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 - SUPPLEMENTARY APPENDIX

Multivariate logistic regression analysis:

A multivariate sequential 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% (our threshold for unacceptable failure rate). Potentially confounding variables were selected to be included in the first saturated model based on biological plausibility, directed acyclic graphs (DAGs) and having an association with the dependent and independent variables under study with a two-tailed $P < 0.2$. Variables not contributing to the multivariate model at a two-tailed $P < 0.05$ where removed from the model if their elimination did not alter the coefficient of other variables or the R^2 of the model in order to obtain the most parsimonious model. The association with diagnostic EBUS was measured using the odds ratio (OR) with its 95% confidence interval (95%CI) calculated using the bootstrapping method. All associations where considered statistically significant at a two-tailed $P < 0.05$. Goodness of fit was assesses using the Hosmer-Lemeshow test.

CUSUM Analysis to assess pathologist performance and learning curves:

We used binary CUSUM analysis to assess pathologist performance (including the existence of a learning curve), which requires setting an acceptable failure rate (level of

error if the procedure is carried out correctly, due to the inherent variability of the test) 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 accusing a pathologist of being incompetent, designated α) of 0.1 and a type II error (odds of falsely certifying someone as competent, designated β) of 0.1. [1-3]

We constructed binomial CUSUM charts using Microsoft Excel 2013 (Microsoft Corporation, Redmond, WA, USA). This method consists in the cumulative sum of failure minus success with each case. The CUSUM score was calculated using the equation:

$$S_n = \sum (X_i - s)$$

Where S_n = CUSUM, X_i = 1 for failure and 0 for success; s is a score calculated based on probability of acceptable and unacceptable failure rates. For each failure a score of (1 - s) was added and for each success a score of ($-s$) was added (in other words, s was subtracted). The score (s) was calculated using the equation:

$$s = \frac{\ln((1-p_0)/(1-p_1))}{\ln((1-p_0)/(1-p_1))+\ln(p_1/p_0)}$$

Where p_0 was the acceptable failure rate (10%) and p_1 was the unacceptable failure rate (20%). Therefore, s from the above equation equals 0.1452, which means that for each failure we added 0.8548 (1- 0.1452=0.8548) and for each success we subtracted 0.1452. 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. When the CUSUM curve crosses a decision limit (decision threshold) from above it is inferred that the failure rates were within the predetermined acceptable rate of 10% (absence of a statistically significant difference with the predetermined acceptable failure rate of 10%, i.e. 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% (absence of a statistically significant difference with the predetermined unacceptable failure rate of 20%, i.e. bad performance); if the CUSUM curve remained between two decision limits continued observation is indicated (stable performance within acceptable levels). Therefore, competence is assumed when the CUSUM curve is sloping downward or remains stable, but when the curve is sloping upward it indicates a below than acceptable success rate.

The decision limits (h_1 y h_0) are calculated based on the risk of type I (α) and II (β) errors using the following equations:

$$h_1 = \frac{\ln\left((1-\beta)/\alpha\right)}{\ln((1-p_0)/(1-p_1)) + \ln\left(p_1/p_0\right)}$$

$$h_0 = \frac{-\ln\left((1-\alpha)/\beta\right)}{\ln((1-p_0)/(1-p_1)) + \ln\left(p_1/p_0\right)}$$

When $\alpha = \beta$, then $h_0 = h_1$ and the decision limits are multiple of h_0. In our case, as $\alpha = \beta = 0.1$; $p_0 = 10\%$ and $p_1 = 20\%$; then $h_0 = h_1 = 2.71$. Therefore, we marked the decision limits in our CUSUM graphs as horizontal lines from the *y* axis at intervals of 2.71.

Supplementary Appendix Table 1. General characteristics of the cohort by Pathologist Group (reference group versus low performance group)

		Pathologists having diagnostic accuracy > 80%	Pathologists with statistically significant lower performance (6, 7, 9, 15)
N (nodes/patients)		367/201	287/163
Sex			
Male, n (%)		264 (71.9%)	204 (71.1%)
Age, mean (SD)		63.5 (12.6)	64.1 (13.4)
Indications of EBUS-TBNA	Adenopathies (intra-thoracic cancer), n (%)	248 (67.6%)	199 (69.8%)
	Adenopathies (extra-thoracic cancer), n (%)	41 (11.2%)	23 (8.1%)
	Inflammatory adenopathies, n (%)	21 (5.7%)	30 (10.5%)
	Cancer staging, n (%)	28 (7.6%)	17 (6.0%)
	Infectious adenopathies, n (%)	6 (1.6%)	0 (0.0%)
	Histological re-evaluation of cancer, n (%)	2 (0.5%)	0 (0.0%)
	Other, n (%)	21 (5.7%)	16 (5.6%)
Adenopathy size, median (IQR)		13.0 (10.0-20.0)	14.0 (10.0-20.0)
PET-CT	Negative, n (%)	35 (9.5%)	26 (9.1%)
	Positive (SUV max. > 2.5), n (%)	139 (37.9%)	102 (35.5%)
	Not performed, n (%)	193 (52.6%)	159 (55.4%)
Nodal station	7, n (%)	108 (29.4%)	86 (30.0%)
	4R, n (%)	96 (26.2%)	66 (23.0%)
	10R, n (%)	61 (16.6%)	40 (13.9%)
	4L, n (%)	28 (7.6%)	33 (11.5%)
	10L, n (%)	27 (7.4%)	20 (7.0%)
	2R, n (%)	12 (3.3%)	12 (4.2%)
	11R, n (%)	10 (2.7%)	8 (2.8%)
	11L, n (%)	9 (2.5%)	7 (2.4%)
	12R, n (%)	5 (1.4%)	2 (0.7%)
	2L, n (%)	3 (0.8%)	2 (0.7%)
	8, n (%)	2 (0.5%)	2 (0.7%)
	12L, n (%)	0 (0%)	1 (0.3%)
	5, n (%)	1 (0.3%)	1 (0.3%)
	3, n (%)	0 (0%)	1 (0.3%)
	Mass, n (%)	5 (1.4%)	6 (2.1%)
Final diagnosis	Non-small cell lung cancer	125 (34%)	89 (31%)
	Small cell lung cancer	20 (5%)	7 (2%)
	Lymphoma	6 (2%)	4 (1%)
	Other cancer	44 (12%)	35 (12%)
	Normal	102 (28%)	51 (18%)

	Sarcoidosis	6 (2%)	13 (5%)
	Tuberculosis	4 (1%)	0 (0%)
	Granulomas non sarcoidosis non tuberculosis	8 (2%)	6 (2%)
	Other benign conditions	13 (4%)	7 (2%)
Anaesthetic modality	Conscious sedation, n (%)	364 (99,2%)	287 (100%)
	General anaesthesia, n (%)	3 (0,8%)	0 (0,0%)
	Local anaesthesia, n (%)	0 (0,0%)	0 (0,0%)

Notes: PET-CT: positron emission tomography - computed tomography scan; EBUS-TBNA: endobronchial ultrasonography - transbronchial needle aspiration; SUV max.: maximum standardized uptake value; IQR: interquartile range (percentile 25 to percentile 75); CT: computed tomography

Supplementary Appendix table 2. Pathologists' experience and accuracy

Pathologist	Pathologist's Experience in years			Pathologist's Experience in number of EBUS			Diagnostic accuracy (EBUS was diagnostic)		
	Median	P25	P75	Median	P25	P75	n	%	N total of EBUS
1.00	4.1	3.1	4.9	60	30	90	99	81.8%	121
2.00	2.9	2.9	2.9	3	1	4	6	100.0%	6
3.00	2.9	1.8	3.7	64	32	96	111	86.7%	128
4.00	8.3	8.2	8.3	2	1	3	5	100.0%	5
5.00	0.2	0.2	0.3	5	2	7	10	100.0%	10
6.00	2.3	0.7	2.8	26	13	39	38	73.1%	52
7.00	4.0	3.2	5.0	47	23	70	73	77.7%	94
8.00	0.3	0.3	0.7	6	3	9	12	100.0%	12
9.00	1.9	1.2	2.4	37	18	55	55	74.3%	74
10.00	0.6	0.5	0.6	3	1	4	5	83.3%	6
11.00	1.3	1.1	1.7	11	5	16	22	100.0%	22
12.00	2.0	1.5	2.0	6	3	9	10	83.3%	12
13.00	5.4	4.8	5.7	18	9	27	31	86.1%	36
14.00	2.8	2.8	2.9	4	2	6	9	100.0%	9
15.00	1.5	0.8	1.8	33	16	50	43	64.2%	67
16.00	3.4	3.3	3.5	4	2	5	4	66.7%	6
Notes: pathologists' experience was measured for each EBUS; therefore, in this table we report the median and interquartile range (percentile 25 –P25– to percentile 75 –P75–) of pathologists' experience.									

REFERENCES

1. Bolsin S, Colson M. The use of the Cusum technique in the assessment of trainee competence in new procedures. *International journal for quality in health care : journal of the International Society for Quality in Health Care.* 2000;12(5):433-438.
2. Kemp SV, El Batrawy SH, Harrison RN, et al. Learning curves for endobronchial ultrasound using cusum analysis. *Thorax.* 2010;65(6):534-538.
3. 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 : official journal of the British Society for Clinical Cytology.* 2007;18(3):143-150.