Pooled cancer prevalence among nodules was 60% (n=5105 nodules). Individual study cancer prevalence varied from 21% to 86% across studies. Ten of 70 studies documented endemic infectious lung disease.10,11,14,15,24-29

The overall agreement for study eligibility between reviewers was 94.8% and the κ was 0.72 (using the method by Cohen), showing moderate agreement between reviewers. Consensus was used when reviewers disagreed; agreement was not reviewed quantitatively. Despite contacting corresponding authors, missing data on mean or median nodule size remained in 12 studies. Among the 49 studies reporting a mean or median lesion size, the median lesion size across studies was 2 cm (interquartile range, 1.7-2.8 cm).

Meta-analysis

An unadjusted pooled analysis of the 70 studies showed evidence of significant heterogeneity among studies in sensitivity (P of 87% [95% CI, 85%-90%]) and specificity (P of 82% [95% CI, 78%-86%]). Pooled sensitivity of FDG-PET for diagnosing lung cancer was 89% (95% CI, 86%-91%) and pooled specificity was 75% (95% CI, 71%-79%).

Ten studies reporting endemic disease had an unadjusted pooled specificity of 54% (95% CI, 37%-69%)10,11,14,15,24-29 compared with 78% (95% CI, 74%-81%) in the remaining 60 studies. The asymmetry test (using the method by Deeks et al48) did not show evidence of publication bias (P = .14). No trend over time or between periods in diagnostic accuracy was observed (eResults in the Supplement).

A random-effects model that included a random intercept for each study and various fixed effects for the study characteristics (eMethods, eTable 5, and the eFigure in the Supplement) was used to account for the observed heterogeneity. The model yielded an average adjusted estimate of sensitivity of 89% (95% CI, 87%-91%) and specificity of 75% (95% CI, 71%-78%) (Figure 2).30-89
The area under the HSROC curve (Figure 3) was 0.90 (95% CI, 0.87–0.92). The PIs show the extreme amount of heterogeneity among studies that remains after adjusting for study characteristics. The sensitivity of a randomly chosen study was predicted as 89% (95% PI, 70%–97%) and specificity was 75% (95% PI, 45%–91%). Similar increases in PI length were observed for all analyses. The results presented in the following sections are adjusted results from the random-effects model using multiple imputation.

**Infectious Lung Disease**

Ten studies reporting infectious lung disease endemic to the local population comprised 1431 individuals, of whom 1082 had cancer (76%). Granulomas as a percentage of benign diagno-
sensitivity ranged from 45% to 75%. Studies of populations in China, South Africa, and Japan reported tuberculosis as the common etiology for granulomatous disease.

The remaining North American studies found histoplasmosis, coccidioidomycosis, inflammation, and unspecified granuloma as common etiologies for pathologically diagnosed benign disease. Four of the 10 studies were retrospective and 7 studies determined diagnosis with pathology only.

The specificity was estimated to be 16 percentage points lower in populations with endemic infectious lung disease. This lower specificity persisted at 14% even for rigorously conducted and well-controlled studies. The average adjusted estimate of specificity in regions with endemic disease was 61% (95% CI, 49%-72%) compared with 77% (95% CI, 73%-80%) for nonendemic regions (Figure 3). For rigorously conducted and well-controlled studies, the estimates of specificity in endemic and nonendemic regions were 66% (95% CI, 51%-78%) and 80% (95% CI, 74%-85%), respectively.

The average adjusted sensitivity did not significantly differ by endemic status (94% [95% CI, 90%-96%] vs 88% [95% CI, 85%-90%] in nonendemic regions). The adjusted estimate of sensitivity in rigorously conducted and well-controlled studies was slightly higher in endemic regions (96% [95% CI, 92%-98%] vs 90% [95% CI, 86%-93%] in nonendemic regions).

Size of Lesion
Among the 34 studies reporting average or median lesion diameters less than or equal to 2 cm, the average adjusted sensitivity was 87% (95% CI, 84%-90%). In comparison, 23 studies with average or median diameters of greater than 2 cm had a slightly higher average adjusted sensitivity (91% [95% CI, 89%-93%]). Specificity of FDG-PET to diagnose lung cancer was not significantly influenced by lesion size (74% for studies with lesions ≤2 cm and 75% for studies with larger average lesion size).

Type of FDG-PET Scan
The 40 studies using FDG-PET/CT demonstrated slightly better average adjusted sensitivity (90% [95% CI, 88%-92%]) compared with the 19 studies using only FDG-PET (89% [95% CI, 84%-92%]) or the studies combining FDG-PET with another method of imaging (82% [95% CI, 75%-89%]). Scanner type had little association with specificity. The average adjusted specificity for studies that used a FDG-PET or FDG-PET/CT in combination with another imaging modality (75%) was similar to studies using only PET/CT (76%).

PET not combined with CT or with other imaging modalities had a slightly worse average adjusted specificity of 70% (95% CI, 62%-77%) compared with FDG-PET/CT or PET/CT plus another imaging modality. Among the other imaging modalities reported, 2 studies used single-photon emission CT as the alternative secondary scanning modality.

Study Quality
Study quality scores ranged from 3 to 14 (median score was 10 of 15 possible points). The quality metric that most studies failed to meet was patients receiving the same reference standard regardless of index test result (67%). Studies often lacked sufficient numbers of benign cases (29 studies had < 25 benign cases). Most studies (81%) emulated the use of the FDG-PET scan in current clinical practice.

Lower-quality studies had reduced average adjusted sensitivity (87% [95% CI, 85%-90%]) compared with higher-quality studies (91% [95% CI, 88%-93%]) after controlling for other study characteristics in the regression model. Average adjusted specificity was similar across lower-quality studies (75%) compared with higher-quality studies.

*References 14, 24, 26-29, 32, 39, 45, 46, 48, 50, 52-54, 58, 60, 62, 64, 67-69, 72, 74, 75, 77, 80-83, 85-88
†References 10, 11, 15, 25, 30, 31, 34, 35, 37, 41, 43, 44, 55-57, 61, 63, 73, 76, 78, 79, 84, 86
‡References 11, 15, 25, 30, 32-37, 39, 40, 47, 49-51, 54, 57, 75, 90
§References 38, 41-43, 53, 62, 75, 79, 80, 87

Figure 3. Performance of FDG-PET to Diagnose Lung Nodules by Endemic Status for 70 Studies

The operating points for endemic and nonendemic infectious lung disease studies and 95% confidence and prediction intervals for those operating points are shown. The horizontal box and whiskers plot represents the distribution of study specificity, and the vertical box and whiskers plot represents the distribution of study sensitivity. The box limits are the closest data point to the interquartile range of 25% and 75% with the bar being the median (50%). Error bar whiskers represent the data point closest to 1.5 times the interquartile range and the dots outside the whiskers represent outlier study values. FDG indicates fluodeoxyglucose F18; HSROC, hierarchical summary receiver operator curve; PET, positron emission tomography.
(74%). Studies that relied on either pathological diagnosis or less than 1 year of follow-up had similar average adjusted sensitivity (87% [95% CI, 83%-90%]) compared with those that did not (90% [95% CI, 88%-92%]). The average adjusted specificity among studies that relied on a combination of prolonged surveillance and pathological diagnosis had higher average adjusted specificity (77% [95% CI, 73%-81%]) than those that exhibited possible verification bias (69% [95% CI, 61%-75%]). Additional study quality results are provided in the eResults in the Supplement.

Sensitivity Analysis for the Effect of Individual Studies on Pooled Estimates

Removal of the largest study by Bryant and Cerfolio (n = 585), which reported endemic infectious lung disease, reduced the average adjusted specificity from 61% to 56% in studies with endemic infectious lung disease. Its removal had little influence on the sensitivity of the test in either average adjusted results or in the endemic infectious lung disease populations. A sensitivity analysis using the distance method by Cook identified 1 study (García Vicente et al80) as potentially overly influential. However, its exclusion did not noticeably change results. No individual study unduly influenced the estimated sensitivity or specificity of FDG-PET.

Discussion

For the last decade, molecular imaging with FDG-PET has become part of the diagnostic arsenal of tests considered for the evaluation of suspicious lung nodules. This method of imaging is suggested based on low-quality evidence (grade 2C) for the diagnosis of solid nodules larger than 8 mm.2 The limitation of FDG-PET in the diagnosis of smaller lesions is well documented and this meta-analysis also found studies reporting lesions smaller than 2 cm had lower sensitivity compared with studies reporting on larger nodules.3,91,92 Previous meta-analyses found FDG-PET to be highly sensitive (94% to 96%) and reasonably specific (78% to 86%) in the diagnosis of lung cancer.5,6 Compared with prior studies, the sensitivity and specificity in our meta-analysis was lower. The HSROC was 0.9, which is similar to that reported by Gould et al6 and our study also exhibited heterogeneity across studies.

In the 2001 meta-analysis,6 727 of the 1474 lesions (49%) were from Japanese or European populations. Also, a portion of the studies in the meta-analysis were populations from the northeast or other areas of the United States where granulomatous disease is less common. Similarly in the study by Cronin et al,8 860 of the 1190 lesions (72%) reported in the 22 studies reviewed were from geographic areas where infectious lung disease is rare.

In regions where infectious lung disease is highly prevalent, the specificity of FDG-PET scans to diagnose lung nodules suspicious for lung cancer in our study was approximately 61%. However, the best specificity in endemic regions (from either the average adjusted or adjusted results) was 66%.

Therefore, in individuals being evaluated for a suspicious lung lesion, and who reside in a region with significant endemic infectious lung disease, FDG-PET/CT does not reliably distinguish benign disease from lung cancer.

We have shown that the specificity of FDG-PET/CT for the diagnosis of lung cancer was overstated in regions with endemic infectious lung disease and could lead to unnecessary biopsies or thoracotomies for indeterminate lung nodules. Knowledge of this limitation in such regions is especially important if low-dose CT screening for lung cancer is widely adopted and should be reflected in current nodule management guidelines.5,3

Our review included more studies and had greater heterogeneity in both sensitivity and specificity compared with the earlier meta-analyses by Gould et al6 and Cronin et al.3 Some heterogeneity across studies arises from the scanning method, the size of the lesion examined, whether the study relied only on pathological verification of cancer, and the prevalence of endemic infectious lung disease in the study population. However, there remained substantial variability among studies in test performance that was not accounted for by these factors.

We observed a transition in the literature from scanners using FDG-PET only to FDG-PET/CT since their introduction into clinical practice in 2001.93,94 Recently, radiologists have undertaken significant efforts to find a complement or replacement for the PET radionuclide or the positron emission image-generating scanner.42,95-98

We attempted to include the breadth of research in PET for lung nodule diagnosis by searching for studies that compared new modalities or radionuclides with existing FDG-PET or PET plus CT. Multimodality studies collectively had a higher specificity (80%) compared with studies using either FDG-PET or CT alone, but as a group they may be susceptible to publication bias that potentially decreases the accuracy of FDG-PET compared with the newer imaging methods.

To date, no replacement for FDG has been suggested for the diagnosis of lung nodules suspicious for lung cancer.93-99 In addition, a majority of participants (n = 4615) were in studies in which the mean or median lesion size was less than or equal to 2 cm. The lower sensitivity observed in this analysis arises, in part, from the application of this diagnostic modality to a broader clinical population with both smaller lesions and a greater likelihood of infectious disease.

After adjusting for study characteristics in our model, the precision of estimated sensitivity and specificity is quite good (shown by the narrow 95% CIs in Figure 3). However, variability remains even after adjusting for known study characteristics as shown in the distribution of individual study sensitivity and specificity estimates and combined PIs.

The range of test performance observed in practice is quite large (shown by the wide PIs in Figure 3). These reflect a lack of consistency in the application of FDG-PET diagnosis for lung nodules that is concerning and this meta-analysis suggests significant variability in practice patterns. Accordingly, technical standards and consistent adherence
to imaging protocols and image interpretation should be strictly followed to reduce these inconsistencies. This is especially important in smaller lesions (<2 cm) and in regions with endemic infectious lung disease to prevent false-positive and negative test results that could cause harm to patients.

The limitations of this analysis are those common to meta-analyses (eg, publication bias, selection bias, limited information from study reports, and potential for ecological fallacy). Even though we did not find evidence of significant publication bias, this does not exclude its possibility. Because FDG-PET was recommended for the diagnosis of lung cancer, a publication bias to report poor FDG-PET accuracy or negative results may exist.

Studies reporting results from scanners with FDG-PET only may no longer reflect clinical practice and arguably should not be included in this analysis. However, we controlled for the shortcomings of scanners with FDG-PET only in the regression model so that additional studies reporting results from smaller lesions and from regions with endemic infectious lung disease could be explored.

Although the accuracy of FDG-PET/CT is superior to the accuracy of FDG-PET only, we included both modalities because they are still in use in the United States and elsewhere, and, as previously stated, we did not find a significant difference in specificity based on these 2 scanner types. To avoid selection bias, this meta-analysis attempted to broadly review studies reporting use of FDG-PET to characterize lung nodules and examined studies in which FDG-PET/CT was compared with other imaging modalities for the diagnosis and staging of lung cancer. We controlled for study heterogeneity using a random-effects regression model with a number of clinically important covariates; however, residual confounding may still be present.

In this large meta-analysis, the observed association between lower specificity and endemic infectious lung disease appeared robust across sensitivity analyses. We found that studies that fully used the metabolic and anatomic information from a FDG-PET/CT scan in a semiquantitative interpretation (rather than a simplified dichotomizing of a standard uptake value) demonstrated improved test accuracy. Even in regions of endemic disease, robust reading methods by experienced readers generated accurate scans.26,27

Until this expertise and method is more uniformly applied among scan readers, FDG-PET for the diagnosis of lung cancer in patients who reside in a region with significant endemic infectious lung disease should be recognized as having lower specificity (approximately 61%) than previously reported. Knowledge of this reduction in specificity should limit the use of FDG-PET to diagnose lung cancer unless substantial institutional expertise in FDG-PET interpretation has been proven. Should low-dose CT screening for lung cancer become the diagnostic standard, knowledge of FDG-PET/CT performance is even more critical because the vast majority of indeterminate lung nodules detected through screening are benign.100

Conclusions

The accuracy of FDG-PET for diagnosing lung nodules was extremely heterogeneous. Use of FDG-PET/CT was less specific in diagnosing malignancy in populations with endemic infectious lung disease compared with nonendemic regions. These data do not support the use of FDG-PET to diagnose lung cancer in endemic regions unless an institution achieves test performance accuracy similar to that found in nonendemic regions.

ARTICLE INFORMATION

Author Affiliations: Veterans Affairs Hospital, Tennessee Valley Healthcare System, Nashville (Deppen, Massion, Grogan); Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, Tennessee (Deppen, Aldrich, Putnam, Grogan); Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee (Blume); Department of Surgery, Vanderbilt University Medical Center, Nashville, Tennessee (Kensinger); School of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Morgan); Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Aldrich, McPheeters); Division of Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Massion); Department of Medical Imaging, Tennessee Valley Healthcare System-Veterans Affairs, Nashville (Walker); Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, Tennessee (Aldrich, McPheeters); Division of Obstetrics and Gynecology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (McPheeters).

Author Contributions: Drs Deppen and Grogan had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Deppen, Aldrich, Massion, Putnam, Grogan. Acquisition, analysis, or interpretation of data: Deppen, Blume, Kensinger, Morgan, Massion, Walker, McPheeters, Putnam. Drafting of the manuscript: Deppen, Blume, Walker. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: Deppen, Blume. Obtained funding: Deppen, Putnam, Grogan. Administrative, technical, or material support: Deppen, Kensinger, Walker. Study supervision: Blume, Aldrich, Massion, Walker, McPheeters, Putnam, Grogan.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: This work was supported by grant R03HS021554-01 from the Agency for Healthcare Research and Quality (Dr Grogan). This work was also supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service Career Development award 10-024 (Dr Grogan), grant K07CA172294 from the National Institutes of Health and the National Cancer Institute (Dr Aldrich), Vanderbilt Institute for Clinical and Translational Research grant UL1TR000011 from the National Center for Advancing Translational Sciences at the National Institutes of Health (REDCap database), the lung SPORE grant CA90949 from the National Cancer Institute (Dr Massion), and grant U01CA152662 from the Early Detection Research Network (an initiative of the National Cancer Institute) (Dr Massion).

Role of the Funder/Sponsor: The funders/sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank Rebecca Jerome, MLS, MPH (research librarian at Vanderbilt University Medical Center), for her uncompensated assistance in the literature review.

Correction: This article was corrected on September 30, 2014, to fix 16% to 16 percentage points for specificity in regions with endemic disease in the abstract and Results section.

Downloaded From: https://jama.jamanetwork.com/ by a Non-Human Traffic (NHT) User on 03/29/2019
REFERENCES


