LUNG CANCER

Accuracy of transbronchial needle aspiration for mediastinal staging of non-small cell lung cancer: a meta-analysis

J-E C Holty, W G Kuschner, M K Gould

Thorax 2005;**60**:949–955. doi: 10.1136/thx.2005.041525

See end of article for authors' affiliations

Correspondence to: Dr J-E C Holty, Center for Primary Care and Outcomes Research, Stanford University, 117 Encina Commons, Stanford, CA 94305-6019, USA; jholty@stanford.edu

Received 27 January 2005
Accepted 8 June 2005
Published Online First 30 June 2005

Background: The reported accuracy of transbronchial needle aspiration (TBNA) for mediastinal staging in non-small cell lung cancer (NSCLC) varies widely. We performed a meta-analysis to estimate the accuracy of TBNA for mediastinal staging in NSCLC.

Methods: Medline, Embase, and the bibliographies of retrieved articles were searched for studies evaluating TBNA accuracy with no language restriction. Meta-analytical methods were used to construct summary receiver-operating characteristic curves and to pool sensitivity and specificity.

Results: Thirteen studies met inclusion criteria, including six studies that surgically confirmed all TBNA results and enrolled at least 10 patients with and without mediastinal metastasis (tier 1). Methodological quality varied but did not affect diagnostic accuracy. In tier 1 studies the median prevalence of mediastinal metastasis was 34%. Using a random effects model, the pooled sensitivity and specificity were 39% (95% CI 17 to 61) and 99% (95% CI 96 to 100), respectively. Compared with tier 1 studies, the median prevalence of mediastinal metastasis (81%; p = 0.002) and pooled sensitivity (78%; 95% CI 71 to 84; p = 0.009) were higher in non-tier 1 studies. Sensitivity analysis confirmed that the sensitivity of TBNA depends critically on the prevalence of mediastinal metastasis. The pooled major complication rate was 0.3% (95% CI 0.01 to 4).

Conclusions: When properly performed, TBNA is highly specific for identifying mediastinal metastasis in patients with NSCLC, but sensitivity depends critically on the study methods and patient population. In populations with a lower prevalence of mediastinal metastasis, the sensitivity of TBNA is much lower than reported in recent lung cancer guidelines.

Non-small cell lung cancer (NSCLC) is the most common malignancy in the world and accounts for an estimated 1 million deaths each year.[1] The overall 5 year survival is approximately 15%.[2] However, the survival rate approaches 70% in some patients with resectable disease.[3] Metastasis to the mediastinal lymph nodes is one of the most important factors in determining resectability and prognosis.[4] Careful mediastinal staging is essential to identify appropriate candidates for surgery and to avoid futile thoracotomy in patients with more advanced disease.

Currently, computed tomography (CT) is the most frequently used preoperative staging modality. However, large benign hyperplastic lymph nodes are commonly found in patients with NSCLC[5] and normal sized lymph nodes frequently harbour metastases.[6] Preoperative clinical staging with CT differs from surgical staging in as many as 40% of cases.[7 8] Newer imaging modalities such as positron emission tomography (PET) have limitations in diagnostic accuracy as well.[9] Given the limitations of CT and PET, invasive surgical staging techniques such as mediastinoscopy are often used to exclude or confirm mediastinal lymph node metastasis, especially in patients who are candidates for surgical resection. However, mediastinoscopy is associated with a complication rate of 2–3% and a surgical mortality rate of around 0.1%.[10–12]

Transbronchial needle aspiration (TBNA) using a flexible fibreoptic bronchoscope was developed in the early 1980s to obviate the need for more risky surgical staging procedures. Compared with mediastinoscopy, TBNA is generally more convenient, less risky, and less expensive.[13] A recent systematic review of mediastinal staging with TBNA found the sensitivity to be similar to that of mediastinoscopy (76% *v* 81%).[14] This analysis, however, was not restricted to patients with NSCLC, did not assess study methodological quality, and did not attempt to identify sources of variation in study results. We performed a meta-analysis to estimate the diagnostic accuracy of TBNA in patients with NSCLC and to identify technical factors and patient characteristics that have an impact on accuracy.

METHODS

A more detailed description of our methods is available as an online data supplement on the *Thorax* website at www.thoraxjnl.com/supplemental.

Literature search and identification of studies

Medline and Embase (January 1966 to July 2003; Medline updated through April 2004) were searched to identify studies that examined TBNA for mediastinal staging in NSCLC (fig S1A and B, online supplement), and reference lists of included studies and review articles were manually searched. All articles were considered, regardless of language.

Selection of studies

We included studies that (1) examined TBNA using a flexible bronchoscope for mediastinal staging in patients with NSCLC; (2) enrolled at least 10 subjects with and/or 10 subjects without mediastinal metastasis; (3) provided sufficient data to permit calculation of sensitivity and/or specificity; and (4) enrolled no more than 10% of patients with a diagnosis other than NSCLC or provided separate data

Abbreviations: NSCLC, non-small cell lung cancer; TBNA, transbronchial needle aspiration

for patients with NSCLC. More rigorous (tier 1) studies enrolled at least 10 subjects with and 10 subjects without mediastinal lymph node involvement, surgically confirmed all TBNA results (for example, with mediastonoscopy, mediastonomy and/or thoracotomy), and used the patient as the unit of analysis. The authors of abstracts and studies not reporting sufficient data were contacted to request additional information.

Study quality

An existing instrument was adapted to describe the methodological quality of studies,[15] as reported previously (fig S2, online supplement).[9 16] We developed criteria for the technical quality of TBNA based on our clinical experience and by reviewing published guidelines.[17–19]

Data abstraction

One investigator abstracted primary data regarding patient characteristics and the sensitivity and/or specificity of TBNA for identifying mediastinal metastasis in patients with NSCLC.

When possible, we separated staging characteristics of TBNA for patients with and without enlarged lymph nodes on the CT scan and for biopsies performed at hilar, subcarinal, paratracheal, or other lymph node stations. We also separately tabulated test characteristics for studies using ''real time'' imaging—for example, CT fluoroscopy, endobronchial ultrasound, or transthoracic ultrasound.

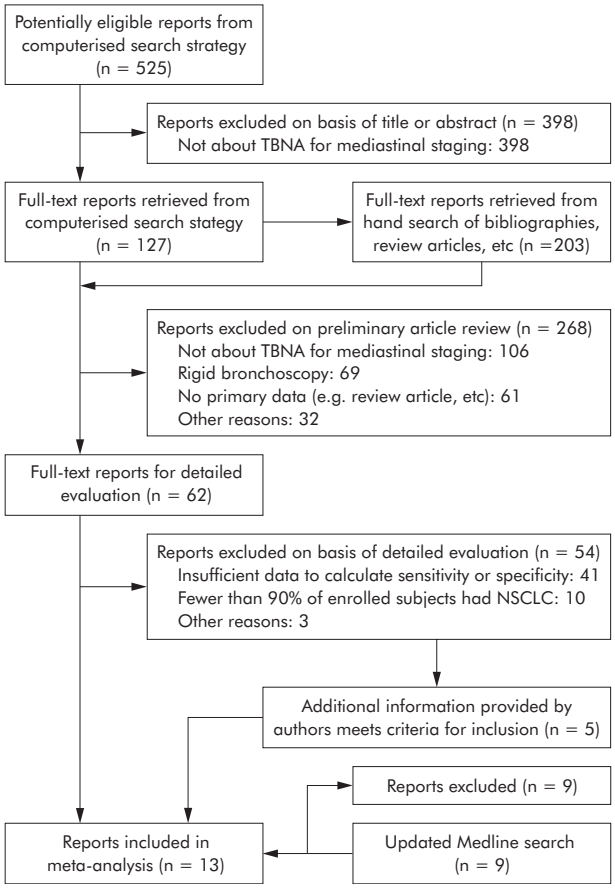

Figure 1 Literature search and selection. Studies could meet one or more exclusion criteria. For simplicity, only one primary exclusion criterion per study is shown.

Data synthesis and statistical analysis

We constructed a 2×2 contingency table for each study to summarise the results of TBNA and the reference test(s). For each study we calculated the true positive rate (TPR; sensitivity), the false positive rate (FPR; 1−specificity), and the log odds ratio (LOR). When necessary, we added 0.5 as a correction factor to calculate the LOR.

Because many studies of TBNA did not confirm positive test results surgically, they were unable to report false positive rates. We therefore calculated a weighted kappa-1 coefficient which does not require information about the false positive rate to measure test accuracy with respect to avoiding false negative results.[20 21]

A random effects model was used to pool sensitivity, specificity, LOR and kappa-1.[22] When pooling sensitivity and specificity, we excluded studies with <10 subjects with or without mediastinal lymph node involvement, respectively, in the calculations. Summary receiver operating characteristic (SROC) curves as described by Moses *et al*[23] were constructed to summarise the results quantitatively.

To assess sources of variation in study results we performed sensitivity analyses, discriminate function analyses, and meta-regressions. Sensitivity analysis included stepwise single study elimination, adjusting the correction factor, and varying the reference test result in studies that employed a suboptimal reference standard. To compare sensitivity and specificity jointly in studies grouped by tier and prevalence we used discriminant function analysis. Multivariate analysis of variance (ANOVA)[24 25] was used to compare reported sensitivities and LORs in studies with high and low prevalences of lymph node metastasis (⩾60% or <60%) and year of study publication (⩾1995 or <1995). To assess for the presence of publication bias we constructed inverted funnel plots of standard error versus estimated effect size (LOR) for each individual study.[26] We also assessed how the exclusion of small cell cancer cases from the included studies impacted on the accuracy of TBNA.

All biostatistical models were programmed with Excel 8.0 for Windows (Microsoft Corporation, Redmond, Washington, USA). Discriminant function analysis was performed using SAS 9.0 for Windows (SAS Corp, Cary, NC, USA). We calculated 95% confidence intervals (CIs) for the TPR and the FPR by using the quadratic method.[27] A normal approximation to the binomial of the standard error was used in calculating all other confidence intervals, as appropriate. When making comparisons between groups of studies, an unpaired *t* test or the Mann-Whitney U test was used, as appropriate. A two tailed p value of <0.05 was considered statistically significant.

RESULTS

Literature search and study selection

Our literature search identified 525 potentially eligible studies (fig 1); 398 studies judged not to be relevant after carefully reviewing their titles and abstracts were eliminated. A hand search of the bibliographies of the remaining 127 articles identified 203 additional studies that were potentially relevant. A preliminary review of these 330 articles eliminated 268 studies, leaving 62 articles for detailed analysis (table S1, online supplement). After detailed review, 13 studies met the inclusion criteria (table 1).[28–40] Studies were most often excluded because they provided insufficient data to calculate sensitivity or specificity (76%) or enrolled more than 10% of subjects with a diagnosis other than NSCLC (60%). Inter-rater agreement for study inclusion was high (mean kappa ~80%; table S2, online supplement). Five authors provided additional information that enabled us to include their studies.[28 29 32 33 37]

Table 1 Characteristics of studies in the meta-analysis*

Study (reference)	Year	Prevalence§	NSCLC (%)	TBNA results‡‡ (%) TP	FN	FP	TN	Inclusion criteria and comments
Tier 1 studies†								
Harrow *et al*[28]‡	2000	34	100	8	17	1	48	Patients with suspected lung cancer were included. Patients without lung cancer, or TBNA from a lymph node confluent with the tumour mass were excluded by the authors.
Bilaceroglu *et al*[29]‡	1998	60	100	24	9	0	22	Patients with potentially resectable extrabronchial or endobronchial mass suggestive of lung cancer and without extrathoracic metastases were included. We excluded patients with N0 or N1 disease on pre-TBNA CT.¶
Disdier *et al*[30]	1998	52	100	5	9	1	12	Patients with potentially resectable lung cancer without evidence of extrathoracic metastases and with enlarged mediastinal LAD on CT or CXR were included.
Ratto *et al*[31]	1988	30	96	2	12	0	33	Patients with potentially resectable lung cancer (no preoperative SCLC) without evidence of extrathoracic metastases were included. Only subcarinal lymph nodes were biopsied.
Schenk *et al*[32]‡	1986	29	100	8	13	2	50	Patients with CXR evidence of lung cancer and who were potentially resectable without evidence of extrathoracic metastases were included. One FP had scanty neoplastic cells without lymphocytes.
Studies not meeting tier 1 criteria								
Herth *et al*[33]‡ **	2002	82	100	30	6	0	8	Patients with central lung cancer and enlarged mediastinal LAD by CT without evidence of extrathoracic metastases were included. Needle placement via endobronchial ultrasound.
Wang *et al*[34]	2002	100	93††	10	0	0	0	Inclusion criteria not specified. Unable to calculate specificity (100% prevalence).
Patelli *et al*[35]	2002	NA	100	127	52	–	15	Patients with NSCLC, N2 disease by CT scanning and a negative bronchoscopy for lung cancer were included. Statistical analysis was by TBNA specimen. Data were collected by retrospective chart review. Non-surgically confirmed negative TBNA specimens (total 49) were assumed to be false negative.
Katis *et al*[36]	1998	95	100	28	8	–	2	Patients with CXR evidence of lung cancer and enlarged mediastinal LAD on CT and who were potentially resectable and without evidence of extrathoracic metastases were included.
Rong *et al*[37]‡ **	1998	79	100	26	2	0	5	Patients undergoing thoracotomy for mediastinal adenopathy on CT and suspected lung cancer were included. Bronchoscopist was not experienced with TBNA. Real-time CT assisted needle placement.
Schenk *et al*[38]	1993	81	100	32	6	–	9	Patients with resectable lung cancer and paratracheal LAD without extrathoracic metastases were included. Four of the 32 TPs were confirmed surgically.
Schenk *et al*[39]	1989	81	100	14	3	–	4	Patients with CXR evidence of lung cancer and who were potentially resectable and without evidence of extrathoracic metastases were included.
Wang *et al*[40]	1983	55	100	13	3	–	13	Patients with suspected lung cancer and who were potentially resectable and without evidence of extrathoracic metastases were included.

TBNA, transbronchial needle aspiration; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; LAD, lymphadenopathy; CT, chest computed tomography; CXR: chest radiograph; NA, not applicable; TP, true positive; FN, false negative; FP, false positive; TN, true negative; –, studies in which positive TBNA results were not confirmed surgically but were assumed to be true positives.
*Five studies[30 31 33 34 36] reported age (median 60 years) and six studies [30 31 33–36] reported sex characteristics (median proportion male 91%).
†Studies meeting tier 1 criteria surgically confirmed all TBNAs had at least 10 subjects with and without mediastinal lymph node metastasis and used the patient as the unit of analysis.
‡Additional information obtained from original study author(s).
§Prevalence of mediastinal lymph node metastasis. Studies that did not surgically confirm all TBNA results assumed that the false positive rate was zero.
¶All mediastinal and hilar lymph nodes (≥8 mm) identified by CT scanning in this study were sampled by TBNA. To estimate TBNA diagnostic accuracy for identifying mediastinal metastasis, patients with N0 or N1 disease on pre-TBNA CT scanning were excluded from the analysis (the combined mediastinal/hilar TBNA results for NSCLC were: TP 49, FN 18, FP 0, TN 16). The sensitivity (73% *v* 73%; p = 0.97) and specificity (100% *v* 100%; p = 1.0) were similar whether or not N0/N1 disease on CT scanning was excluded.
**Two studies used real-time radiological assistance to guide needle placement during TBNA.
††Three of 42 patients undergoing TBNA had SCLC. Thirty two patients who had hilar lymph node biopsies (not mediastinal) were excluded.
‡‡Number of patients (except for the study by Patelli *et al*,[35] where statistical analysis was by TBNA specimen).

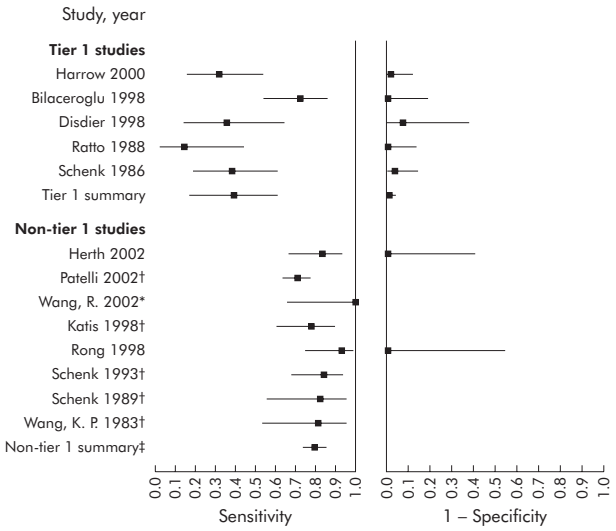

Figure 2 Individual study estimates of sensitivity and 1 − specificity of TBNA for identifying mediastinal metastasis. Error bars represent 95% confidence intervals (CIs). Point estimates and 95% CIs for tier 1 studies and studies meeting inclusion criteria but not tier 1 criteria are shown. Tier 1 and non-tier 1 summary point estimates and corresponding 95% CIs are shown and were calculated using a random effects model. *Specificity was not calculated for the study by Wang R *et al*[34] because the prevalence of mediastinal lymph node metastasis was 100%. †Specificity is not shown for studies that did not surgically confirm all TBNA results but instead assumed that all positive results were true positives. ‡Summary non-tier 1 −specificity not shown because only two studies allowed calculation of specificity (both 100%),[33 37] although both of these studies had fewer than 10 patients without mediastinal metastases.

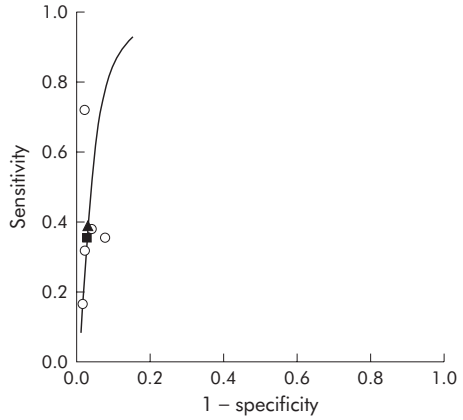

Figure 3 Summary receiver-operating characteristic curve for mediastinal staging with TBNA for tier 1 studies. Individual study estimates of sensitivity and 1 − specificity are shown (open circles). Median (solid square) and pooled sensitivities (solid triangle) are also shown.

Study description

The median number of participants per study was 44 (range 10–183). Six studies[30 31 33 34 36 37] reported statistics about the age of participants (median age 60 years) and seven studies [30 31 33–37] reported sex characteristics (median proportion male 89%). One study reported results by using individual lymph nodes as the unit of analysis.[35] For the other studies that reported results by using the patient as the unit of analysis, the median prevalence of mediastinal metastasis was 70% (interquartile range 47–83). The size and type of TBNA needle used and the number of aspirate passes per lymph node station varied between studies (table S3, online supplement). None of the studies stratified results according to nodal station or lymph node size on the CT scan in patients with NSCLC. In eight studies all positive and negative TBNA results were confirmed by mediastinoscopy, mediastinotomy, or thoracotomy.[28–34 37] Six studies enrolled fewer than 10 subjects without mediastinal lymph node involvement.[33 34 36–39] Two studies[33 37] used real-time imaging (CT or endobronchial ultrasound) to guide needle placement during TBNA. Five studies met criteria for tier 1 analysis.[28–32]

Study quality

Studies met between 12 and 23 of the 34 prespecified criteria for methodological quality. Seven studies met at least 50% of the criteria.[28–30 32 33 36 38] Table S4 (online supplement) shows selected aspects of methodological quality for each study. In general, tier 1 studies met more criteria (mean 18.8; 95% CI 15.8 to 21.8) than non-tier 1 studies (mean 15.8; 95% CI 14.0 to 17.5), but this difference was not statistically significant ($p = 0.13$).

Diagnostic accuracy of TBNA

Tier 1 analysis (5 studies)

In these studies the median prevalence of mediastinal metastasis was 34% (range 29–60). The median sensitivity and specificity of TBNA were 36% (interquartile range 32–38) and 98% (interquartile range 96–100), respectively (table 2). The pooled (random effects) sensitivity was 39% (95% CI 17 to 61) and the pooled specificity was 99% (95% CI 96 to 100) (table 2, fig 2). The corresponding positive and negative likelihood ratios were 29.0 and 0.62, respectively. The summary ROC curve is shown in fig 3.

The pooled (random effects) kappa-1 coefficient was 30% (95% CI 15 to 46), suggesting that the accuracy of TBNA with respect to false negative results was poor to fair in tier 1 studies.

Non-tier 1 analysis (8 studies)

Two non-tier 1 studies used real-time radiological needle guidance during TBNA.[33 37] In the remaining six studies the median prevalence of mediastinal metastasis was 81% (range 55–100; $p = 0.002$ for comparison with tier 1 studies). None of these six studies provided sufficient information to calculate specificity (for example, they did not surgically confirm positive TBNA results). The median sensitivity of TBNA in studies not using real-time radiological needle guidance was 82% (interquartile range 79–84; table 2). The pooled (random effects) sensitivity was 78% (95% CI 71 to 84; table 2, fig 2). The pooled kappa-1 coefficient (random effects) was 40% (95% CI 19 to 62; table 2), suggesting that the accuracy of TBNA with respect to false negative results was fair in non-tier 1 studies.

The median prevalence of mediastinal metastasis in the two non-tier 1 studies that used real-time radiological needle guidance was 83% ($p = 0.84$ for comparison with the six other non-tier 1 studies). The pooled (85%) and median sensitivities (88%) in these two studies were not significantly different ($p = 0.36$ and $p = 0.38$, respectively) from the pooled and median sensitivities of the six non-tier 1 studies that did not use real-time radiological guidance.

Summary analysis (11 studies)

The Q statistic from the random effects model showed that there was statistically significant heterogeneity in sensitivity ($p<0.001$) but not in specificity ($p = 0.90$). Discriminant function analysis confirmed that there was a statistically significant difference in the joint sensitivity and specificity of tier 1 and non-tier 1 studies ($p = 0.002$, parametric Wilks' lambda test; fig 4). We therefore did not pool the results of tier 1 and non-tier 1 studies.

Table 2 Summary of results*

	Median sensitivity (IQR)	Median specificity (IQR)¶	Pooled sensitivity (95% CI)†	Pooled specificity (95% CI)† ¶	Likelihood ratio (95% CI)‡ ¶	
					Positive	Negative
Tier 1 studies	0.36 (0.32–0.38)	0.98 (0.96–1.00)	0.39 (0.17 to 0.61)	0.99 (0.96 to 1.00)	29.0 (21.9 to 38.4)	0.62 (0.53 to 0.72)
Non-tier 1 studies	0.82 (0.79–0.84)	–	0.78 (0.71 to 0.84)	–	–	–
p value§	0.001	–	0.009	–	–	–

IQR, interquartile range.
*For studies that did not use real-time radiological guidance for needle placement.
†Pooled sensitivity and specificity were calculated using a random effects model.
‡Positive and negative likelihood ratios were calculated from the pooled sensitivity and specificity.
§p value for the comparison between tier 1 and non-tier 1 studies.
¶Median and pooled specificities for non-tier 1 studies were not calculated because only two studies allowed calculation of specificity (both 100%);[33 37] however, both of these studies had fewer than 10 patients without mediastinal metastases. We were thus unable to calculate likelihood ratios for non-tier 1 studies.

Complication rate

One study did not report complications.[30] Of the remaining studies, two reported major complications,[28 40] including two major bleeds and one pneumothorax requiring a chest tube. Two other cases of pneumothoraces[35] and one case of pneumomediastinum[28] spontaneously resolved under observation. The mean rate of major complications per patient in tier 1 and non-tier 1 studies was 0.32% (95% CI 0.01 to 6) and 0.25% (95% CI 0.01 to 6), respectively (p = 0.65). The overall major complication rate was 0.26% (95% CI 0.01 to 4).

Sensitivity analysis and meta-regressions

An inverted funnel plot showed no evidence of publication bias (fig S4, online supplement). Stepwise single study elimination did not substantially affect the magnitude of the pooled LOR or sensitivity in tier 1 or non-tier 1 studies (table S5, online supplement). In one study, one of two false positive results had scanty neoplastic cells and no lymphocytes.[32] Re-categorising this result as a true negative had no effect on pooled sensitivity, specificity, LOR, or the kappa-1 coefficient. Varying the correction factor from 0.5 to 0.1 had no impact on the LOR or the kappa-1 coefficient. Using a 0.1 correction tended to shift the summary ROC curve to the left (increasing specificity), but had little discernable impact on sensitivity.

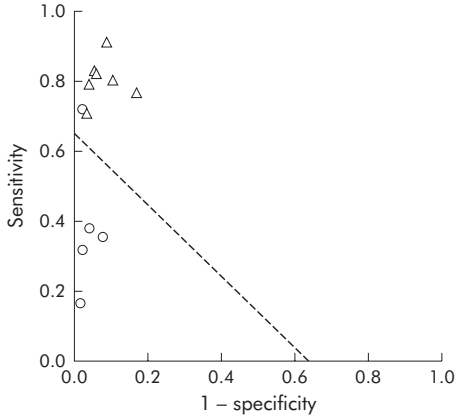

Figure 4 Discriminant function analysis for mediastinal staging with TBNA. Individual study estimates of sensitivity and 1 − specificity are shown for tier 1 studies (circles) and all other studies meeting inclusion criteria (triangles). The one study with 100% prevalence of mediastinal lymph node metastasis was not included (that is, specificity undefined).[34] The discriminant function that separates tier 1 from all other included studies is shown by the broken line. This confirms that differences in diagnostic accuracy between tier 1 and non-tier 1 studies are statistically significant when sensitivity and specificity are considered jointly (p = 0.002, parametric Wilks' lambda test). The one tier 1 study[29] that fell on the ''wrong'' side of the line had a relatively high prevalence (60%) of mediastinal metastasis.

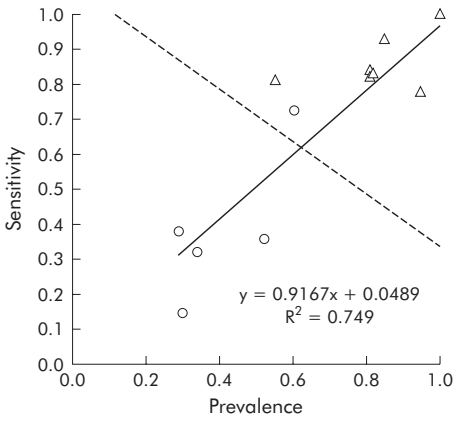

Figure 5 Sensitivity as a function of the prevalence of mediastinal lymph node metastasis. Individual study estimates of prevalence and sensitivity are shown for tier 1 (circles) and non-tier 1 (triangles) studies. Both the linear regression equation and R^2 are shown. One study reported results by using individual lymph nodes as the unit of analysis and was not included (unable to calculate prevalence of mediastinal metastasis).[35] The discriminant function that separates tier 1 from non-tier 1 studies (broken line) was significant (p = 0.002, parametric Wilks' lambda test).

Study sensitivity was positively correlated with the prevalence of lymph node metastasis (fig 5). When the prevalence rose from 40% to 80%, sensitivity increased from 42% to 78%. For the seven studies in which the prevalence of mediastinal disease was ⩾60%, the median sensitivity (83% *v* 36%; p = 0.005) and pooled sensitivity (84% *v* 40%; p = 0.005) were higher than the five remaining studies in which prevalence was <60% (fig S3, online supplement). Discriminant function analysis confirmed that the joint sensitivity and specificity were different in studies with high versus low prevalence (p = 0.01, parametric Wilks' lambda test).

For the eight studies published since 1995, the pooled sensitivity (71% *v* 60%; p = 0.52) was not significantly different from the five remaining studies published before 1995. However, the median prevalence of lymph node metastasis in more recent studies (82% *v* 55%; p = 0.09) was higher than in the five earlier studies.

These and other potential sources of heterogeneity were assessed by a multivariate ANOVA to compare reported sensitivities and LORs in studies with respect to the prevalence of lymph node metastasis (⩾60% or <60%) and year of publication (⩾1995 or <1995). Because only two included studies used real-time radiological needle guidance, we were unable to assess this potential source of heterogeneity and excluded these two studies from the analysis. Sensitivity was higher in studies with a higher prevalence of

lymph node metastasis (difference 60%; 95% CI 51 to 69) and in more recently published studies (difference 10%; 95% CI 1 to 18). The prevalence of lymph node metastasis, but not year of publication, had a significant effect on the LOR.

Excluding patients with small cell lung cancer from the included studies had no impact on the pooled sensitivity in tier 1 (39% *v* 41%, p = 0.92) or non-tier 1 (78% *v* 80%, p = 0.71) studies.

DISCUSSION

TBNA is highly specific for identifying mediastinal metastasis in patients with NSCLC, but sensitivity depends critically on the prevalence of mediastinal disease. Specificity is excellent, but not perfect. In three of eight studies that surgically confirmed all TBNA results, four false positive results were reported. One of the four false positive results would have been avoided if biopsy specimens were considered negative when they lacked nodal tissue or when the cytopathologist identified the specimen as ''contaminated'' or containing ''atypical'' cells. It is essential to avoid contamination of the bronchoscope channel and to follow stringent criteria to define positive or negative biopsy specimens in order to minimise the risk of false positive TBNA results. We found that TBNA is generally safe with a major complication rate of approximately 0.3%.

We identified several sources of variation in study results. Sensitivity was much lower in tier 1 studies than non-tier 1 studies. Tier 1 studies surgically confirmed all TBNA results, enrolled at least 10 patients with and without mediastinal metastasis, and used the patient as the unit of analysis. Sensitivity was also lower in studies with a low prevalence (<60%) of mediastinal metastasis. Not surprisingly, TBNA appears to be less sensitive than mediastinoscopy for identifying mediastinal metastasis. A recent meta-analysis of 14 studies of mediastinoscopy reported a pooled sensitivity of 81% (95% CI 76 to 85).[14] In these studies the pooled prevalence of mediastinal disease was 37%, which is similar to the median prevalence (34%) of lymph node metastasis in tier 1 studies of TBNA.

The difference in diagnostic accuracy between tier 1 and non-tier 1 studies was statistically significant. We believe that this difference is probably related to a lower prevalence of mediastinal metastasis in tier 1 than in non-tier 1 studies. Higher disease prevalence and enrolment of patients with a more severe spectrum of disease are sources of variation in studies of diagnostic accuracy leading to an increase in sensitivity.[41 42] We speculate that the higher prevalence of mediastinal metastasis in non-tier 1 studies may reflect enrolment of study cohorts with a more severe spectrum of mediastinal disease, resulting in more positive TBNA results. For example, non-tier 1 (high prevalence) studies may have enrolled a greater number of patients with bulky lymphadenopathy in whom TBNA was being used to confirm the diagnosis of unresectable disease. In contrast, tier 1 (lower prevalence) studies may have enrolled potential surgical candidates with less impressive lymph node enlargement. A recent meta-analysis of 39 studies comparing PET with CT scanning for mediastinal staging in NSCLC found that the median prevalence of malignant lymph nodes in enrolled patients was 32% (range 5–64), which is similar to the median prevalence of mediastinal metastasis in the tier 1 studies in our analysis.[9] Most of the studies of PET and CT scanning enrolled patients with potentially resectable NSCLC. Furthermore, the bronchoscopist's technique may be more proficient when the pretest probability of obtaining a positive result is high (higher prevalence of mediastinal disease within the study cohort). For example, more diligence may be taken to identify endobronchial landmarks, more TBNA needle passes attempted, and more aggressive sedation given to minimise cough and patient movement during the procedure.

The difference in pooled sensitivities between tier 1 and non-tier 1 studies may also be the result of methodological differences. Non-tier 1 studies used suboptimal methodological criteria by not confirming all TBNA results against a reference standard (verification bias), having insufficient numbers of participants with and without mediastinal metastasis, and/or not using the patient as the unit of analysis. Verification bias has been shown to lead to overestimates of test sensitivity.[41]

A previous meta-analysis showed that the pooled sensitivity of 12 studies analysing TBNA in patients with either small-cell lung cancer or NSCLC was 76%.[14] Our estimates of sensitivity were lower for tier 1 studies (39%) because several studies that were included in this previous meta-analysis did not meet the criteria for our tier 1 analysis. Interestingly, the exclusion of patients with small cell lung cancer from the studies included in our analysis did not significantly affect sensitivity.

Despite the relatively low sensitivity of TBNA in detecting mediastinal metastasis compared with other invasive staging procedures, TBNA continues to be an appropriate diagnostic test in the sampling of mediastinal lymph nodes, especially if concurrently performed with routine bronchoscopic examination for suspected lung cancer. TBNA is generally more convenient, less risky, and less expensive than other invasive staging procedures such as mediastinoscopy.[13] A formal assessment of the cost effectiveness of staging TBNA is beyond the scope of this analysis.

Although we were unable directly to assess how newer needles, use of on-site cytological analysis, and/or improved techniques may impact on TBNA accuracy, our multivariate ANOVA showed that more recent studies—which presumably used more up to date techniques and equipment—had a slightly higher sensitivity when we controlled for prevalence of mediastinal metastasis.

Our study has several limitations. Firstly, only a small number of studies met our inclusion criteria (five tier 1 and eight non-tier 1 studies). Most studies enrolled fewer than 100 participants and were performed at single centres where experience with TBNA is likely to be extensive. Large multicentre prospective studies of TBNA should be performed in consecutively enrolled patients with NSCLC. Studies should explicitly define inclusion criteria and should report separate results for patients with non-bulky and bulky lymphadenopathy. Secondly, because needle type and size, as well as the number of aspiration passes varied between studies, we were unable to control for these test characteristics. Likewise, because most studies did not report age or sex characteristics, we were unable to control for these demographic features. Thirdly, few of the included studies provided information on whether TBNA results altered patient management. Clearly, positive results on TBNA obviate the need for mediastinoscopy because specificity and positive predictive value are high. However, simple calculations based on our results indicate that, when prevalence is relatively low (~35%), approximately 85% of patients will have negative TBNA results and 25% of such results will be falsely negative. Fourthly, despite an exhaustive search, we may not have identified all studies, especially those with unpublished results. We identified one potentially relevant abstract but we were unable to obtain sufficient additional information to assess it for inclusion.[43] However, an inverted funnel plot showed no evidence of publication bias. Finally, the 13 included studies used a variety of reference tests (cervical mediastinoscopy, anterior mediastinotomy and/or thoracotomy with ipsilateral lymph node sampling), raising the possibility of differential verification

bias.[41] Because none of the reference tests has perfect sensitivity, the true sensitivity of TBNA may be even lower than our estimates. Future studies of the diagnostic accuracy of TBNA should require thoracotomy with systematic sampling of both normal and abnormal appearing lymph nodes at all accessible mediastinal stations to exclude the presence of lymph node metastasis.[44]

In conclusion, we found that TBNA is highly specific for detecting mediastinal lymph node metastasis in patients with NSCLC, but that sensitivity depends critically on the prevalence of mediastinal lymph node involvement. In patient populations with a relatively low prevalence of mediastinal disease (such as those with potentially resectable NSCLC), the sensitivity of TBNA is poor.

ACKNOWLEDGEMENTS

The authors thank Helena C Kraemer for providing statistical advice; Corinna Haberland, Annette Langer-Gould, Pavel Strnad, Kelvin Tan, and Tomasz Ziedalski for reviewing non-English language studies; and the authors of the studies included in our meta-analysis who provided additional unpublished data.

A more detailed description of the methods used in this study is available as an online supplement on the *Thorax* website at www.thoraxjnl.com/supplemental

Authors' affiliations

J-E C Holty, W G Kuschner, M K Gould, VA Palo Alto Health Care System, Palo Alto, CA and Stanford University School of Medicine, Stanford, CA, USA
J-E C Holty, M K Gould, Center for Primary Care and Outcomes Research, Stanford University, Stanford, CA, USA

Dr Holty is supported by the VA Special Fellowships Program. Dr Gould is a recipient of an Advanced Research Career Development Award from the VA Health Services Research and Development Service.

Drs Holty, Gould and Kuschner have no financial conflict of interest or competing interests to disclose.

REFERENCES

1 **Spiro SG**, Porter JC. Lung cancer - where are we today? Current advances in staging and nonsurgical treatment. *Am J Respir Crit Care Med* 2002;**166**:1166–96.
2 **Ries LAG**. Influence of extent of disease, histology and demographic factors on lung cancer survival in the SEER population-based data. *Semin Surg Oncol* 1994;**10**:21–30.
3 **Mountain CF**. Revisions in the international system for staging lung cancer. *Chest* 1997;**111**:1710–7.
4 **Patterson GA**. Surgical techniques in the diagnosis of lung cancer. *Chest* 1991;**100**:523–6.
5 **Arita T**, Matsumoto T, Kuramitsu T, *et al*. Is it possible to differentiate malignant mediastinal nodes from benign nodes by size? Reevaluation by CT, transesophageal echocardiography and nodal specimen. *Chest* 1996;**110**:1004–8.
6 **Takeshi A**, Tatsuya K, Mitsutoshi M. Bronchogenic carcinoma: incidence of metastases to normal sized lymph nodes. *Thorax* 1995;**50**:1267–9.
7 **Lahde S**, Paivansalo M, Rainio P. CT for predicting the resectability of lung cancer. A prospective study. *Acta Radiol* 1991;**32**:449–54.
8 **Dales RE**, Stark RM, Raman S. Computed tomography to stage lung cancer. Approaching a controversy using meta-analysis. *Am Rev Respir Dis* 1990;**141**:1096–101.
9 **Gould MK**, Kuschner WG, Rydzak CE, *et al*. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer. *Ann Intern Med* 2003;**139**:879–92.
10 **Rami Porta R**. Surgical exploration of the mediastinum by mediastinoscopy, parasternal mediastinotomy and remediastinoscopy: indications, technique and complications. *Ann Ital Chir* 1999;**70**:867–72.
11 **Hujala KT**, Sipila JI, Grenman R. Mediastinoscopy: its role and value today in the differential diagnosis of mediastinal pathology. *Acta Oncol* 2001;**40**:79–82.
12 **Detterbeck FC**, DeCamp MM Jr, Kohman LJ, *et al*. Lung cancer. Invasive staging: the guidelines, *Chest* 2003;**123**(1 Suppl):167–75S.
13 **Shure D**. Fiberoptic bronchoscopy: diagnostic applications. *Clin Chest Med* 1987;**8**:1–13.
14 **Toloza EM**, Harpole L, Detterbeck F, *et al*. Invasive staging of non-small cell lung cancer: a review of the current evidence. *Chest* 2003;**123**(1 Suppl): 157–66S.
15 **Kent DL**, Haynor DR, Larson EB, *et al*. Diagnosis of lumbar spinal stenosis in adults: a meta-analysis of the accuracy of CT, MR, and myelography. *AJR Am J Roentgenol* 1992;**158**:1135–44.
16 **Gould MK**, Maclean CC, Kuschner WG, *et al*. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions. A meta-analysis. *JAMA* 2001;**285**:914–24.
17 **Harkin TJ**, Wang KP. Bronchoscopic needle aspiration of mediastinal and hilar lymph nodes. *J Bronchology* 1997;**4**:238–49.
18 **Dasgupta A**, Mehta AC, Wang KP. Transbronchial needle aspiration. *Semin Respir Crit Care Med* 1997;**18**:571–81.
19 **Harrow EM**, Wang KP. The staging of lung cancer by bronchoscopic transbronchial needle aspiration. *Chest Surg Clin North Am* 1996;**6**:223–35.
20 **Kraemer HC**, Periyakoil VS, Noda A. Tutorial in biostatistics: kappa coefficients in medical research. *Stat Med* 2002;**21**:2109–29.
21 **Thompson JR**. Estimating equations for kappa statistics. *Stat Med* 2001;**20**:2895–906.
22 **DerSimonian R**, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.
23 **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.
24 **Hedges LV**, Olkin I. Regression models in research synthesis. *Am Stat* 1983;**37**:137–40.
25 **Olkin I**, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 1998;**54**:317–22.
26 **Cooper H**, Hedges LV. *The Handbook of research synthesis*. New York: Russell Sage Foundation, 1994.
27 **Fleiss JL**. *Statistical methods for rates and proportions*, 2nd edn. New York: John Wiley & Sons, 1981.
28 **Harrow EM**, Abi-Saleh W, Blum J, *et al*. The utility of transbronchial needle aspiration in the staging of bronchogenic carcinoma. *Am J Respir Crit Care Med* 2000;**161**:601–7.
29 **Bilaceroglu S**, Cagirici U, Gunel O, *et al*. Comparison of rigid and flexible transbronchial needle aspiration in the staging of bronchogenic carcinoma. *Respiration* 1998;**65**:441–9.
30 **Disdier C**, Varela G, Sanchez de Cos J, *et al*. Utilidad de la puncion tranbronquial y la mediastinoscopia en la estadificacion ganglionar mediastinica del carcinoma broncogenico no microcitico. Estudio preliminar. *Arch Bronconeumol* 1998;**34**:237–44.
31 **Ratto GB**, Mereu C, Motta G. The prognostic significance of preoperative assessment of mediastinal lymph nodes in patients with lung cancer. *Chest* 1988;**93**:807–13.
32 **Schenk DA**, Bower JH, Bryan CL, *et al*. Transbronchial needle aspiration staging of bronchogenic carcinoma. *Am Rev Respir Dis* 1986;**134**:146–8.
33 **Herth F**, Hecker E, Hoffman H, *et al*. [Endobronchial ultrasound for local tumour and lymph node staging in patients with centrally growing lung cancer]. *Ultraschall Med* 2002;**23**:251–66.
34 **Wang R**, Han JJ, Yao J, *et al*. [The role of transbronchial needle aspiration in the staging of bronchogenic carcinoma]. *Zhongguo fei ai za zhi (Chinese Journal of Lung Cancer)* 2002;**5**:284–6.
35 **Patelli M**, Lazzari L, Poletti V, *et al*. Role of fiberoptic transbronchial needle aspiration in the staging of N2 disease due to non-small cell lung cancer. *Ann Thorac Surg* 2002;**73**:407–11.
36 **Katis K**, Katrogiannis G, Paraskevopoulos I, *et al*. Bronchoscopic needle aspiration in mediastinal staging of patients with bronchogenic carcinoma. *J Bronchology* 1998;**5**:195–9.
37 **Rong F**, Cui B. CT scan directed transbronchial needle aspiration biopsy for mediastinal nodes. *Chest* 1998;**114**:36–9.
38 **Schenk DA**, Chambers SL, Derdak S, *et al*. Comparison of the Wang 19-gauge and 22-gauge needles in the mediastinal staging of lung cancer. *Am Rev Respir Dis* 1993;**147**:1251–8.
39 **Schenk DA**, Strollo PJ, Pickard JS, *et al*. Utility of the Wang 18-gauge transbronchial histology needle in the staging of bronchogenic carcinoma. *Chest* 1989;**96**:272–4.
40 **Wang KP**, Brower R, Haponik EF, *et al*. Flexible transbronchial needle aspiration for staging of bronchogenic carcinoma. *Chest* 1983;**84**:571–6.
41 **Whiting P**, Rutjes AWS, Reitsma JB, *et al*. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;**140**:189–202.
42 **Brenner H**, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;**16**:981–91.
43 **Wiersema LM**, Edell ES, Midthun DE, *et al*. A blinded prospective study comparing transbronchial needle aspiration (TBNA) and endosonography guided needle aspiration (EUS FNA) in the staging of mediastinal lymphadenopathy (MLN) in patients with non small cell lung cancer (NSCLC). *Gastrointest Endosc* 2001;**53**:AB68.
44 **Gould MK**, Silvestri GA, Detterbeck F. Multidisciplinary management of lung cancer. *N Engl J Med* 2004;**350**:2008–10.

Accuracy of transbronchial needle aspiration for mediastinal staging of non-small cell lung cancer: a meta-analysis

Jon-Erik C. Holty, M.D.

Ware G. Kuschner, M.D.

Michael K. Gould, M.D., M.S.

Online data supplement

METHODS

We used systematic methods to identify relevant studies, apply inclusion and exclusion criteria, evaluate study quality and summarize the diagnostic accuracy of TBNA for mediastinal lymph node involvement in patients with non-small cell lung cancer.

Literature search

An investigator (J.C.H.) and a professional librarian developed a computerized search strategy to identify relevant studies published between January 1966 and 1 July 2003 in the Medline and Embase electronic databases. This strategy employed key words (both controlled vocabulary and free text terms) and was divided into three parts each connected by the [AND] bullion. The first part mapped the search to non-small cell lung cancer and included key words such as lung neoplasm/cancer, bronchial neoplasm/cancer or carcinoma/cancer non-small-cell lung. The second part of the search strategy mapped the search to staging concepts and included key words such as neoplasm staging, lymphatic/lymph-node metastasis, neoplasm metastasis or mediastinal cancer/neoplasm. The final part mapped the search to TBNA and included key words such as bronchoscopy/transbronchial/TBNA and biopsy/needle biopsy/aspirate/sentinel lymph node biopsy/FNA. We first searched for articles in the Medline database. All duplicate articles found in the Embase database were excluded as well as non-human studies. A detailed description of our initial search strategies is shown in fig S1A and B.

We updated the literature search in Medline through 6 April 2004 by employing the key words (both controlled vocabulary and free text terms) *transbronchial* or *TBNA*. In addition, we manually searched reference lists of included studies and review articles and reviewed practice

guidelines and systematic reviews. All articles regardless of language were considered for inclusion.

Inclusion and exclusion criteria

The initial search strategy (fig S1A and B) yielded a total of 525 articles (fig 1). A careful review of titles and abstracts eliminated 398 articles not dealing with TBNA. A hand search of the bibliographies of the remaining 127 articles identified 203 additional potentially relevant studies. We excluded studies that examined rigid bronchoscopy. An initial review of full reports by one investigator (J.E.H.) excluded 268 studies for the following reasons: not a study of staging (n=106); study of rigid bronchoscopy (n=69); review article or no primary data presented (n=61); or miscellaneous reasons (n=32). Sixty-two potentially eligible studies underwent further review (table S1).

Each non-excluded English-language study was reviewed by at least two investigators (J.C.H., M.K.G. or W.G.K.) to assess whether they met inclusion criteria. To be included, a study had to (1) examine TBNA using a flexible bronchoscope for mediastinal staging in patients with known or suspected NSCLC; (2) enroll at least 10 subjects with and/or 10 subjects without mediastinal lymph node involvement; (3) provide sufficient original data to permit calculation of sensitivity and/or specificity; and (4) for studies that did not provide separate data for participants with disorders other than NSCLC, include no more than ten percent (≤10%) of subjects with primary diagnoses other than NSCLC. This final inclusion criteria was chosen because TBNA may be more sensitive in patients with small cell lung cancer,[1] and because staging and treatment differs between NSCLC and small cell lung cancer.[2][3] These criteria were designed to identify studies that met minimal standards of acceptability.

Table S1. Potentially eligible studies*

Exclusion and inclusion criteria	Study
Full-text reports for detailed evaluation (n=62)	
Study did not examine TBNA using a flexible bronchoscope for mediastinal staging in patients with NSCLC	[4][5][6][7][8][9][10][11][12]
Fewer than 10 subjects with or without mediastinal metastasis	[4] [6][7][8][9][10] [13][14][15][16][17][18][19][20][21][22][23]
Insufficient data to calculate sensitivity or specificity	[5][6][7][8][9][10][11][12][13][14][15][17][18][19][20][21][22][23][24][25][26][27][28][29][30][31][32][33][34][35][36][37][38][39][40][41][42][43][44][45][46][47]
Fewer than 90% of enrolled subjects had NSCLC or did not provide separate data for patients with NSCLC	[1] [4] [6][7][8][9][10] [13][14] [17][18] [22][23][24][25][26][27][28][29] [32] [35] [40] [45][46][47][48][49][50][51][52][53][54][55]
Abstract	[56]
Met inclusion criteria	[57][58][59][60][61][62][63][64]
Authors provided additional information that enabled us to include their studies	[1] [50][51] [54][55]
Updated Medline search (n=9)	
Study did not examine TBNA using a flexible bronchoscope for mediastinal staging in patients with NSCLC	[65][66][67][68]
Insufficient data to calculate sensitivity or specificity	[65][66][67][68][69][70]
Fewer than 90% of enrolled subjects had NSCLC or did not provide separate data for patients with NSCLC	[65] [71]
No primary data (e.g. review article)	[72][73]

* Studies could meet one or more exclusion criteria.

Because many studies did not require pathologic confirmation of positive TBNA results, and because this may result in biased estimates of test performance, we divided studies *a priori* into two tiers. Tier 1 studies enrolled at least 10 subjects with and without mediastinal lymph node involvement, surgically confirmed all TBNA results and used the patient as the unit of analysis. Non-tier 1 studies did not meet one or more of these criteria, but met all other criteria for inclusion.

We included abstracts only when the study authors provided full reports of their methods and results. We requested additional information from the authors of all studies that did not report data sufficient to calculate sensitivity and/or specificity for NSCLC staging. If two or more publications by the same authors with overlapping patients were identified, only the most recent study was included. Disagreements were resolved by discussion and/or by consulting with a third investigator (J.C.H., M.K.G. or W.G.K.). A single reviewer evaluated non-English-language studies. A Cohen's kappa for interrater reliability was calculated to assess agreement about study eligibility between raters (table S2).[74] Based on our selection criteria, we excluded 54 studies that either provided insufficient data to calculate sensitivity or specificity (76%) or enrolled <90% of subjects having NSCLC (60%). We obtained additional unpublished data from seven authors that enabled us to include five additional studies.[1] [50][51] [54][55]

Table S2. Cohen's kappa measure of agreement between reviewers on included studies

	Reviewer 2 (M.K.G.)	Reviewer 3 (W.G.K.)
Reviewer 1 (J.E.H.)	0.83	0.77

Study quality

We adapted criteria for methodological quality proposed by Kent et al.,[75] to identify high-quality studies of TBNA for lung cancer staging (fig S2). The revised criteria cover seven assessment categories: technical quality of TBNA, technical quality of the reference test, application of the reference test, independence of interpretations, clinical description, cohort assembly and sample size. These criteria have been used to evaluate several different diagnostic modalities, including CT for lumbar stenosis,[75] polymerase chain reaction for the diagnosis of human immunodeficiency virus infection,[76] and positron emission tomography for the evaluation of focal pulmonary lesions[77] and in the staging of lung cancer.[78] Independent criteria for the technical quality TBNA were developed based on the clinical experience of two of the authors experienced in the use of TBNA (M.K.G and W.G.K) and by reviewing published guidelines.[79][80][81] All English-language studies were assessed by at least two investigators (J.C.H., M.K.G. or W.G.K.) for methodological quality with all disagreements resolved by discussion. One reviewer assessed the quality of all non-English language studies.

Data abstraction

One investigator (J.C.H) abstracted primary data regarding patient characteristics and the sensitivity and/or specificity of TBNA for identifying mediastinal metastasis in patients with NSCLC. To assess the accuracy of TBNA, data abstraction was performed for patients with a primary diagnosis of NSCLC that was confirmed histologically.

When possible, we separated staging characteristics of TBNA for patients with and without enlarged lymph nodes on CT, and for biopsies performed at hilar, subcarinal, paratracheal or other lymph node stations. We also separately tabulated test characteristics for

studies utilizing 'real-time' imaging (e.g. CT fluoroscopy, enbronchial ultrasound or transthoracic ultrasound).

Data synthesis and sensitivity/specificity calculations:

We constructed 2 x 2 contingency tables for each study to summarize the results of TBNA and the reference test(s). For each study, the true positive rate (TPR; sensitivity), the false-positive rate (FPR; 1-specificity), the log odds ratio (LOR; log odds TPR - log odds FPR) and the kappa-1 statistic were calculated. To calculate log odds ratios, we added a correction factor of 0.5 to each cell in any 2 x 2 table that contained one or more zero values.

Because many studies did not confirm positive TBNA results surgically, the false-positive rates for these studies are unknown. We calculated a weighted kappa-1 coefficient (a generalization of the unweighted or Cohen's kappa coefficient) to assess accuracy with regards to avoiding false negative results.[82][83] Calculation of the kappa-1 coefficient does not require the false positive rate (1-specificity) (but does require knowledge of the marginal probabilities) thus providing a less biased method to compare diagnostic accuracy in studies that did and did not surgically confirm all positive test results. We calculated a pooled kappa-1 coefficient by using both fixed[84] and random effects models.[85] The asymptotic variance of the weighted kappa coefficient can be estimated.[86] Weighted kappa coefficient values close to one suggest good test accuracy, while values less than 0.40 suggest only fair to poor test accuracy.[83] [87]

Summary receiver operating characteristic (SROC) curves as described by Moses *et al.*,[88][89] were constructed to quantitatively summarize the results of studies. These curves demonstrate the trade-off between sensitivity and specificity as the threshold for defining a positive test varies. This method assumes that individual study estimates of sensitivity and

specificity represent unique points on a common SROC curve. We performed ordinary least squares regression by using the log odds ratio as the dependent variable and an implied function of the test threshold (logit TPR + logit FPR) as the independent variable, and plotted the relationship between the TPR and FPR after performing a reverse transformation, as we have done previously.[77][78] Any studies that enrolled no patients with or without mediastinal lymph node involvement were excluded from the SROC analysis.

When the SROC curve is symmetrical (e.g. $\beta \approx 0$), the studies may be summarized by a common LOR. Beta coefficients were not statistically significantly different from zero for tier 1 studies (0.26; 95% CI, -1.75 to 2.28), non-tier 1 studies (-0.28; 95% CI, -1.34 to 0.79) and all studies combined (0.29; 95% CI, -0.13 to 0.70). The LOR represents the odds of a positive test in subjects with disease relative to the odds of a positive test in subjects without disease. To estimate the common log OR, we employed both fixed[84] and random effects models,[84][85] adding 0.5 to each cell in any 2 x 2 table that contained one or more zero values.

Pooled sensitivity and specificity were calculated by fixed and random effects models. When pooling sensitivity and specificity, studies with <10 subjects with or without mediastinal lymph node involvement were excluded, respectively, in the calculations. Heterogeneity was assessed by the Q-statistic. When there was statistically significant heterogeneity, a random effects model, as described by DerSimonian and Laird,[85] was used to summarize trial results.

Studies often have different thresholds to define positive and negative test results. Thus TPR and FPR are typically positively correlated and therefore not independent. Pooling sensitivity and specificity is often inappropriate.[90][91] However, we believe that separate pooling of the sensitivity and specificity for studies of TBNA for mediastinal staging is appropriate for several reasons. Unlike most other diagnostic tests, the TPR and FPR of TBNA

do not appear to be correlated. Implicit and explicit thresholds are unlikely to affect sensitivity differently from specificity (e.g. TBNA technical criteria that decrease false negative results are unlikely to result in more false positives). In fact, linear regression analysis of tier 1 studies demonstrated a weak (slope near zero), but negative correlation between TPR and FPR. Furthermore, TBNA test results have a dichotomous rather than continuous outcome. Only one false positive was reclassified as a true negative when more stringent reference standards were employed in studies with suboptimal reference standards.

Sensitivity analysis

We performed sensitivity analysis to evaluate potential sources of heterogeneity between studies. Sensitivity analysis included stepwise single study elimination to assess significant changes in the LOR or pooled sensitivity. We also assessed whether varying the correction factor from 0.5 to 0.1 had any significant impact on LOR or sensitivity at median specificity on the SROC curve. Finally, we varied assumptions about the definition of a positive or negative biopsy result in certain studies. For example, we reexamined the individual study test results and considered any biopsy specimen negative (when possible) if the biopsy was aborted, if nodal tissue was not located, if there was insufficient tissue for pathologic examination (e.g. inadequate or inconclusive specimens), if the cytopathologist identified the specimen as "contaminated" or if the specimen contained "atypical" cells. All biopsy results that were "suspicious" for malignancy were considered positive as long as the aspirate did not contain an abundance of columnar epithelial cells, if scarce malignant cells were identified or if lymphocytes were absent. We than reanalyzed the pooled sensitivity, specificity, LOR and kappa-1 coefficient based on

these changes and compared this to our base-case analysis (the sensitivity and specificity reported by the study authors).

To date, empirical studies of statistical methods (i.e. funnel plots) in assessing publication bias have focused on randomized clinical control trials of treatment effect and not on diagnostic studies. Thus, there are no universally accepted methods to assess publication bias in trials of diagnostic accuracy. In the absence of any generally accepted method, we constructed inverted funnel plots of standard error versus estimated effect size (LOR) for each individual study to assess for the presence of publication bias.[84] If additional small studies were conducted, but not published due to unfavorable results (e.g. low sensitivity), the funnel plot should be asymmetric.

We also assessed differences in diagnostic accuracy between tier 1 and non-tier 1 studies via discriminant function analysis.[92] This analysis was repeated with respect to studies that confirmed or did not confirm all TBNA results, and studies with high (≥60%) versus low (<60%) prevalence of mediastinal metastasis. P-values were calculated via a parametric Wilks' Lambda test.

Meta-regressions

We performed a multivariate analysis of variance (ANOVA) to examine the effect of specific study characteristics on sensitivity and overall diagnostic accuracy (LOR).[93][94] Study characteristics included prevalence of lymph node metastasis (≥60% or <60%) and year of study publication (≥1995 or <1995).

Statistical models

All biostatistical models were programmed in Excel 8.0 for Windows (Microsoft Corporation, Redmond, Washington, USA). Discriminant function analysis was performed in SAS 9.0 for Windows (SAS Corp, Cary, North Carolina). We calculated 95% confidence intervals (CI's) for the TPR and the FPR by using the quadratic method.[95] A normal approximation to the binomial of the standard error was used in calculating all other CI's, as appropriate. When making comparisons between groups of studies we used an unpaired t-test or the Mann-Whitney U test as appropriate. A two-tailed p-value <0.05 was considered statistically significant.

Table S3. Characteristics of studies in the meta-analysis[§]

Study	Fiberoptic bronchoscope type	Needle type & size	Number of passes*	Real-time imaging	Surgical reference
Tier 1 studies[‡]					
Harrow et al.	N/S	SW-121, SW-122 or MW-319 needle, 21 or 22-gauge (cytology) or 19-gauge (histology)	2 to 3 aspirates	No	Mediastinoscopy, mediastinotomy, or standard or video-assisted thoracotomy
Bilaceroglu et al	Olympus	10-mm long SW-221 needles, 21-gauge	3 to 5	No[†]	Mediastinoscopy, mediastinotomy or thoracotomy
Disdier et al.	Olympus (1T20D or P20D) or Pentax 2000E	13-mm long Olympus NA2C needles, 21-gauge	1 to 3 punctures	No	Mediastinoscopy, mediastinotomy or thoracotomy
Ratto et al.	N/S	Olympus NM-1K 21-Gauge	N/S	No	Mediastinoscopy or thoracotomy
Schenk et al.	Olympus BF-4B2	Wang aspiration fixed Type 1 or retractable Type 2 biopsy needles	3 to 5 aspirates	No	Cervical or parasternal mediastinal exploration or thoracotomy
Studies not meeting tier 1 criteria					
Herth et al.	N/S	MW-522 needles, 22-gauge	N/S	Endobronchial ultrasound	Thoracotomy
Wang R. et al.	N/S	Needle 1.85 mm	N/S	No	Thoracotomy
Patelli et al.	N/S	22-gauge cytology or 19-gauge histology needles	N/S	No	Mediastinoscopy or video-assisted thoracotomy
Katis et al.	N/S	1.3-cm Olympus NA-1C 21-gauge or 1.5-cm MW-220-1 20 gauge needles	2 to 3	No	Mediastinoscopy or mediastinotomy
Rong et al.	Olympus BF-1T20	N/S	3	Real-time CT	Thoracotomy
Schenk et al.	N/S	Wang 22-gauge cytology needle and Wang 19-gauge histology needle	3 to 4 aspirates	No	Mediastinoscopy, mediastinotomy, thoracotomy or percutaneous needle aspiration
Schenk et al.	Olympus (BF4B2 or P-10)	Wang 18-gauge	3 to 4 aspirates	No	"Surgical mediastinal exploration"
Wang K. P. et al.	N/S	N/S	N/S	No	Mediastinoscopy or thoracotomy

***** Number of passes per lymph node station.

† In this study fluoroscopy was occasional used to guide needle placement.

‡ Studies meeting tier 1 criteria surgically confirmed all TBNA's, had at least 10 subjects with and without mediastinal lymph node metastasis and used the patient as the unit of analysis.

§ Abbreviations: N/S= not specified.

Table S4. Number of quality criteria met in studies of TBNA for mediastinal staging*

Study	Study quality measure							Total
	Index test technical quality (by test)†	Reference test quality (by study)	Application of reference test (by study)	Independence of test interpretation	Clinical description & characteristics	Cohort assembly	Sample size	
Harrow 2000	8	0	1	1	2	4	2	18
Disdier 1998	9	2	1	2	4	3	0	21
Bilaceroglu 1998	7	2	1	1	4	6	2	23
Ratto 1988	3	2	2	2	2	3	0	14
Schenk 1986	7	1	1	0	2	6	1	18
Patelli 2002	9	0	0	0	2	3	1	15
Herth 2002†	10	1	1	0	3	2	1	18
Wang R. 2002	8	1	1	0	3	3	0	16
Katis 1998	7	0	1	0	4	5	1	18
Rong 1998†	6	1	1	0	3	2	0	13
Schenk 1993	10	0	0	1	2	5	1	19
Schenk 1989	7	0	1	0	2	5	0	15
Wang K. P. 1983	4	0	1	1	2	4	0	12
Total number of criteria	13	2	2	3	4	8	2	34

***** Highest possible score is 34.

† Studies that utilized 'real-time' imaging to guide needle placement had three additional criteria.

Table S5. Sensitivity analysis: effect of excluding individual studies

	Pooled LOR (95% CI)	Pooled sensitivity (95% CI)
Tier 1 studies		
Harrow et al. 2000	16.8 (5.4 to 52.7)	41% (13% to 68%)
Bilaceroglu et al. 1998	13.9 (4.8 to 40.7)	29% (18% to 40%)
Disdier et al. 1998	22.6 (7.4 to 69.3)	40% (13% to 66%)
Ratto et al. 1988	14.0 (4.5 to 43.9)	45% (24% to 67%)
Schenk et al. 1986	19.6 (5.6 to 69.6)	39% (12% to 67%)
Tier 1 summary*	18.0 (6.6 to 49.2)	39% (17% to 61%)
Non-tier 1 studies†		
Wang R. et al. 2002	55.9 (14.5 to 215.9)	79% (67% to 91%)
Patelli et al. 2002	51.2 (11.0 to 238.2)	84% (71% to 97%)
Katis et al. 1998	73.6 (16.5 to 329.1)	81% (69% to 92%)
Schenk et al. 1993	48.7 (10.7 to 221.9)	80% (67% to 92%)
Schenk et al. 1989	61.3 (13.7 to 273.4)	80% (69% to 92%)
Wang K. P. et al. 1983	48.1 (10.7 to 216.7)	80% (67% to 94%)
Non-tier 1 summary*	52.1 (14.2 to 193.4)	78% (71% to 84%)

***** These are the pooled overall results without study elimination.

† The two non-tier 1 studies that utilized 'real-time' radiological needle guidance were excluded.[50][51]

References

1. Schenk DA, Bower JH, Bryan CL, et al. Transbronchial needle aspiration staging of bronchogenic carcinoma. *Am Rev Respir Dis* 1986;134:146-148.

2. Ries LAG. Influence of extent of disease, histology and demographic factors on lung cancer survival in the SEER population-based data. *Semin Surg Oncol* 1994;10:21-30.

3. Mountain CF. Revisions in the international system for staging lung cancer. *Chest* 1997;111(6):1710-1717.

4. Horsley JR, Miller RE, Amy RW, et al. Bronchial submucosal needle aspiration performed through the fiberoptic bronchoscope. *Acta Cytologica* 1984;28(3):211-7.

5. Marel M, Melinova L, Potockova L, et al. [Comparison of the results of transbronchial puncture of the lymph nodes with surgical findings in patients with lung cancer]. *Cas Lek Cesk* 1990;129(23):727-9.

6. Davenport RD. Rapid on-site evaluation of transbronchial aspirates. *Chest* 1990;98(1):59-61.

7. Wang KP. Transbronchial needle aspiration to obtain histology specimen. *J Bronchology* 1994;1:116-122.

8. Gluskowski J, Zajaczkowska J. The role of bronchoscopy in the diagnosis of enlargement of the hilar and-or mediastinal lymph nodes. *Pol Med J* 1972;11(5):1180-7.

9. Schaberg T, Mai J, Thalmann U, et al. [Transbronchial lymph node puncture via the fibre bronchoscope in pulmonary neoplasms]. *Pneumologie* 1986;40:306-311.

10. Torzsok L. [The transbronchial puncture]. *Prax pneumol* 1975;29:162-165.

11. Filippow VP, Borisow VV, Timashewa ED. [Technique and diagnosis with the transbronchial biopsy of mediastinal lymph nodes]. *Probl Tuberk* 1972;50:80-81.

12. Kharchenko VP, Galil-Ogly GA, Kuz'min IV, et al. [Classification, diagnosis and treatment of x-ray-negative lung cancer]. *Sov Med* 1988(12):70-5.

13. Wang KP, Haponik EF, Gupta PK, et al. Flexible transbronchial needle aspiration. Technical considerations. *Ann Otol Rhinol Laryngol* 1984;93(3 Pt 1):233-6.

14. Gasparini S, Zuccatosta L, Zitti P, et al. Integration of TBNA and TCNA in the diagnosis of peripheral lung nodules. Influence on staging. *Ann Ital Chir* 1999;70(6):851-5.

15. Blainey AD, Curling M, Green M. Transbronchial aspiration of subcarinal lymph nodes. *Br J Dis Chest* 1988;82(2):149-54.

16. Garpestad E, Goldberg S, Herth F, et al. CT fluoroscopy guidance for transbronchial needle aspiration: an experience in 35 patients. *Chest* 2001;119(2):329-32.

17. Diette GB, White P, Jr., Terry P, et al. Utility of on-site cytopathology assessment for bronchoscopic evaluation of lung masses and adensopathy. *Chest* 2000;117(4):1186-1190.

18. Gittlen SD, Erozan Y, Wang KP. A new versatile transbronchial cytology needle for the staging and diagnosis of bronchogenic carcinoma. *Chest* 1988;94(3):561-5.

19. Wang EY, Wragg T, Nguyen GK. Role of transbronchial fine-needle aspiration in the investigation of mediastinal lymphadenopathy in patients suspected to have lung cancers. *Diagn Cytopathol* 2002;26(2):132-4.

20. Harrow E, Millard PS, Wight G, et al. Survival duration among patients with lung cancer staged by bronchoscopic needle aspiration. *J Bronchology* 1996;3:96-101.

21. Gay PC, Brutinel WM. Transbronchial needle aspiration in the practice of bronchoscopy. *Mayo Clin Proc* 1989;64:158-162.

22. Herth F, Becker HD. Endobronchial ultrasound (EBUS) - assessment of a new diagnostic tool in bronchoscopy. *Onkologie* 2001;24:151-154.

23. Salathe M, Soler M, Bolliger CT, et al. Transbronchial needle aspiration in routine fiberoptic bronchoscopy. *Respiration* 1992;59(1):5-8.

24. Chin R, Jr., McCain TW, Lucia MA, et al. Transbronchial needle aspiration in diagnosing and staging lung cancer: how many aspirates are needed? *Am J Respir Crit Care Med* 2002;166(3):377-81.

25. Castella J, Buj J, Puzo C, et al. Diagnosis and staging of bronchogenic carcinoma by transtracheal and transbronchial needle aspiration. *Ann Oncol* 1995;6 Suppl 3:S21-4.

26. Crocket JA, Wong EY, Lien DC, et al. Cost effectiveness of transbronchial needle aspiration. *Can Respir J* 1999;6(4):332-5.

27. Herth FJ, Becker HD, Ernst A. Ultrasound-guided transbronchial needle aspiration: an experience in 242 patients. *Chest* 2003;123(2):604-7.

28. Utz JP, Patel AM, Edell ES. The role of transcarinal needle aspiration in the staging of bronchogenic carcinoma. *Chest* 1993;104(4):1012-6.

29. Harrow E, Halber M, Hardy S, et al. Bronchoscopic and roentgenographic correlates of a positive transbronchial needle aspiration in the staging of lung cancer. *Chest* 1991;100(6):1592-6.

30. Harrow EM, Oldenburg FA, Jr., Lingenfelter MS, et al. Transbronchial needle aspiration in clinical practice. A five-year experience. *Chest* 1989;96(6):1268-72.

31. Wang KP. Flexible transbronchial needle aspiration biopsy for histologic specimens. *Chest* 1985;88(6):860-3.

32. Shure D, Fedullo PF. The role of transcarinal needle aspiration in the staging of bronchogenic carcinoma. *Chest* 1984;86(5):693-6.

33. Rong F, Chui B, Guo S. [Transbronchial needle aspiration under CT guide for enlarged mediastinal lymph node biopsy]. *Zhonghua Zhong Liu Za Zhi (Chinese journal of oncology)* 1996;18(6):458-60.

34. Rong F, Guo S, Chen J. [Transbronchial needle aspiration in routine bronchoscope practice]. *Zhonghua Jie He He Hu Xi Za Zhi (Chinese journal of tuberculosis and respiratory diseases)* 2000;23(1):37-9.

35. Benov E, Michev K, Kostadinov D, et al. Transbronchial needle aspiration in the staging of bronchogenic carcinoma. *Diagn Therap Endoscopy* 1996;3(1):61-65.

36. Ceron L, Ceccetto A, Manzato M, et al. [Transbronchial needle aspiration (TBNA): personal experience]. *E J Pathol Histol* 1997;3(4):70-75.

37. Ceron L, Ceccetto A, Manzato M, et al. [Diagnosis and staging of mediastinal-pulmonary disease by transbronchial needle aspiration (TBNA)]. *L'Internista* 2000;8:248-252.

38. Okazaki M, Takakura S, Fujii H, et al. [Endoscopic approach to pulmonary diseases: Transbronchial needle aspiration]. *Kekkaku* 2000;75(1):41-46.

39. Ceron L, Wang KP. Involvement of N-components in lung cancer. *Monaldi Arch Chest Dis* 1997;52(6):540-3.

40. Wang KP, Selcuk ZT, Erozan Y. Transbronchial needle aspiration for cytology specimens. *Monaldi Arch Chest Dis* 1994;49(3):265-7.

41. Mlynarczyk W, Wruk M, Meissner R. [Transbronchial biopsy of mediastinal lymph nodes in the diagnosis of neoplastic changes in the lungs]. *Pneumonol Pol* 1980;48(12):803-7.

42. Debeljak A, Mermolja M, Orel J, et al. [Transbronchial needle aspiration with fiberoptic and rigid bronchoscope in the diagnosis and staging of lung cancer]. *Radiol Oncol* 1994;28(4):309-15.

43. Ceron L, Ceccetto A, Manzato M, et al. [Transbronchial needle aspiration (TBNA) for the diagnosis of mediastinal-hilar disease for the staging of lung cancer: personal experience]. *Rass Patol App Respir* 1999;14:103-9.

44. Fernandez Villar JA, Iglesias Rio F, Barreiro Barreiro JM, et al. [Clinical usefulness and cost-effectiveness of transbronchial needle aspiration for the diagnosis of mediastinal adenopathy]. *Rev Clin Esp* 2001;201(4):169-173.

45. Shields GW, Coggeshall JW, Witt WS, et al. Transbronchoscopic needle aspiration in the diagnosis and staging of lung cancer. *South Med J* 1986;79(6):694-5.

46. Harrow EM, Oldenburg FA, Smith AM. Transbronchial needle aspiration in clinical practice. *Thorax* 1985;40(10):756-9.

47. Yu Ho K, In Seon C, Ik Ju J, et al. [Clinical evaluation on transbronchial needle aspiration (TBNA) of subcarinal lymph node in lung cancer]. *Kyorhaek mit hohupki chirhwan (Tuberculosis and respiratory diseases)*, 1993:177-184.

48. Baker JJ, Solanki PH, Schenk DA, et al. Transbronchial fine needle aspiration of the mediastinum. Importance of lymphocytes as an indicator of specimen adequacy. *Acta Cytol* 1990;34(4):517-23.

49. Shannon JJ, Bude RO, Orens JB, et al. Endobronchial ultrasound-guided needle aspiration of mediastinal adenopathy. *Am J Respir Crit Care Med* 1996;153(4 Pt 1):1424-30.

50. Herth F, Hecker E, Hoffman H, et al. [Endobronchial ultrasound for local tumour and lymph node staging in patients with centrally growing lung cancer]. *Ultraschall Med* 2002;23(4):251-266.

51. Rong F, Cui B. CT scan directed transbronchial needle aspiration biopsy for mediastinal nodes. *Chest* 1998;114(1):36-9.

52. Rodriguez de Castro F, Diaz Lopez F, Serda GJ, et al. Relevance of training in transbronchial fine-needle aspiration technique. *Chest* 1997;111(1):103-5.

53. Rodriguez de Castro F, Rey A, Caminero J, et al. Transbronchial fine needle aspiration in clinical practice. *Cytopathology* 1995;6(1):22-9.

54. Bilaceroglu S, Cagirici U, Gunel O, et al. Comparison of rigid and flexible transbronchial needle aspiration in the staging of bronchogenic carcinoma. *Respiration* 1998;65:441-449.

55. Harrow EM, Abi-Saleh W, Blum J, et al. The utility of transbronchial needle aspiration in the staging of bronchogenic carcinoma. *Am J Respir Crit Care Med* 2000;161(2 Pt 1):601-7.

56. Wiersema LM, Edell ES, Midthun DE, et al. A blinded prospective study comparing transbronchial needle aspiration (TBNA) and endosonography guided needle aspiration (EUS FNA) in the staging of mediastinal lymphadenopathy (MLN) in patients with non small cell lung cancer (NSCLC) [Abstract #556]. *Gastrointest Endosc* 2001;53(5):AB68.

57. Schenk DA, Chambers SL, Derdak S, et al. Comparison of the Wang 19-gauge and 22-gauge needles in the mediastinal staging of lung cancer. *Am Rev Respir Dis* 1993;147:1251-1258.

58. Patelli M, Lazzari L, Poletti V, et al. Role of fiberoptic transbronchial needle aspiration in the staging of N2 disease due to non-small cell lung cancer. *Ann Thorac Surg* 2002;73:407-411.

59. Disdier C, Varela G, Sanchez de Cos J, et al. Utilidad de la puncion tranbronquial y la mediastinoscopia en la estadificacion ganglionar mediastinica del carcinoma broncogenico no microcitico. Estudio preliminar. *Arch Bronconeumol* 1998;34:237-244.

60. Schenk DA, Strollo PJ, Pickard JS, et al. Utility of the Wang 18-gauge transbronchial histology needle in the staging of bronchogenic carcinoma. *Chest* 1989;96(2):272-274.

61. Ratto GB, Mereu C, Motta G. The prognostic significance of preoperative assessment of mediastinal lymph nodes in patients with lung cancer. *Chest* 1988;93(4):807-813.

62. Wang KP, Brower R, Haponik EF, et al. Flexible transbronchial needle aspiration for staging of bronchogenic carcinoma. *Chest* 1983;84(5):571-6.

63. Wang R, Han JJ, Yao J, et al. [The role of transbronchial needle aspiration in the staging of bronchogenic carcinoma]. *Zhongguo fei ai za zhi (Chinese journal of lung cancer)* 2002;5(4):284-6.

64. Katis K, Kotrogiannis G, Paraskevopoulos I, et al. Bronchoscopic needle aspiration in mediastinal staging of patients with bronchogenic carcinoma. *J Bronchology* 1998;5:195-199.

65. Sharafkhaneh A, Baaklini W, Gorin AB, et al. Yield of transbronchial needle aspiration in diagnosis of mediastinal lesions. *Chest* 2003;124:2131-2135.

66. Cetinkaya E, Yildiz P, Altin S, et al. Diagnostic value of transbronchial needle aspiration by Wang 22-gauge cytology needle in intrathoracic lymphadenopathy. *Chest* 2004;125(2):527-531.

67. Win T, Stewart S, Groves AM, et al. The role of transbronchial needle aspiration in the diagnosis of bronchogenic carcinoma. *Respir Care* 2003;48(6):602-5.

68. Khoo KL, Chua GS, Mukhopadhyay A, et al. Transbronchial needle aspiration: initial experience in routine diagnostic bronchoscopy. *Respir Med* 2003;97(11):1200-4.

69. Chokhani R, Gasparini S. Transbronchial needle aspiration (TBNA) in the early diagnosis and staging of bronchogenic carcinoma. *Indian J Chest Dis Allied Sci* 2003;45(2):111-5.

70. Herth F, Becker HD, Ernst A. Conventional vs endobronchial ultrasound-guided transbronchial needle aspiration: a randomized trial. *Chest* 2004;125(1):322-325.

71. Hsu LH, Liu CC, Ko JS. Education and experience improve the performance of transbronchial needle aspiration: a learning curve at a cancer center. *Chest* 2004;125(2):532-540.

72. Kramer H, Groen HJ. Current concepts in the mediastinal lymph node staging of nonsmall cell lung cancer. *Ann Surg* 2003;238(2):180-8.

73. Silvestri GA, Hoffman B, Reed CE. One from column A: choosing between CT, positron emission tomography, endoscopic ultrasound with fine-needle aspiration, transbronchial needle aspiration, thoracoscopy, mediastinoscopy, and mediastinotomy for staging lung cancer. *Chest* 2003;123(2):333-5.

74. Orwin RG. Evaluating coding decisions. In: Cooper, H, Hedges, LV, editors. *The handbook of research synthesis*. 1st ed. New Yory: The Russel Sage Foundation, 1994:139-162.

75. Kent DL, Haynor DR, Larson EB, et al. Diagnosis of lumbar spinal stenosis in adults: a meta-analyis of the accuracy of CT, MR, and myelography. *AJR Am J Roentgenol* 1992;158:1135-1144.

76. Owens DK, Holodniy M, Garber A. 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;124:803-815.

77. Gould MK, Maclean CC, Kuschner WG, et al. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions. A meta-analysis. *JAMA* 2001;285(7):914-924.

78. Gould MK, Kuschner WG, Rydzak CE, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer. *Ann Intern Med* 2003;139(11):879-892.

79. Harkin TJ, Wand KP. Bronchoscopic needle aspiration of mediastinal and hilar lymph nodes. *J Bronchology* 1997;4:238-249.

80. Dasgupta A, Mehta AC, Wang KP. Transbronchial needle aspiration. *Seminars in Respiratory and Critical Care Medicine* 1997;18(6):571-581.

81. Harrow EM, Wang KP. The staging of lung cancer by bronchoscopic transbronchial needle aspiration. *Chest Surg Clin N Am* 1996;6(2):223-35.

82. Kraemer CK, Periyakoil VS, Noda A. Tutorial in biostatistics: Kappa coefficients in medical research. *Stat Med* 2002;21:2109-2129.

83. Thompson JR. Estimating equations for kappa statistics. *Stat Med* 2001;20:2895-2906.

84. Cooper H, Hedges LV. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, 1994.

85. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.

86. Fleiss JL, Cohen J, Everitt BS. Large-sample standard errors of kappa and weighted kappa. *Psychol Bull* 1969;72:323-327.

87. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-174.

88. 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.

89. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. *Med Decis Making* 1993;13:313-321.

90. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;120(8):667-76.

91. Irwig L, Macaskill P, Glasziou P, et al. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;48(1):119-130.

92. Press SJ, Wilson S. Choosing between logistic regression and discriminant analysis. *J Amer Stat Assoc* 1978;73:699-705.

93. Hedges LV, Olkin I. Regression models in research synthesis. *Am Stat* 1983;37(2):137-140.

94. Olkin I, Sampson A. Comparison of Meta-analysis versus analysis of variance of individual patient data. *Biometrics* 1998;54:317-322.

95. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York: John Wiley & Sons, 1981.

FIGURE LEGENDS

Figure S1A. TBNA Medline search strategy.

Figure S1B. TBNA Embase search strategy.

Figure S2. Methodological quality questionnaire.

Figure S3. Individual study estimates of sensitivity and 1-specificity of TBNA for identifying mediastinal metastasis. Error bars represent 95% CI's. Point estimates and 95% CI's for studies with high mediastinal metastasis prevalence (≥60%) and low prevalence (<60%) are shown.

***** Specificity was not calculated for the study by Wang, R. et. al.,[85] because the prevalence of mediastinal lymph node metastasis was 100%.

† Specificity is not shown for studies that did not surgically confirm all TBNA results.

‡ Summary high prevalence specificity not shown because only three studies [50][51] allowed calculation of specificity (all 100%), and two of these studies had less than 10 patients without mediastinal metastases.[50][51]

Figure S4. Inverted funnel plot. Log odds ratio (LOR) versus standard error.

Figure S1A.

(lung neoplasms[Exploded MeSH] OR bronchial neoplasms[MeSH] OR carcinoma, non-small-cell lung[MeSH] OR Neoplasms, squamous cell[Exploded MeSH] OR (carcinoma, squamous cell[Exploded MeSH] AND lung[Exploded MeSH]) OR (Mediastinum[Exploded MeSH] and neoplasms[Exploded MeSH]) OR mediastinal neoplasms[MeSH] OR lung(W)cancer?[title/abstract word] OR non(W)small(W)cell[title/abstract word] OR pulmonary(N)neoplasm?[title/abstract word] OR pulmonary(N)malignan?[title/abstract word] OR lung(N)neoplasm?[title/abstract word] OR lung(N)malignan?[title/abstract word])

AND

Neoplasm staging[MeSH] OR lymphatic metastasis[MeSH] OR lymph nodes[Exploded MeSH] OR lymphatic diseases[Exploded MeSH] Stag????[title/abstract word] OR lymph(W)node[title/abstract word] OR lymphatic[title/abstract word]

AND

Bronchoscopy[MeSH] OR bronchoscope?[title/abstract word] OR transbronchial[title/abstract word] OR tbna[title/abstract word]

AND

biopsy, needle[MeSH] OR biopsy[Exploded MeSH] OR sentinel lymph node biopsy[MeSH] OR (fine(W)needle?[title/abstract word] OR aspirat?[title/abstract word] OR biops?[title/abstract word] OR tbna[title/abstract word])

AND

NOT (animal! NOT human/de)

Figure S1B.

lung neoplasms! OR lung cancer! OR bronchial neoplasms/de OR bronchus cancer! OR carcinoma, non-small-cell lung/de OR lung non small cell cancer/de OR carcinoma!

OR

Neoplasms, squamous cell! OR (carcinoma, squamous cell! AND lung!) OR squamous cell carcinoma/de

OR

(Mediastinum! and neoplasms!) OR mediastinal neoplasms/de OR (mediastinum! AND neoplasm!) OR mediastinum cancer!

OR

lung(W)cancer?/ti,ab OR non(W)small(W)cell/ti,ab OR pulmonary(N)neoplasm?/ti,ab OR pulmonary(N)malignan?/ti,ab OR lung(N)neoplasm?/ti,ab OR lung(N)malignan?/ti,ab

AND

Neoplasm staging/de OR cancer staging/de OR lymphatic metastasis/de OR lymph node metastasis/de OR lymph nodes! OR lymph node! OR lymphatic diseases! OR lymphatic system disease!

OR

Stag????/ti,ab OR lymph(W)node/ti,ab OR lymphatic/ti,ab

AND

Bronchscopy/de OR bronchoscope?/ti,ab OR transbronchial/ti,ab OR tbna/ti,ab

AND

Biopsy, needle/de OR needle biopsy/de OR biopsy! OR sentinel lymph node biopsy/de OR (fine(W)needle?/ti,ab OR aspirat?/ti,ab OR biops?/ti,ab OR tbna/ti,ab)

AND

NOT (animal! Not human/de)

Figure S2.

A. Index test technical quality (by test)

1. Was the TBNA technique described in sufficient detail to reproduce the procedure? (Reference to earlier published work that includes a complete technical description is allowable.)

2. Was the type and size of needle used to perform the TBNA noted?

3. Were the following procedure(s) in place to avoid contamination:
   - Was TBNA performed prior to: (1) all brushings, washings and endobronchial biopsies AND (2) inspection of distal airways?

   - Was a separate needle used for each biopsy/aspirate?

   - Was suctioning avoided to prevent aspiration of contaminated respiratory secretions (e.g. was the bronchoscope introduced into the endobronchial tree without suctioning and was suction released prior to withdrawal of the needle)?

   - Were aspirates containing an abundance of columnar epithelial cells, few abnormal cells (e.g. scarce malignant cells) and few lymphocytes considered 'negative' aspirates?

4. Were physician(s) performing TBNA experienced in the procedure as noted by:
   - Were all TBNA's performed by or under the direct supervision by a trained attending physician?

5. Was computed tomography (CT) assessment of mediastinal lymphadenopathy appropriate as noted by:
   - Were significant mediastinal lymph nodes defined as >1 cm in the short axis diameter?

   - Was IV contrast used during the scan of the pulmonary hila?

   - Was needle placement based on measurements taken from the CT scan?

   - Were accessible nodes defined as being within 1 cm or less from the tracheal wall?

   - If TBNA was performed blindly (e.g. CT or other imaging study was not performed or results of CT or other imaging study were negative for lymphadenopathy), were TBNA's performed (at a minimum) at the side of the tracheal carina AND/OR regional mediastinum corresponding to the primary tumor?

6. Were TBNA samples collected for BOTH cytological and histological review?

A2. Index test quality – If TBNA was performed using 'real-time' imaging (e.g. CT Fluoroscopy, enbronchial ultrasound or transthoracic ultrasound):

- Was the imaging device and procedure used to locate lymph nodes clearly described in detail?

- Was an image obtained to confirm needle position prior to aspiration/biopsy?

- Were accessible nodes defined as being within 1 cm or less from the tracheal wall?

B. Reference test quality (by study)

1. Were both positive and negative TBNA staging results compared to a gold standard (a surgical staging procedure with lymph node evaluation/dissection and biopsy/surgical pathologic review)?

2. Was the confirmatory surgical staging procedure clearly described?

C. Application of reference test (by study)

1. Did all patients with a NEGATIVE TBNA undergo a surgical staging procedure with biopsy (e.g. cervical or parasternal mediastinal exploration AND/OR mediastinal exploration at thoracotomy)?

2. Did all patients with a NEGATIVE TBNA undergo a thoracotomy with systematic sampling of normal and abnormal lymph nodes at all accessible lymph node stations?

D. Independence of test interpretation

1. Was the pathologist blinded to the histologic diagnosis (either prior to TBNA or after confirmatory surgical staging procedure) before the cytology from the TBNA was reviewed?

2. If patients underwent surgical confirmation of mediastinal lymphadenopathy, was the surgeon blinded to the results of the TBNA?

3. If a CT was performed prior to TBNA, was the radiologist blinded to the patient's clinical status?

E. Clinical description & characteristics

1. Did the study include complete demographic information as per the following:
   - Age of patients enrolled (either overall or per individual subject) was noted?

   - Sex of patients enrolled (either overall or per individual subject) was noted?

2. Was the individual subject used as the unit of analysis (e.g. not aspiration samples)?

3. Were inclusion and exclusion criteria clearly specified?

E. Cohort assembly

1. Were subjects enrolled prospectively?

2. Were subjects enrolled consecutively?

3. Was the relevant cohort assembled as per the following:
   - Did all enrolled patients have a pathologic diagnosis of NSCLC?

   - Were patients with evidence of nonresectability EXCLUDED (e.g. evidence of metastatic disease, malignant pleural effusions, etc.,)?

   - Were patients at risk for a false positive result EXCLUDED (e.g. evidence of tumor within 2 cm of the carina)?

4. Was there NO evidence of workup bias (e.g. were only patients with a positive TBNA enrolled)?

5. Were patients with previous lymph node biopsy attempts excluded (e.g. negative TBNAs or other biopsy procedure prior to study)?

6. Was this a multi-center trial?

F. Sample size

1. Did at least 35 participants with NSCLC have mediastinal lymph node involvement?

2. Did at least 35 participants with NSCLC have NO mediastinal lymph node involvement?

Figure S3.

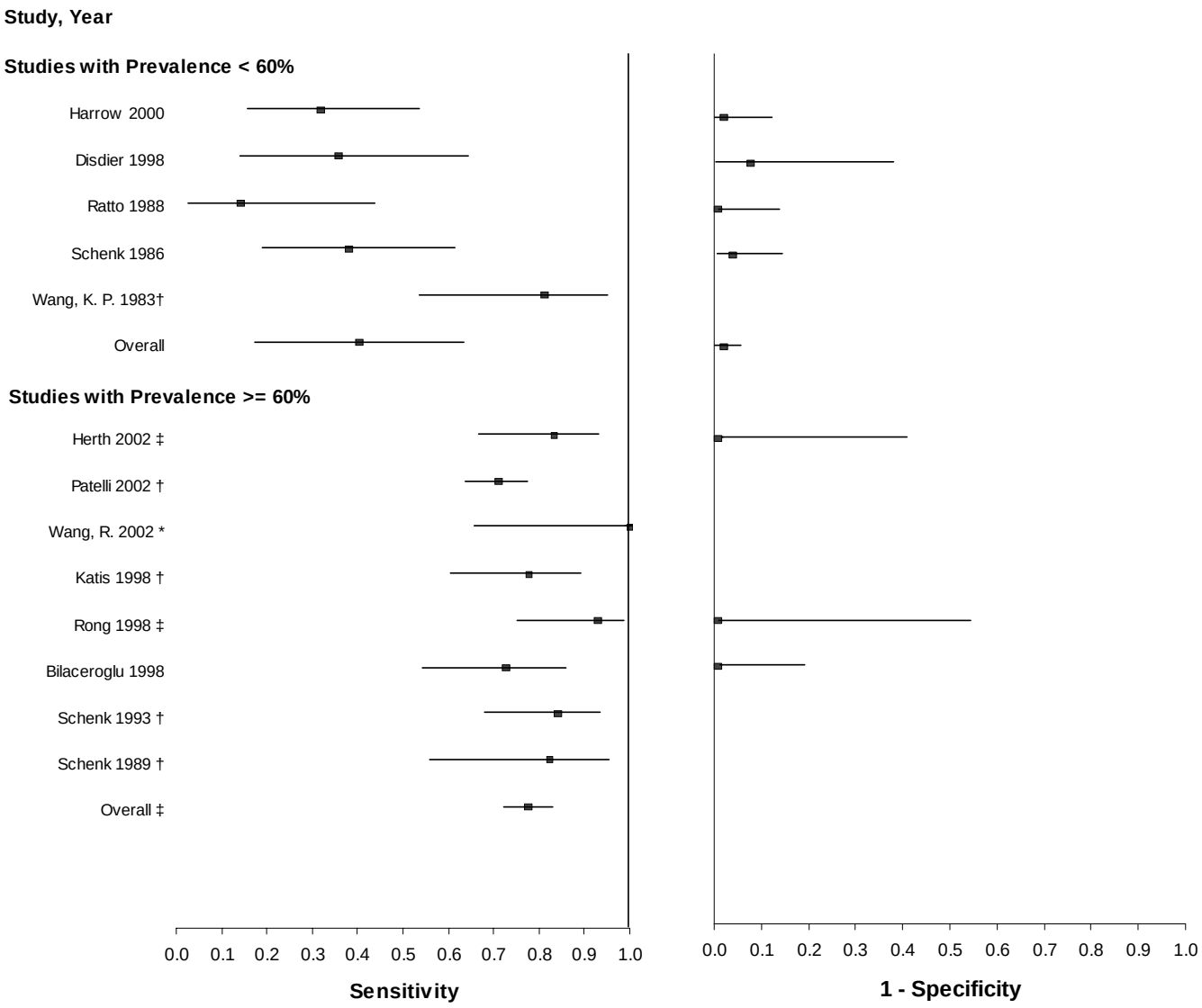
Study, Year
Studies with Prevalence < 60%
Harrow 2000
Disdier 1998
Ratto 1988
Schenk 1986
Wang, K. P. 1983†
Overall
Studies with Prevalence >= 60%
Herth 2002 ‡
Patelli 2002 †
Wang, R. 2002 *
Katis 1998 †
Rong 1998 ‡
Bilaceroglu 1998
Schenk 1993 †
Schenk 1989 †
Overall ‡
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Sensitivity
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
1 - Specificity

Figure S4.

Log odds ratio versus standard error