


BMJ Open Learning curves and association of pathologist's performance with the diagnostic accuracy of linear endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA): a cohort study in a tertiary care reference centre

Javier Flandes,¹ Luis Fernando Giraldo-Cadavid ^{2,3}, Maria Teresa Perez-Warnisher,⁴ Andres Gimenez,⁵ Iker Fernandez-Navamuel,⁵ Javier Alfayate,⁵ Alba Naya,⁶ Pilar Carballosa,⁶ Elena Cabezas,⁶ Susana Alvarez,⁵ Ana Maria Uribe-Hernandez,⁷ Luis Seijo⁸

To cite: Flandes J, Giraldo-Cadavid LF, Perez-Warnisher MT, *et al.* Learning curves and association of pathologist's performance with the diagnostic accuracy of linear endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA): a cohort study in a tertiary care reference centre. *BMJ Open* 2022;**12**:e051257. doi:10.1136/bmjopen-2021-051257

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-051257>).

Received 24 April 2021

Accepted 28 September 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Luis Fernando Giraldo-Cadavid;
luisf.giraldo@unisabana.edu.co

ABSTRACT

Objectives We aimed to assess the learning curves and the influence of the pathologist's performance on the endobronchial ultrasound transbronchial needle aspiration's (EBUS-TBNA's) diagnostic accuracy in a real-world study.

Design/setting Cohort study conducted in a tertiary care university hospital (single centre) with patients referred for EBUS-TBNA.

Participants/intervention We initially evaluated 376 patients (673 lymph nodes), 368 (660 lymph nodes) of whom were recruited. The inclusion criterion was EBUS-TBNA indicated for the study of mediastinal or hilar lesions. The exclusion criteria were the absence of mediastinal and hilar lesions during EBUS confirmed by a normal mediastinum and hilum on chest CT (except in cases of mediastinal staging of cancer) and lost to follow-up.

Primary and secondary outcome measures Diagnostic accuracy and related outcomes.

Methods We included patients from a prospectively constructed database. We performed a logistic regression multivariate analysis to adjust for potential confounders of the association between pathologist performance and EBUS-TBNA accuracy. The Cumulative Summation (CUSUM) analysis was used to assess pathologists' performance and learning curves.

Results Most indications for EBUS were suspicion of malignancy, including intrathoracic tumours (68.3%), extrathoracic tumours (9.8%) and cancer staging (7.0%). The patients' mean age was 63.7 years, and 71.5% were male. Overall EBUS-TBNA accuracy was 80.8%. In the multivariate logistic regression model, the factors independently associated with EBUS-TBNA accuracy included certain pathologists (ORs ranging from 0.16 to 0.41; $p < 0.017$), a lymph node short-axis diameter < 1 cm (OR: 0.36; 95% CI 0.21 to 0.62; $p < 0.001$), and the

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ We prospectively generated endobronchial ultrasound (EBUS) database and performed a multivariate regression analysis to control confounding bias.
- ⇒ The characteristics of our institution, which include the participation of pathologists harbouring variable levels of experience, allowed us to compare different levels of experience and plot The Cumulative Summation (CUSUM) curves to determine the number of cases required to attain competence.
- ⇒ As with any observational study, our study is exposed to the risk of confounding bias.
- ⇒ Due to the invasive nature of thoracic surgery, surgical biopsies (the best reference standard) were performed only in cases of non-diagnostic EBUS or for therapeutic purposes.

aetiology of lymph node enlargement (ORs ranging from 7 to 37; $p < 0.001$). CUSUM analysis revealed four different learning curve patterns, ranging from almost immediate learning to a prolonged learning phase, as well as a pattern consistent with performance attrition.

Conclusions Pathologists' proficiency conditioned EBUS-TBNA accuracy. This human factor is a potential source of error independent of factors conditioning tissue sample adequacy.

INTRODUCTION

Since the development of linear endobronchial ultrasound transbronchial needle aspiration (EBUS-TBNA) in the late 1990s,¹⁻³ the technique has become the standard of care in the mediastinal staging of lung cancer, extrathoracic tumours⁴⁻⁶ and the detection

and workup of radiologically occult mediastinal spread of lung cancer.^{4–8} A growing body of literature supports EBUS-TBNA as the initial procedure for the mediastinal staging of lung cancer and other mediastinal tumours.^{3,9} However, the accuracy of the procedure varies among the different studies published.^{5,8–10} Such variability might be explained by differences in the factors influencing its diagnostic accuracy.¹¹ The search for variables conditioning EBUS-TBNA accuracy has focused primarily on factors affecting the quality of the samples obtained, including lymph node size,^{12–14} nodal station,^{12,13} bronchoscopist expertise,^{13,15,16} presence of a trainee,¹⁷ needle type,^{13,18} the indication of the procedure,^{10,14} type of sedation,¹⁰ number of needle passes,¹⁰ patient's age¹⁹ and CT findings.^{7,8,20}

The experience of the pathologist was found to be a potential source of variability in a study combining slides from oesophageal endoscopic ultrasound fine needle aspiration (EUS-FNA) and EBUS-TBNA.²¹ However, the pathologists' learning curve patterns for EBUS-TBNA and the role of pathologist's expertise and skills in the accuracy of EBUS-TBNA have not been formally studied, in spite of the potential importance of this human factor.

We hypothesised that some degree of EBUS-TBNA accuracy should be attributable to the pathologist's proficiency (including their learning curves) and aimed to assess it in a cohort of patients referred for EBUS-TBNA at a tertiary care university hospital, adjusting for other variables potentially affecting EBUS-TBNA accuracy.

METHODS

A cohort was consecutively recruited from a prospectively constructed database of patients referred for EBUS-TBNA at a tertiary care university hospital from January 2010 to March 2015. Each patient was followed up for at least 3 years, obtaining his/her information from their electronic medical records. The inclusion criterion was EBUS-TBNA indicated for the study of mediastinal or hilar lesions. The exclusion criteria were the absence of mediastinal and hilar lesions during EBUS confirmed by a normal mediastinum and hilum on chest CT (except in cases of mediastinal staging of cancer) and lost to follow-up. Sixteen pathologists were included in the study from the beginning of their EBUS-TBNA cytopathology learning curve (they had no additional certifications in cytopathology or lung cancer). Four senior bronchoscopists obtained the EBUS-TBNA samples; they received formal training in EBUS and had an experience of >100 EBUS (two SD above the mean number of 43 EBUS needed to attain proficiency.^{22,23} Further, 31 bronchoscopist trainees also participated in the EBUS-TBNAs. We used numeric codes to anonymise the individual bronchoscopists and pathologists participating in the study for blinding purposes. The experience of both bronchoscopists and pathologists was quantified as the number of EBUS-TBNA performed or reported, respectively.

Endobronchial ultrasound transbronchial needle aspiration

All EBUS-TBNAs were performed on an outpatient basis, with at least one senior EBUS bronchoscopist present and one trainee bronchoscopist assisting. The procedures were performed under conscious sedation with either midazolam and fentanyl or general anaesthesia, adding local lidocaine to the airways. The linear EBUS bronchoscope (BF-UC 180F, Olympus, Tokyo, Japan) used had an outer diameter of 6.9 mm, a 2 mm working channel, a 30° viewing lens and a convex ultrasound transducer placed at the distal tip covered with an inflatable balloon. An Olympus ultrasound processor (EU-ME2) was used for image processing. The bronchoscope was introduced orally in all cases using a mouthpiece to protect the endoscope. During the mediastinal staging of tumours, lymph nodes were sampled if they had a short axis diameter >5 mm or a diameter <5 mm with at least one sign of malignancy.^{9,24} TBNA was performed using the jabbing technique²⁵ under real-time EBUS guidance (Scanning frequency: 7.5 MHz) with 21G needles (Olympus NA-201SX-4021), applying –15 cm H₂O negative pressure, and moving the needle back and forth 15 to 20 times inside the lesion. Three to four passes were performed per lymph node, ensuring that at least one histological core was obtained. Rapid on-site cytological evaluation was not used. Doppler ultrasound was used as needed to avoid puncturing the blood vessels.

Three kinds of samples were available for analysis from each puncture: a cellblock was initially obtained using the internal stylet to release the core from the needle and fixed immediately in formalin, then the contents of the needle were expressed on cytological glass slides using a syringe and fixed, and a needle lavage using saline was performed to obtain more samples for cytological analysis. The pathologists had access to the same clinical information as in real-world clinical practice, as recommended for diagnostic accuracy studies.^{26,27}

Reference standard

Specific diagnoses from EBUS-TBNA or, in cases of mediastinal staging of lung cancer, an abundance of lymphocytes were considered diagnostic EBUS-TBNA. EBUS-TBNA was considered non-diagnostic in the absence of a specific diagnosis or, in cases of the mediastinal staging of lung cancer, the absence of a specific diagnosis, plus a scarcity of lymphocytes. In cases of non-diagnostic EBUS, the final diagnosis was established by open thoracotomy, thoracoscopy or mediastinoscopy. Additionally, in those patients who underwent surgery after an adequate EBUS-TBNA sample (ie, therapeutic surgery for lung tumour resection), the final diagnosis was established based on the histological examination of the surgical specimen. Patients were also followed up for at least 3 years with clinical and imaging examinations to assess if such follow-up data were concordant or discordant with the EBUS results. Therefore, the reference standard to classify the cases as true positives (TP), true negatives (TN), false positives (FP) and false negatives

(FN) was built using surgical, clinical and imaging information during the follow-up, and their definitions are detailed below.

A TP was defined as a TBNA cytology showing a specific diagnosis (cancer or a benign condition) undisputed by the findings of surgical biopsy/resection, follow-up or postmortem analysis, or in the case of benign disease follow-up and imaging findings consistent with the benign TBNA cytological diagnosis. A TN was defined as a TBNA cytology negative for cancer or other conditions but showing a lymphocyte fund, undisputed by the findings of surgical biopsy/resection, postmortem analysis or follow-up. An FP was defined as TBNA cytology compatible with cancer, or a specific benign condition disputed by further definitive testing, such as surgical biopsy/resection or postmortem analysis, showing a different diagnosis. An FN was defined as non-diagnostic TBNA, or inconclusive TBNA cytology (including suspicious for malignancy reports) with subsequent biopsy/resection, follow-up or postmortem analysis compatible with cancer or a specific diagnosis. We also considered EBUS procedural failures with the impossibility of obtaining EBUS-TBNA samples as FN.

Patient and public involvement

It was not possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Statistical analysis

Information about variables potentially affecting EBUS-TBNA accuracy (eg, confounders) was collected, including age, sex, lymph node size, nodal station, bronchoscopist (including his/her experience), pathologist (including his/her experience), indication for the procedure, chest CT and fluorodeoxyglucose positron emission tomography CT (PET-CT) findings (independent variables). Missing values were imputed using multiple imputation (data were missing in 5.9% of cases, there were no missing data in dependent variables, pathologists and bronchoscopists).

Categorical variables were reported as absolute frequencies and percentages; continuous variables were reported with means±SD or medians and IQR (IQR: 25th to 75th percentile) depending on their distribution. Bivariate analyses for categorical variables were performed using the χ^2 test (exact method), while the Student's t-test or the Mann-Whitney U test were used for continuous variables depending on their distribution. Normality was tested using the Kolmogorov-Smirnov test. Overall accuracy was calculated by the equation:
$$\frac{(TP+TN)}{(TP+FP+FN+TN)}$$

EBUS-TBNA was considered a diagnostic success (diagnostic EBUS) in the case of a TP or a TN result and considered a diagnostic failure in the case of an FP or an FN.

A multivariate binary logistic regression model was built using diagnostic EBUS as the dichotomous dependent variable and introducing as independent variables

all other variables potentially affecting EBUS accuracy. Each individual pathologist was compared with a reference group of pathologists with a diagnostic yield >80% (diagnostic yield ≤80% was our threshold for an unacceptable failure rate). Potentially confounding variables were selected based on biological plausibility and having an association with the dependent and independent variables under study (see details in online supplemental appendix). All associations were considered statistically significant at a two-tailed $p < 0.05$.

We used binary Cumulative Summation (CUSUM) analysis to assess pathologist performance (including their learning curve), which requires setting an acceptable failure rate (level of error if the procedure is carried out correctly) and an unacceptable failure rate (maximum acceptable level of error). We considered an acceptable failure rate of 10% (accuracy=90%) and an unacceptable failure rate of 20% (accuracy=80%) and defined a type I error (odds of falsely classifying someone as incompetent, designated α) of 0.1 and a type II error (odds of falsely classifying someone as competent, designated β) of 0.1.^{15 28 29}

We constructed binomial CUSUM charts using Microsoft Excel 2013 (Microsoft Corporation, Redmond, Washington, USA). This method consists of the cumulative sum of failure minus success with each case. We drew the CUSUM curve plotting the cumulative sum score after each case (y axis) versus the index number of that case (x axis). Consecutive errors drive the CUSUM curve upward, while consecutive success drives the CUSUM curve downward. The CUSUM graph includes horizontal lines called decision limits (h_1 and h_0), which are the boundaries of an acceptable or unacceptable error rate and are calculated based on the risk of α and β errors. When the CUSUM curve crosses a decision limit from above, it is inferred that the failure rates were within the predetermined acceptable rate of 10% (good performance); when the CUSUM curve crosses a decision limit from below, it is inferred that the failure rates have reached the predetermined unacceptable rate of 20% (bad performance); if the CUSUM curve remained between two decision limits, continued observation is indicated (stable performance). Therefore, competence is assumed when the CUSUM curve slopes downward or remains stable, but when the curve slopes upward, it indicates a below-than-acceptable success rate (see details in online supplemental appendix).

We estimated that a minimum sample size of 600 lymph nodes and 300 patients would be required for the logistic regression multivariate analysis with the following considerations: at least 10 events (diagnostic failures) per variable included in the model,^{30 31} the multivariate model having 12 variables, that is, a total of 120 events required, an estimated overall accuracy of 80% (20% of diagnostic failures), and an average of 2 nodes studied per patient. Even though CUSUM charts do not require a specific sample size for the continuum control of someone's performance,³² to avoid the risk of prematurely

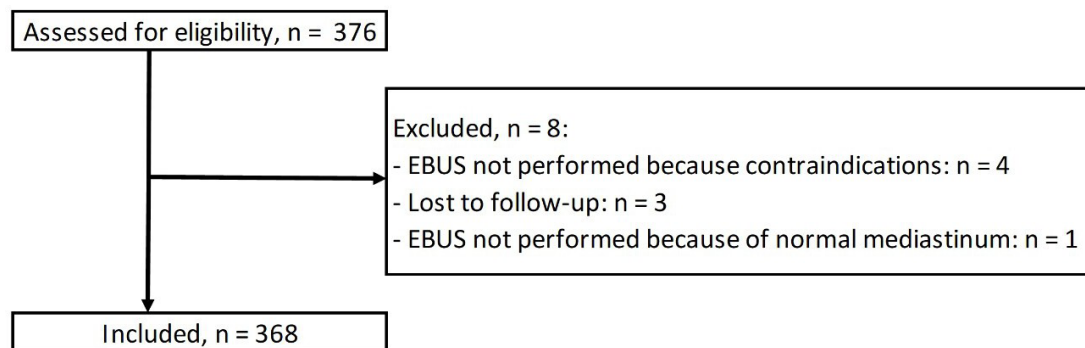


Figure 1 Participant flow diagram. EBUS, endobronchial ultrasound.

classifying someone as competent at the beginning of the learning curve (type II error), we modelled the CUSUM curve of a pathologist with perfect performance, finding that with our CUSUM analysis assumptions a minimum of 19 EBUS were required to obtain a statistically significant perfect performance. Therefore, to determine the number of EBUS pathology reports required to attain competence, we only analysed pathologists who had read more than 19 EBUS.

Statistical analyses were performed using Microsoft Excel 2013 (Microsoft Corporation), MedCalc V.16.8 (MedCalc Software bvba, Ostend, Belgium) and IBM-SPSS Statistics V.22.

RESULTS

A total of 376 patients (673 lymph nodes) were initially evaluated for the study, from which 368 patients (660 lymph nodes) using EBUS-TBNA were recruited (figure 1). The mean age was 63.7 ± 12.9 years, and 71.5% of the patients were men. Most procedures were cancer-related: 451 (68.3%) for intrathoracic tumours, 65 (9.8%) for extrathoracic tumours and 46 (7.0%) for the staging of existing biopsy-proven cancer. Other indications included suspicion of infectious (6/660; 0.9%) or inflammatory (51/660; 7.7%) diseases. The median short-axis diameter of the studied lymph nodes was 13.5 mm (IQR 10.0–20.0). The most frequently sampled stations were 7 (29.5%) and 4R (24.8%). EBUS-TBNA was the only diagnostic test for the 428 (64.8%) lymph nodes sampled. Additional testing included mediastinoscopy for 98 lymph nodes (14.8%), surgical biopsy for 40 lymph nodes (6.1%) and biopsy of an alternative site in 24 (3.6%) (table 1).

Indications for additional procedures included therapeutic resection of tumours and diagnostic uncertainty. The most common final diagnoses after combining EBUS-TBNA results with additional testing were 215 (32.6%) non-small cell lung cancers (NSCLC), 83 (12.6%) other cancers and 154 (23.3%) normal lymph nodes (table 1). No major EBUS complications were reported.

Overall, EBUS-TBNA accuracy was 80.8%, achieving a diagnosis in 533/660 lymph nodes sampled (127 diagnostic failures). We had four FP cases; two of them were

reported as anthracosis on EBUS and the final diagnosis was squamous cell carcinoma, one was reported as anthracosis on EBUS and the final diagnosis was undifferentiated carcinoma and one was reported as undifferentiated carcinoma and the final diagnosis was adenocarcinoma. The variables associated with EBUS accuracy in bivariate analysis included nodal size and station, some pathologists, some bronchoscopists and final diagnosis (table 2).

In the multivariate binary logistic regression analysis, we found that the pathologist interpreting EBUS-TBNA results was independently associated with diagnostic accuracy after adjusting for potential confounders (including pathologist experience quantified as the number of EBUS-TBNA cases evaluated) (table 3). Four pathologists were significantly associated with lower accuracy in the bivariate and multivariate models (adjusting for confounders, including node size, location, bronchoscopist, EBUS indication, final diagnosis): #6 (OR 0.27; 95% CI 0.11 to 0.68), #7 (OR 0.32; 95% CI 0.16 to 0.62), #9 (OR 0.41; 95% CI 0.20 to 0.85), #15 (OR 0.16; 95% CI 0.07 to 0.37). The general characteristics of the cases reported by good and poor performing pathologists were similar (see online supplemental table 1). Lower accuracy was also correlated with a lymph node size <1 cm (OR 0.36; 95% CI 0.21 to 0.62), trainee endoscopist #9 (OR 0.28; 95% CI 0.10 to 0.75) and lymph node station 11R (OR 0.22; 95% CI 0.08 to 0.62). A final diagnosis of granulomatous disease was associated with a trend for lower accuracy (OR 0.28; 95% CI 0.08 to 1.00), while greater accuracy was associated with lymph node station 4R (OR 1.99; 95% CI 0.08 to 0.62), the experience of the senior bronchoscopist (OR 1.003; 95% CI 1.001 to 1.005), trainee endoscopist #13 (OR 7.7; 95% CI 1.02 to 58.86), a diagnosis of NSCLC (OR 7.07; 95% CI 3.89 to 12.84), malignancy in general (OR 37.1; 95% CI 8.39 to 163.97), as well as normal appearing lymph nodes (OR 8.01; 95% CI 4.01 to 16.01). Pathologist experience with TBNA sampling was associated with diagnostic accuracy only in the better performance quartile of pathologists (OR 1.03; 95% CI 1.01 to 1.06), the detailed information about pathologists' experience can be found in online supplemental table 2. The senior bronchoscopists performing

Table 1 General characteristics of the cohort

N (Nodes/patients)		660/368	
Sex			
Male, n (%)		263	(71.5%)
Age (years), mean (SD)		63.7	(12.9)
Indications of EBUS-TBNA	Adenopathies (intrathoracic cancer), n (%)	451	(68.3%)
	Adenopathies (extrathoracic cancer), n (%)	65	(9.8%)
	Inflammatory adenopathies, n (%)	51	(7.7%)
	Cancer staging, n (%)	46	(7.0%)
	Infectious adenopathies, n (%)	6	(0.9%)
	Histological re-evaluation of cancer, n (%)	2	(0.3%)
	Other, n (%)	39	(5.9%)
Adenopathy size, median (IQR)		13.5	(10–20)
PET-CT	Negative, n (%)	61	(9.2%)
	Positive (SUV max>2.5), n (%)	245	(37.1%)
	Not performed, n (%)	354	(53.6%)
Nodal station	7, n (%)	195	(29.5%)
	4R, n (%)	164	(24.8%)
	10R, n (%)	102	(15.5%)
	4L, n (%)	62	(9.4%)
	10L, n (%)	47	(7.1%)
	2R, n (%)	24	(3.6%)
	11R, n (%)	18	(2.7%)
	11L, n (%)	16	(2.4%)
	12R, n (%)	7	(1.1%)
	2L, n (%)	5	(0.8%)
	8, n (%)	4	(0.6%)
	12L, n (%)	2	(0.3%)
	5, n (%)	2	(0.3%)
	3, n (%)	1	(0.2%)
Additional tests for final diagnosis	Mass	11	(1.7%)
	None, n %	428	(64.8%)
	EBUS-TBNA repeated, n %	12	(1.8%)
	Mediastinoscopy, n %	98	(14.8%)
	Follow-up by PET-CT n %	39	(5.9%)
	Surgical biopsy, n %	40	(6.1%)
	Bone biopsy, n %	6	(0.9%)
	Other biopsies, n %	18	(2.7%)
	Electromagnetic navigation bronchoscopy, n %	7	(1.1%)
	Conventional bronchoscopy with biopsies, n %	2	(0.3%)
	CT guided transthoracic needle aspiration, n %	3	(0.5%)
Fine needle aspiration of neck adenopathies, n %	3	(0.5%)	
Upper gastrointestinal endoscopy, n %	4	(0.6%)	

Continued

Table 1 Continued

Final diagnosis	Lung cancer, non-small cell, n %	215	(32.6%)
	Lung cancer, small cell, n %	27	(4.1%)
	Other cancer, n %	83	(12.6%)
	Normal, n %	154	(23.3%)
	Other benign conditions, n %	22	(3.3%)
	Sarcoidosis, n %	19	(2.9%)
	Granulomas non-sarcoidosis non-tuberculosis, n %	14	(2.1%)
	Lymphoma, n %	10	(1.5%)
	Tuberculosis, n %	4	(0.6%)
	Insufficient material for diagnosis, n %	108	(16.4%)
	Inconclusive histology, n %	4	(0.6%)
Complications	None, n (%)	662	(99.8%)
	Oxygen desaturation, n (%)	1	(0.2%)
Anaesthetic modality	Conscious sedation, n (%)	660	(99.5%)
	General anaesthesia, n (%)	3	(0.5%)
	Local anaesthesia, n (%)	0	(0.0%)

EBUS-TBNA, endobronchial ultrasonography transbronchial needle aspiration; PET-CT, positron emission tomography CT scan; SUV max, maximum standardised uptake value.

the EBUS-TBNA were not associated with EBUS-TBNA accuracy after adjusting for confounders ($p>0.23$).

CUSUM analysis for eight pathologists with a cumulative experience of more than 19 EBUS-TBNA cases revealed that the median number of procedures needed to reach competence (defined as $<10\%$ of error: accuracy $>90\%$) was 32 (IQR: 26–46). The CUSUM curves showed four main performance patterns: an almost immediate achievement of competence exemplified by pathologist 13 (short learning phase pattern), observed in 14% of pathologists; a pattern of continuous unacceptable failures (CUSUM curve never crossing the acceptable threshold), as exemplified by pathologist 15 (prolonged learning phase pattern), observed also in 14% of pathologists; a pattern shared by most pathologists (43% of them) of a stable phase of acceptable performance following the initial learning curve (intermediate learning phase pattern). An initial pattern of competence followed by deteriorating performance, as exemplified by pathologists 1 and 7 (attrition performance pattern), observed in 28% of pathologists (figure 2).

DISCUSSION

In our cohort study, pathologists showed four disparate learning curve patterns, ranging from an almost immediate achievement of competence to performance attrition. Additionally, EBUS-TBNA accuracy was conditioned by the pathologist interpreting the lymph node sampling. Such pathologists' influence was independent of other factors, including the pathologist's experience with the procedure and potential confounders. Although one

previous study looked at the performance of cytopathologists in abdominal lesions studied using EUS-FNA, to our knowledge, this is the first study to investigate how pathologists might condition EBUS-TBNA accuracy in the study of mediastinal and hilar lesions.^{29 33} We also confirmed that lymph node size and station, bronchoscopist skills and underlying cause of nodal enlargement influence test accuracy, as described in previous studies of EBUS-TBNA.^{1 34–36}

One potential explanation for some pathologists having poorer results is that they were sent cases that were more difficult to report, including smaller lymph node sizes and more difficult stations, which could act as potential confounders. To control for such potential confounders of pathologist's performance, we carried out a multivariate logistic regression analysis adjusting for confounders such as lymph node size, station, final diagnosis, bronchoscopist and bronchoscopist's experience, and we found that the lower diagnostic accuracy rates of some pathologists were independent of such confounders and even of their experience with the procedure. This finding points to sources of human error in the interpretation of EBUS-TBNA results dependent on the pathologist's proficiency and that are not predicted by the pathologist's experience, which should be the focus of future studies.

The pathologists' CUSUM curves revealed four different patterns, including three previously reported by Kemp *et al*¹⁵ in a study focusing on the learning curves of EBUS-TBNA bronchoscopists. Kemp identified a bronchoscopist who attained competence almost immediately, mimicking our pathologist's short learning phase pattern.

Table 2 Variables associated with EBUS-TBNA diagnostic accuracy (bivariate analysis)

Variable	Diagnostic success, N=533		Diagnostic failure, N=127		OR	95% CI	P value
Nodal station 10L, n (%)	32	(6%)	15	(12%)	0.477	(0.250 to 0.910)	0.025
Nodal station 10R, n (%)	73	(14%)	29	(23%)	0.536	(0.331 to 0.869)	0.011
Nodal station 11R, n (%)	10	(2%)	8	(6%)	0.284	(0.110 to 0.736)	0.010
Nodal station 4R, n (%)	146	(27%)	18	(14%)	2.285	(1.340 to 3.896)	0.002
Size <1 cm, n (%)	87	(16%)	33	(26%)	0.540	(0.340 to 0.858)	0.009
Pathologist #6*, n (%)	38	(7%)	14	(11%)	0.399	(0.201 to 0.791)	0.009
Pathologist #7*, n (%)	73	(14%)	21	(17%)	0.511	(0.288 to 0.906)	0.022
Pathologist #9*, n (%)	55	(10%)	19	(15%)	0.425	(0.232 to 0.778)	0.006
Pathologist #15,* n (%)	43	(8%)	24	(19%)	0.263	(0.146 to 0.473)	<0.001
Trainee bronchoscopist #12, n (%)	154	(29%)	52	(41%)	0.589	(0.395 to 0.879)	0.010
Senior bronchoscopist #2, n (%)	269	(50%)	45	(35%)	1.857	(1.243 to 2.774)	0.003
Senior bronchoscopist #4, n (%)	25	(5%)	14	(11%)	0.397	(0.200 to 0.788)	0.008
Lung cancer (NSCLC), n (%)	194	(36%)	21	(17%)	2.889	(1.752 to 4.764)	<0.001
Other cancer (non-lung cancer), n (%)	80	(15%)	3	(2%)	7.299	(2.266 to 23.509)	0.001
Normal lymph node, n (%)	141	(26%)	13	(10%)	3.154	(1.722 to 5.777)	<0.001
Granulomatous disease (non-sarcoidosis, non-TB), n (%)	7	(1%)	7	(6%)	0.228	(0.079 to 0.663)	0.007
Size mm, median (IQR)	14	(10–20)	12	(9–18)	1.554	(1.033 to 2.337)	0.034
Senior bronchoscopist experience (in number of EBUS), median (IQR)	330	(218–458)	317	(180–429)	1.001	(1.000 to 1.003)	0.050

A normal lymph node was defined as a TBNA cytology negative for cancer or other conditions, but showing a lymphocyte fund, undisputed by the findings of surgical biopsy/resection, postmortem analysis or follow-up.

*The reference group was pathologists with a diagnostic accuracy >80%; bivariate analysis performed using binary logistic regression at a significance level of $p < 0.05$ (two-tailed).

EBUS-TBNA, endobronchial ultrasound transbronchial needle aspiration; NSCLC, non-small cell lung cancer; TB, tuberculosis.

However, he also identified two bronchoscopists with CUSUM curves showing a relentless rise, similar to our prolonged learning phase pattern indicating a greater difficulty in reaching competence, as well as a bronchoscopist acquiring competence (crossing the decision limit from above) after 76 procedures, similar to our intermediate learning phase pattern. We did not find studies on EBUS-TBNA showing a pattern of performance attrition, such as the one identified in our study. Based on our results, we would suggest that EBUS should be read by cytopathology fellowship-trained pathologists only with a minimum experience in EBUS reporting of 30–40 cases.

Regarding bronchoscopists, we found an association between greater experience of the senior bronchoscopist and higher diagnostic accuracy, but there were no significant differences in diagnostic accuracy among the senior bronchoscopists. The current practice of EBUS-TBNA at our institution precludes building CUSUM curves for individual bronchoscopists because all EBUS-TBNA procedures are performed by at least one senior endoscopist and one trainee, making it impossible to measure the independent effect of each bronchoscopist on test accuracy. However, the bronchoscopists' learning curve for EBUS performance has been widely studied elsewhere.^{37–40} Most studies agree that the learning curve for

EBUS-TBNA is short, although the number of procedures needed to achieve an expert level varies. Our study found significant variations in the accuracy of EBUS-TBNA attributed to the participation of two trainees in the procedure. Trainee 9 was associated with lower accuracy (OR 0.28), while trainee 13 was associated with higher accuracy (OR 7.75). A previous US study found that 33% of pulmonary fellows did not achieve expert-level performance during training.⁴¹ Similarly, Nguyen *et al*¹⁷ found a 16% drop in the diagnostic yield of EBUS-TBNA when a trainee was present during the procedure. However, we found that some trainees may have improved diagnostic accuracy of the test. This is not an unexpected finding, since we are inevitably dealing with a procedure influenced by human error and skills.

Other variables that affected the diagnostic accuracy in our study were lymph node size (<1 cm with lower accuracy), nodal station (11R with lower accuracy and 4R with higher accuracy) and the underlying disease (malignancy and granulomatous disease were associated with higher and lower accuracy, respectively). All of these have been previously identified as key variables,^{134–36} so our findings support previous studies. Our overall diagnostic accuracy was lower for benign aetiologies, including granulomatous disease, which is also consistent with the literature,

Table 3 Variables associated with EBUS-TBNA diagnostic accuracy (multivariate analysis)

Variable	OR (adjusted)	95% CI	P value
Nodal size<1 cm	0.365	0.213 to 0.625	<0.001
Nodal station			
11R	0.221	0.078 to 0.625	0.004
4R	1.993	1.085 to 3.659	0.026
Senior bronchoscopist experience (in number de EBUS)	1.003	1.001 to 1.005	0.004
Pathologist experience in number of EBUS (quartile of pathologist with better performance)	1.032	1.008 to 1.057	0.009
Trainee bronchoscopist #9	0.276	0.101 to 0.754	0.012
Trainee bronchoscopist #13	7.748	1.020 to 58.858	0.048
NSCLC	7.068	3.890 to 12.843	<0.001
Other cancer (non-lung cancer)	37.092	8.391 to 163.969	<0.001
Normal lymph node	8.010	4.007 to 16.013	<0.001
Granulomatous disease (non-sarcoidosis, non-TB)	0.284	0.081 to 1.001	0.050
Pathologist 6 (vs reference category: pathologists with accuracy>80%)	0.271	0.108 to 0.684	0.006
Pathologist 7 (vs reference category: pathologists with accuracy>80%)	0.318	0.163 to 0.624	0.001
Pathologist 9 (vs reference category: pathologists with accuracy>80%)	0.411	0.198 to 0.853	0.017
Pathologist 15 (vs reference category: pathologists with accuracy>80%)	0.156	0.067 to 0.365	<0.001

Multivariate analysis performed using binary logistic regression at a significance level of $p < 0.05$ (two tailed).

A normal lymph node was defined as a TBNA cytology negative for cancer or other conditions, but showing a lymphocyte fund, undisputed by the findings of surgical biopsy/resection, postmortem analysis or follow-up.

EBUS-TBNA, endobronchial ultrasonography transbronchial needle aspiration; NLCLC, non-small cell lung cancer; TB, tuberculosis.

since EBUS may have lower diagnostic yield in benign mediastinal disease.^{37 42 43}

Our study has several limitations. First, it is an observational study exposed to the risk of confounding bias. Nevertheless, the main EBUS database was generated prospectively, and the variables known to condition EBUS-TBNA accuracy were recorded systematically after each procedure and included in our multivariate analysis for controlling confounding bias. However, the potential influence of an unmeasured confounding factor cannot be entirely ruled out in our study. PET-CT fluorodeoxyglucose uptake was available for 46.4% of the lymph nodes sampled before the procedure was performed. PET-CT findings have been shown to condition EBUS-TBNA diagnostic accuracy,^{36 43} but this limitation is also relatively common in real-world EBUS-TBNA procedures, as evidenced by the AQUIRE registry, with 46.1% of patients having PET-CT available.³⁶ Furthermore, we adjusted for it in our multivariate model. Additionally, the final number of pathologists included in this study was small and the study is limited to a single centre, these results may change in a study including a larger number of pathologists from multiple centres.

According to our definition of FP, intended to address the overall accuracy (for malignant and benign conditions) of the EBUS-TBNA, we had four FP cases: two of

them were reported as anthracosis on EBUS, and the final diagnosis was squamous cell carcinoma; one was reported as anthracosis on EBUS, and the final diagnosis was undifferentiated carcinoma; and one was reported as undifferentiated carcinoma, and the final diagnosis was adenocarcinoma. However, from the point of view of lung cancer staging, the dangerous FP are those cases diagnosed as malignant by EBUS-TBNA and then proved benign because such patients might have been denied the chance of a potentially curative surgical intervention, and we had not such dangerous FP.

Many of our patients who had a specific diagnosis from the EBUS-TBNA undisputed by clinical or imaging findings during the follow-up were not subjected to a surgical biopsy to confirm their diagnosis; this corresponds to a partial verification (verifying only part of the population with the reference standard).⁴⁴ Such a source of bias, which is common to real-world studies in which the reference standard is an invasive procedure (eg, thoracic surgery), has been found to have little influence on the diagnostic capacity of the test.⁴⁴ Nevertheless, we tried to limit its influence in our results, extending our clinical and imaging follow-up to 3 years. Despite these limitations, the characteristics of our institution, which includes the participation of pathologists harbouring variable levels of expertise and knowledge of cytopathology, allowed us

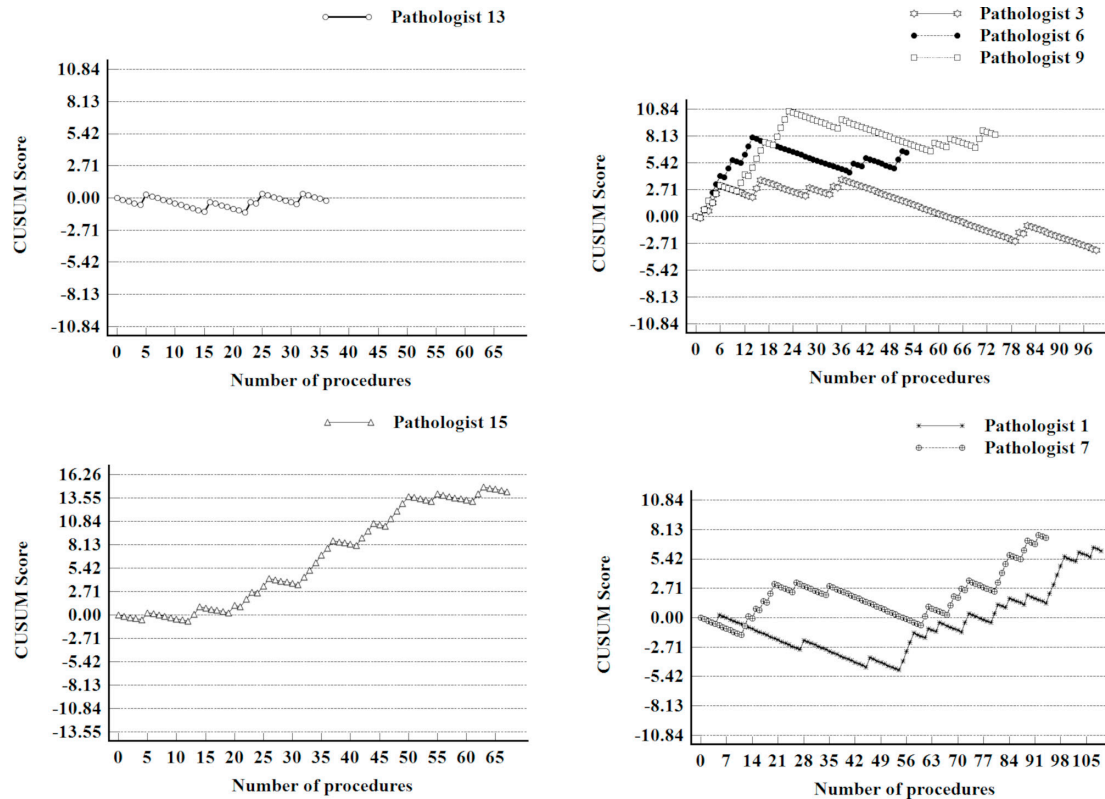


Figure 2 Cumulative Summation (CUSUM) curves of pathologist performance. CUSUM curves of pathologist performance showing four main patterns: short learning phase pattern (pathologist 13); intermediate learning phase pattern (pathologists 3, 6, 9); prolonged learning phase pattern (pathologist 15); performance attrition (pathologists 1 and 7).

to assess the effect of this human factor as a source of error independent of other confounders in multivariate analysis.

In conclusion, we found that the pathologists' learning curves follow four main patterns and that certain pathologists were associated with lower EBUS-TBNA diagnostic accuracy. Such a finding is intriguing since it is independent of the pathologists' experience with the procedure. Therefore, pathologist proficiency must be considered a potential source of human error conditioning EBUS-TBNA accuracy, despite efforts to obtain adequate tissue samples. However, considering that our study was performed in a single centre with a small number of pathologists, these results should not be regarded as definitive.

Author affiliations

¹Interventional Pulmonology, Fundacion Jimenez Diaz-UTE, Madrid, Spain

²School of Medicine, Research Department, Universidad de La Sabana, Chia, Colombia

³Interventional Pulmonology, Fundacion Neumologica Colombiana, Bogota, Colombia

⁴Pulmonary Medicine Departament, Clínica Universidad de Navarra, Madrid, Spain

⁵Interventional Pulmonology, Fundacion Jimenez Diaz, Madrid, Spain

⁶Pulmonary Medicine Departament, Fundacion Jimenez Diaz, Madrid, Spain

⁷Epidemiology and Biostatistics Departament, Universidad de La Sabana, Chia, Colombia

⁸Pulmonary Medicine, Clínica Universidad de Navarra, Madrid, Spain

Twitter Luis Fernando Giraldo-Cadavid @luisfgiraldo

Contributors LFG-C and MTP-W had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis, and are responsible for the overall content as guarantors. JF, AG, IF-N, JA, AN, PC, EC, SA, AMU-H and LS contributed substantially to the study design, data analysis and interpretation, and the writing of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests JF has received personal fees and non-financial support from Olympus, personal fees from Boston Scientific, personal fees from Pentax, all of them outside the submitted work. LFG-C has received professional fees for speaking engagements and industry funding from pharmaceutical companies not generating conflict of interests with this manuscript. MTP-W declares no conflict of interests related to this manuscript content. AG declares no conflict of interests related to this manuscript content. LS has received professional fees for speaking engagements and industry funding from pharmaceutical companies not generating conflict of interests with this manuscript.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Institutional Board of the Fundacion Jimenez-Diaz CEIm-FJD approved this study, approval reference ID number 04/19. Participants gave written informed consent to undergo EBUS-TBNA (including the use of their data on research).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability

of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Luis Fernando Giraldo-Cadavid <http://orcid.org/0000-0002-7574-7913>

REFERENCES

- Shannon JJ, Bude RO, Orens JB, *et al.* Endobronchial ultrasound-guided needle aspiration of mediastinal adenopathy. *Am J Respir Crit Care Med* 1996;153:1424–30.
- Herth F, Becker HD, Manegold C, *et al.* Endobronchial ultrasound (EBUS)-assessment of a new diagnostic tool in bronchoscopy for staging of lung cancer. *Onkologie* 2001;24:151–4.
- Okamoto H, Watanabe K, Nagatomo A, *et al.* Endobronchial ultrasonography for mediastinal and hilar lymph node metastases of lung cancer. *Chest* 2002;121:1498–506.
- Ge X, Guan W, Han F, *et al.* Comparison of endobronchial ultrasound-guided fine needle aspiration and video-assisted mediastinoscopy for mediastinal staging of lung cancer. *Lung* 2015;193:757–66.
- Adams K, Shah PL, Edmonds L, *et al.* Test performance of endobronchial ultrasound and transbronchial needle aspiration biopsy for mediastinal staging in patients with lung cancer: systematic review and meta-analysis. *Thorax* 2009;64:757–62.
- Criner GJ, Eberhardt R, Fernandez-Bussy S, *et al.* Interventional bronchoscopy. *Am J Respir Crit Care Med* 2020;202:29–50.
- Leong TL, Loveland PM, Gorelik A, *et al.* Preoperative staging by EBUS in cN0/N1 lung cancer: systematic review and meta-analysis. *J Bronchology Interv Pulmonol* 2019;26:155–65.
- El-Osta H, Jani P, Mansour A, *et al.* Endobronchial ultrasound for nodal staging of patients with non-small-cell lung cancer with radiologically normal mediastinum. A meta-analysis. *Ann Am Thorac Soc* 2018;15:864–74.
- Silvestri GA, Gonzalez AV, Jantz MA, *et al.* Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ED: American College of chest physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e211S–50.
- Wahidi MM, Herth F, Yasufuku K, *et al.* Technical aspects of endobronchial ultrasound-guided transbronchial needle aspiration: chest guideline and expert panel report. *Chest* 2016;149:816–35.
- Nardecchia E, Cattoni M, Dominioni L. Endobronchial ultrasound-transbronchial needle aspiration for mediastinal staging of non-small cell lung cancer: variability of results and perspectives. *J Thorac Dis* 2017;9:S418–24.
- Kennedy MP, Jimenez CA, Morice RC, *et al.* Factors influencing the diagnostic yield of endobronchial ultrasound-guided transbronchial needle aspiration. *J Bronchology Interv Pulmonol* 2010;17:202–8.
- Phua GC, Rhee K-J, Koh M, *et al.* A strategy to improve the yield of transbronchial needle aspiration. *Surg Endosc* 2010;24:2105–9.
- DePew ZS, Edell ES, Midthun DE, *et al.* Endobronchial ultrasound-guided transbronchial needle aspiration: determinants of sampling adequacy. *J Bronchology Interv Pulmonol* 2012;19:271–6.
- Kemp SV, El Batrawy SH, Harrison RN, *et al.* Learning curves for endobronchial ultrasound using cusum analysis. *Thorax* 2010;65:534–8.
- Hu Y, Puri V, Crabtree TD, *et al.* Attaining proficiency with endobronchial ultrasound-guided transbronchial needle aspiration. *J Thorac Cardiovasc Surg* 2013;146:1387–92.
- Nguyen S, Ferland N, Beaudoin S, *et al.* Influence of trainee involvement on procedural characteristics for linear endobronchial ultrasound. *Thorac Cancer* 2017;8:517–22.
- Saji J, Kurimoto N, Morita K, *et al.* Comparison of 21-gauge and 22-gauge needles for endobronchial ultrasound-guided transbronchial needle aspiration of mediastinal and hilar lymph nodes. *J Bronchology Interv Pulmonol* 2011;18:239–46.
- Mohan A, Naik S, Pandey RM, *et al.* Diagnostic utility of endobronchial ultrasound guided transbronchial needle aspiration for mediastinal lesions: a prospective three year, single centre analysis. *Thorac Cancer* 2011;2:183–9.
- Gu P, Zhao Y-Z, Jiang L-Y, *et al.* Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. *Eur J Cancer* 2009;45:1389–96.
- Skov BG, Baandrup U, Jakobsen GK, *et al.* Cytopathologic diagnoses of fine-needle aspirations from endoscopic ultrasound of the mediastinum: reproducibility of the diagnoses and representativeness of aspirates from lymph nodes. *Cancer* 2007;111:234–41.
- Sehgal IS, Dhooria S, Aggarwal AN, *et al.* Training and proficiency in endobronchial ultrasound-guided transbronchial needle aspiration: a systematic review. *Respirology* 2017;22:1547–57.
- Beamis JF, Becker HD, Cavaliere S, *et al.* ERS/ATS statement on interventional pulmonology. *Eur Respir J* 2002;19:356–73.
- Fujiwara T, Yasufuku K, Nakajima T, *et al.* The utility of sonographic features during endobronchial ultrasound-guided transbronchial needle aspiration for lymph node staging in patients with lung cancer: a standard endobronchial ultrasound image classification system. *Chest* 2010;138:641–7.
- Herth FJF, Ernst A, Eberhardt R, *et al.* Endobronchial ultrasound-guided transbronchial needle aspiration of lymph nodes in the radiologically normal mediastinum. *Eur Respir J* 2006;28:910–4.
- Whiting P, Rutjes AWS, Reitsma JB, *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- Centre for Reviews Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York University of York; 2009. <https://www.york.ac.uk/crd/guidance/> [Accessed 7 May 2012].
- Bolsin S, Colson M. The use of the Cusum technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care* 2000;12:433–8.
- Eltoum IA, Chhieng DC, Jhala D, *et al.* Cumulative sum procedure in evaluation of EUS-guided FNA cytology: the learning curve and diagnostic performance beyond sensitivity and specificity. *Cytopathology* 2007;18:143–50.
- Hosmer DW, Lemeshow S, Sturdivant RX. Special Topics. In: Hosmer DW, Lemeshow S, Sturdivant RX, eds. *Applied logistic regression*. 3rd ed. New York: NY: John Wiley & Sons, Inc, 2013: 401–8.
- Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
- Lim TO, Soraya A, Ding LM, *et al.* Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 2002;14:251–8.
- Robinson IA, Blackham RB, Cozens NJ, *et al.* Good practice in head and neck fine needle aspiration cytology as assessed by CUSUM. *Cytopathology* 2002;13:335–42.
- Bonifazi M, Zuccatosta L, Trisolini R, *et al.* Transbronchial needle aspiration: a systematic review on predictors of a successful aspirate. *Respiration* 2013;86:123–34.
- 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.
- Ost DE, Ernst A, Lei X, *et al.* Diagnostic yield of endobronchial ultrasound-guided transbronchial needle aspiration: results of the AQUIRE bronchoscopy registry. *Chest* 2011;140:1557–66.
- Konge L, Clementsen PF, Annema JT. Diagnostic yield of EBUS-TBNA during the learning curve. *J Bronchology Interv Pulmonol* 2018;25:e26–7.
- Navani N, Nankivell M, Nadarajan P, *et al.* The learning curve for EBUS-TBNA. *Thorax* 2011;66:352–3.
- Wang HE, D'CRUZ C, Yam DC, *et al.* Learning curve for endobronchial ultrasound-guided transbronchial needle aspiration: experience of a community-based teaching hospital. *Mol Med Rep* 2014;10:2441–6.
- Lin C-K, Lai C-L, Chang L-Y, *et al.* Learning curve and advantages of endobronchial ultrasound-guided transbronchial needle aspiration as a first-line diagnostic and staging procedure. *Thorac Cancer* 2018;9:75–82.
- Stather DR, Chee A, MacEachern P, *et al.* Endobronchial ultrasound learning curve in interventional pulmonary fellows. *Respirology* 2015;20:333–9.
- Abu-Hijleh M, El-Sameed Y, Eldridge K, *et al.* Linear probe endobronchial ultrasound bronchoscopy with guided transbronchial needle aspiration (EBUS-TBNA) in the evaluation of mediastinal and hilar pathology: introducing the procedure to a teaching institution. *Lung* 2013;191:109–15.
- Seijo LM, Campo A, de Torres JP, *et al.* FDG uptake and the diagnostic yield of transbronchial needle aspiration. *J Bronchology Interv Pulmonol* 2011;18:7–14.



44 Lijmer JGet *al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.