Diagnostic Yield of Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration

Results of the AQuIRE Bronchoscopy Registry

David E. Ost, MD, MPH; Armin Ernst, MD, FCCP; Xiudong Lei, PhD;
David Feller-Kopman, MD, FCCP; George A. Eapen, MD, FCCP;
Kevin L. Kovitz MD, FCCP; Felix J. F. Herth, MD, FCCP;
and Michael Simoff, MD, FCCP; on behalf of the AQuIRE Bronchoscopy Registry

Background: New transbronchial needle aspiration (TBNA) technologies have been developed, but their clinical effectiveness and determinants of diagnostic yield have not been quantified. Prospective data are needed to determine risk-adjusted diagnostic yield.

Methods: We prospectively enrolled patients undergoing TBNA of mediastinal lymph nodes in the American College of Chest Physicians Quality Improvement Registry, Evaluation, and Education (AQuIRE) multicenter database and recorded clinical, procedural, and provider information. All clinical decisions, including type of TBNA used (conventional vs endobronchial ultrasound-guided), were made by the attending bronchoscopist. The primary outcome was obtaining a specific diagnosis.

Results: We enrolled 891 patients at six hospitals. Most procedures (95%) were performed with ultrasound guidance. A specific diagnosis was made in 447 cases. Unadjusted diagnostic yields were 37% to 54% for different hospitals, with significant between-hospital heterogeneity ($P = .0001$). Diagnostic yield was associated with annual hospital TBNA volume (OR, 1.003; 95% CI, 1.000-1.006; $P = .037$), smoking (OR, 1.55; 95% CI, 1.02-2.34; $P = .042$), biopsy of more than two sites (OR, 0.57; 95% CI, 0.38-0.85; $P = .015$), lymph node size (reference > 1-2 cm, ≤ 1 cm: OR, 0.51; 95% CI, 0.34-0.77; $P = .003$; > 2-3 cm: OR, 2.49; 95% CI, 1.61-3.85; $P < .001$; and > 3 cm: OR, 3.61; 95% CI, 2.17-6.00; $P < .001$), and positive PET scan (OR, 3.12; 95% CI, 1.39-7.01; $P = .018$). Biopsy was performed on more and smaller nodes at high-volume hospitals ($P < .0001$).

Conclusions: To our knowledge, this is the first bronchoscopy study of risk-adjusted diagnostic yields on a hospital-level basis. High-volume hospitals were associated with high diagnostic yields. This study also demonstrates the value of procedural registries as a quality improvement tool. A larger number and variety of participating hospitals is needed to verify these results and to further investigate other determinants of diagnostic yield.   *CHEST 2011; 140(6):1557–1566*

Abbreviations: ACCP = American College of Chest Physicians; AQuIRE = American College of Chest Physicians Quality Improvement Registry, Evaluation, and Education; EBUS-TBNA = endobronchial ultrasound-guided transbronchial needle aspiration; LN = lymph node; QI = quality improvement; TBNA = transbronchial needle aspiration

Assessing mediastinal lymph node (LN) involvement is critical to staging and treating non-small cell lung cancer.[1-3] CT and PET scans are useful for staging the mediastinum but are limited by false-positive and false-negative results. Therefore, in patients with mediastinal LN enlargement and no evidence of distant metastasis, tissue confirmation of nodal involvement is recommended.[2,4,5] Transbronchial needle

aspiration (TBNA) biopsy has been used to obtain tissue for mediastinal staging for decades, although it has been underused because of variations in training, operator skill, perceived risk, and diagnostic yield.[6] However, its use is increasing because of recent advances in bronchoscopic technology. In particular, endobronchial ultrasound-guided TBNA (EBUS-TBNA) biopsy has been found in randomized trials to have

excellent sensitivity and specificity for diagnosing mediastinal LN involvement compared with conventional TBNA biopsy and even mediastinoscopy.[7-9]

The importance and effectiveness of EBUS-TBNA biopsy in clinical practice is still being evaluated. Increasing medical costs have led to a greater need to rigorously evaluate new technologies to justify their expense.[10-13] Recently, bronchoscopic registries have been developed to collect data on both interventional and diagnostic bronchoscopy.[14,15] As a result, the American College of Chest Physicians (ACCP) has implemented a Web-based pilot project for multiinstitutional databases: the ACCP Quality Improvement Registry, Evaluation, and Education (AQuIRE) program. The first AQuIRE module is a bronchoscopy registry.

Because bronchoscopy is pivotal for staging lung cancer and diagnosing many other diseases involving the mediastinum, we decided to use the AQuIRE bronchoscopy registry to evaluate TBNA biopsy. Our primary objective was to identify factors that affect diagnostic yield. Our secondary objectives were to identify factors that affect the adequacy of tissue sampling and to demonstrate the feasibility of using registry data to generate risk-adjusted, hospital-level diagnostic yields as a possible quality improvement tool.

Materials and Methods

All patients undergoing bronchoscopy with LN sampling by TBNA biopsy were entered consecutively into the registry from February 13, 2009, to February 28, 2010. Institutional review board approval was obtained at each participating site, and informed consent or a waiver of consent was obtained according to institutional guidelines. All data were entered via a Web-based interface.[16]

Since this is a registry, all clinical decisions, including type of TBNA used (conventional vs EBUS), were left to the discretion of the attending bronchoscopist. The principal investigator for each site was primarily responsible for data quality for that site. Addiional quality control checks included standardized definitions with hyperlinks to the definitions from the question on the Web page, drop-down menus and buttons with minimal free text response, ACCP staff available for prompt dilemma resolution for data entry, use of data entry controls in the interface to limit entries to clinically plausible ranges, automated checks and reminders for data completion, automated data validity checks to ensure that clinically plausible data were entered (example: if general anesthesia was used, then an airway must have been used), and secondary data validity analysis.

Information included the following: (1) patient demographics; (2) clinical characteristics; (3) provider information, including physician, hospital, and annual hospital TBNA biopsy volume; (4) procedure indication; (5) sedation information; (6) procedural information; (7) biopsy results; and (8) complications. Clinical characteristics, such as PET scan results, were based on the radiology attending physician's reading using the standard of practice for that institution. The use of standard uptake values as compared with qualitative reading of PET scans was not mandated.

The primary outcome was diagnostic yield, which was defined as whether a specific diagnosis was made by TBNA biopsy. If no specific diagnosis was made by TBNA biopsy, but it was made by another method (ie, transbronchial biopsy), the TBNA biopsy results were considered nondiagnostic. A specific TBNA biopsy diagnosis was defined as a recognized disease (eg, non-small cell lung cancer or sarcoidosis) on the basis of histologic, cytologic, or microbiologic findings, as determined from TBNA biopsy specimens, but did not include conditions such as nonspecific inflammation. For the primary outcome, if adequate lymphocytes were obtained but no diagnosis was made then this was considered nondiagnostic. Outcomes were analyzed on a per-patient basis and a secondary analysis was performed on a per-LN basis.

Our secondary outcome was adequacy of sampling. Cytologic or histologic specimens that demonstrated specific diagnoses or adequate lymphocytes were considered adequate. If no diagnoses were made and insufficient lymphocytes were available to verify adequate LN sampling, the samples were considered inadequate. Adequacy was defined on a per-LN basis.

Groups were compared using the χ^2 test or Fisher exact test, as appropriate. We used the Cochran-Armitage test for trend, with one-sided P values. We compared hierarchical models in which hospitals were treated as higher levels with ordinary logistic regression models of independent data by removing the hospital variance component from the models. The hierarchical model for diagnostic yield on a per-patient basis was specified as patients nested within hospitals. Hospital-level variables included annual TBNA biopsy volume. Patient-level variables included sex, smoking status, number of LNs undergoing biopsy, largest LN, anesthesia use, and linear EBUS use. In the multivariable hierarchical logistic regression per-LN models, we included a third level such that LNs from the same patient were nested within that patient. LN-level variables included size, location, on-site cytologic evaluation use, and linear EBUS use. Patient-level variables included sex, smoking status, anesthesia use, and number of LNs undergoing biopsy. We decided a priori to include all variables with P values < .2 on univariate analysis. We also used indirect standardization to compare hospitals in terms of diagnostic yield after adjusting for LN size. In addition, linear regression models were used to test the association between mean number and size of LNs compared with hospital volume. P values < .05 were considered significant and all tests were two-sided except the Cochrane-Armitage test; no adjustments for multiple comparisons were made. All statistical analyses were performed using

SAS software, version 9.1 (SAS Institute; Cary, North Carolina) and Intercooled Stata 9.2 (Statacorp LP; College Station, Texas).

Results

Six hospitals with 11 physicians enrolled 891 patients. Baseline demographics, clinical characteristics, and procedural details are shown in Table 1, and diagnoses are shown in Table 2. EBUS-TBNA was used in 95% of the cases, and conventional TBNA was used in the other 5%. The most common diagnosis was malignancy, followed by sarcoidosis. The percentage of cases with a malignant diagnosis did not significantly differ among hospitals ($P = .16$). A specific diagnosis was made by TBNA biopsy in 447 (50%) patients. Diagnostic yield varied significantly among hospitals, with crude (unadjusted) diagnostic yields of 37% to 54%.

Univariate associations of clinical characteristics with diagnostic yield on per-patient and per-LN bases are shown in Tables 3 and 4. Homogeneity tests demonstrated significant between-hospital variability ($P = .0001$); thus, a multivariate hierarchical logistic regression analysis was used to account for heterogeneity. In terms of the magnitude of effect of between-hospital variations, OR estimates for some variables changed significantly in hierarchical models compared with ordinary logistic regression models. For example, general anesthesia had an OR of 1.77 ($P = .002$) on ordinary multivariate logistic regression analysis compared with 1.11 ($P = .66$) on hierarchical logistic analysis to account for between-hospital variability.

The results of multivariate hierarchical logistic regression analysis of diagnostic yield on a per-patient and per-LN bases are shown in Tables 5 and 6. On a per-patient basis, annual hospital TBNA biopsy volume ($P = .037$), smoking ($P = .042$), biopsy of two or fewer LN sites ($P = .015$), and large LN size (overall $P < .001$) were associated with obtaining a specific diagnosis. On a per-LN basis, hospital volume ($P = .013$), female sex ($P = .045$), smoking ($P = .037$), and large LN size (overall $P < .001$) were associated with obtaining a specific diagnosis. In the subset of patients who underwent PET scanning, positive PET scans were associated with specific diagnoses on per-patient ($P = .018$) and per-LN bases ($P < .001$).

Sampling adequacy on a per-LN basis was 90%, with little between-hospital variation. A univariate analysis demonstrated that only LN size ($P = .01$), LN location ($P < .001$), and EBUS use ($P < .001$) were associated with adequacy (e-Table 1). A multivariate hierarchical logistic regression analysis (e-Table 2) demonstrated that only LN location other than 4L ($P < .001$) and linear EBUS use ($P = .006$) were associated with adequacy.

Table 1—Patient Demographics, Clinical Characteristics, and Procedural Details

Variable	Frequency (N = 891)
Age, y, mean ± SD	62.6 ± 13.5
Male:female sex	456 (51):435 (49)
Ethnicity	
American Indian or Alaskan Native	2 (0.2)
Asian	29 (3.2)
African American	91 (10.2)
Hispanic or Latino	32 (3.6)
Non-Hispanic white	737 (82.7)
Comorbidity	
Hypertension	440 (49.4)
COPD	204 (22.9)
Coronary artery disease	135 (15.2)
Diabetes	131 (14.7)
Congestive heart failure	32 (3.6)
Asthma	45 (5)
Chronic renal failure, no dialysis	11 (1.2)
Chronic renal failure, dialysis	8 (0.9)
Cerebrovascular accident	23 (2.6)
Hematologic malignancy	23 (2.6)
None of the above	280 (31.4)
Service site	
Inpatient hospital	77 (8.6)
Outpatient hospital	813 (91.2)
Physician office	1 (0.1)
Procedure urgency	
Elective	878 (98.5)
Urgent	13 (1.5)
Bronchoscopy type	
Flexible	876 (98.3)
Rigid and flexible	15 (1.7)
Bronchoscopy route	
Nasal	109 (12.2)
Oral	295 (33.1)
Endotracheal tube	59 (6.6)
Laryngeal mask airway	433 (48.6)
Tracheotomy tube	4 (0.4)
Procedure performed	
EBUS-TBNA biopsy	853 (95.7)
Conventional TBNA biopsy	38 (4.3)
Transbronchial biopsy	130 (14.6)
Endobronchial biopsy	118 (13.2)
Brush	138 (15.5)
BAL	194 (21.8)
Lymph node stations undergoing biopsy by TBNA	
1	12 (1.3)
2R	42 (4.7)
2L	4 (0.4)
3	7 (0.8)
4R	415 (46.6)
4L	240 (26.9)
5	7 (0.8)
7	475 (53.3)
10R	94 (10.4)
10L	31 (3.5)
11R	265 (29.7)
11L	203 (22.8)
12R	22 (2.5)
12L	5 (0.6)

Data presented as No. (%) unless otherwise noted. EBUS-TBNA = endobronchial ultrasound-guided transbronchial needle aspiration; TBNA = transbronchial needle aspiration.

Table 2—*TBNA Diagnoses*

Diagnosis	Cases, No. (%)
Primary lung neoplasm	
Small cell lung cancer	46 (10.3)
Non-small cell lung adenocarcinoma	136 (30.4)
Non-small cell lung squamous cell carcinoma	65 (14.5)
Non-small cell lung cancer, unspecified (undifferentiated)	55 (12.3)
Carcinoid lung tumor	12 (2.7)
Other	7 (1.6)
Metastatic lung neoplasm	
Solid tumor with lung metastasis	44 (9.8)
Hematologic with lung metastasis	12 (2.7)
Infection	
Bacterial	5 (1.1)
Actinomyces	1 (0.2)
Aspergillus	1 (0.2)
Histoplasmosis	1 (0.2)
Mycobacterium avium-intracellulare	1 (0.2)
Viral	1 (0.2)
Other	
Sarcoidosis[a]	55 (12.3)
Bronchogenic cyst	3 (0.7)
Other	2 (0.4)

Data presented as No. (%) unless otherwise noted. See Table 1 legend for expansion of abbreviation.

[a]The diagnosis of sarcoidosis required the finding of granulomatous inflammation in the appropriate clinical context. The final diagnosis was based on the decision of the attending pulmonologist.

To further determine how anesthesia use and hospital volume affected practice patterns, we analyzed the association between anesthesia level and LN number and size (Table 7). Deep or general anesthesia was associated with more LNs sampled per patient ($P < .0001$), smaller LNs ($P < .0001$), and more frequent on-site cytologic evaluation use ($P < .001$). This also held true for the subset of patients with lung cancer (Table 8), including LNs that were ≤ 1 cm ($P = .004$). High hospital volume, defined as being above the median annual TBNA biopsy volume per year, was also associated with sampling more LNs per patient ($P < .0001$) and sampling smaller LNs ($P < .0001$) (Table 9).

To estimate risk-adjusted diagnostic yields among hospitals, we performed indirect standardization based on lymph node size (Table 10). Hospitals that used deep or general anesthesia tended to have higher TBNA volumes. Risk-adjusted diagnostic yields varied from 0.35 to 0.58.

Discussion

To our knowledge, this is the first prospective study of risk-adjusted diagnostic yield for EBUS-TBNA biopsy using multicenter registry data. Our findings suggest that diagnostic yield varies significantly among hospitals and is driven by patient- and hospital-level

variables. Patient-level variables such as LN size have been identified previously, but this is the first prospective study to demonstrate a relationship between hospital TBNA biopsy volume and diagnostic yield.

The effect of hospital volume on diagnostic yield is important to quantify. An OR of 1.003 may seem small, but it is important to interpret the coefficients

Table 3—*Patient and Clinical Characteristics by Diagnostic Yield Per Patient*

Characteristic	No Specific Diagnosis Made (n = 444)	Specific Diagnosis Made (n = 447)	P Value
Median age, y	64.5	64	…
Age, y			.81
< 65	222 (49.4)	227 (50.6)	
≥ 65	222 (50.2)	220 (49.8)	
Sex			.048
Female	202 (46.4)	233 (53.6)	
Male	242 (53.1)	214 (46.9)	
Ethnicity			.68
Non-white	79 (51.3)	75 (48.7)	
White	365 (49.5)	372 (50.5)	
ASA score			.57
1 or 2	193 (50.9)	186 (49.1)	
3 or 4	251 (49.0)	261 (51.0)	
Zubrod score			.59
0	54 (49.1)	56 (50.9)	
1	285 (51.1)	273 (48.9)	
2-4	105 (47.1)	118 (52.9)	
Smoking status			.049
Never	140 (56.5)	108 (43.5)	
Current	81 (46.8)	92 (53.2)	
Previous	223 (47.4)	247 (52.6)	
No. of LN sites undergoing biopsy			.019
1 or 2	277 (47.0)	312 (53.0)	
2-6	167 (55.3)	135 (44.7)	
Largest LN, cm			< .001
≤ 1	125 (65.4)	66 (34.6)	
> 1-2	225 (53.4)	196 (46.6)	
> 2-3	57 (36.1)	101 (63.9)	
> 3	37 (30.6)	84 (69.4)	
Anesthesia			.08
Moderate or local	167 (53.9)	143 (46.1)	
Deep or general	277 (47.7)	304 (52.3)	
On-site cytologic evaluation			.16
No	114 (54.0)	97 (46.0)	
Yes	330 (48.5)	350 (51.5)	
Linear EBUS-TBNA biopsy			.044
No	25 (65.8)	13 (34.2)	
Yes	419 (49.1)	434 (50.9)	
Fellow participation			.81
No	63 (50.8)	61 (49.2)	
Yes	381 (49.7)	386 (50.3)	
PET scan on LN			< .001
Negative	62 (72.1)	24 (27.9)	
Positive	125 (42.7)	168 (57.3)	

Data presented as No. (%) unless otherwise noted. ASA = American Society of Anesthesiologists; LN = lymph node. See Table 1 legend for expansion of other abbreviations.

Table 4—Clinical Characteristics by Diagnostic Yield Per LN

Characteristic	No Specific Diagnosis Made (n = 1,250)	Specific Diagnosis Made (n = 569)	P Value
LN location			.008
Not 4L	1,068 (67.6)	512 (32.4)	
4L	182 (76.2)	57 (23.8)	
LN size, cm			< .001
≤ 1	685 (83.4)	136 (16.6)	
> 1-2	451 (61.1)	287 (38.9)	
> 2-3	80 (44.2)	101 (55.8)	
> 3	34 (43.0)	45 (57.0)	
On-site cytologic evaluation			< .001
No	249 (60.4)	163 (39.6)	
Yes	1,001 (71.1)	406 (28.9)	
Linear EBUS-TBNA biopsy			.41
No	32 (74.4)	11 (25.6)	
Yes	1,218 (68.6)	558 (31.4)	
PET scan on LN			< .001
Negative	336 (89.4)	40 (10.6)	
Positive	274 (59.2)	189 (40.8)	

Data presented as No. (%) unless otherwise noted. See Table 1 and 3 legends for expansion of abbreviations.

correctly. The OR expresses the effect of a one-unit change in annual hospital volume, whereas the ORs for other variables in the model are categorical (ie, male vs female). A hospital that performs 100 more EBUS-TBNA biopsy procedures per year than another hospital would be associated with an OR of $1.003^{100} = 1.35$. Thus, each 100-unit increase in hospital volume increases the odds of a diagnosis by another 35%.

Hospital volume was strongly associated with the probability of obtaining a specific diagnosis, but we found no association between hospital volume and LN sampling adequacy. One possible explanation for this finding is that although LN sampling adequacy does not vary on a per-node basis between hospitals, the number of LNs sampled per patient and willingness to sample smaller LNs does vary and affects outcome. Our findings are consistent with this hypothesis because high-volume hospitals and hospitals that use general anesthesia sampled more LNs and were more likely to sample small LNs than were lower-volume hospitals.

Clinically, significant differences exist between procedures that require deep or general anesthesia and those performed under moderate sedation. Moderate sedation procedures are typically time-driven, in that physicians have a finite amount of time before sedation, and patient comfort issues compel them to terminate the procedure. Thus, moderate sedation procedures typically focus on large, high-yield LNs first. In contrast, more systematic and complete sampling can be performed under deep or general anesthesia. In our study, the hospitals that used anesthesia most frequently were also the highest-volume hospitals and had the highest diagnostic yields (Table 10).

The number of hospitals studied and the absence of routine use of deep or general anesthesia at low-volume hospitals limited our power to demonstrate the effect of anesthesia on diagnostic yield. The use of deep or general anesthesia at low-volume hospitals may result in more LN sampling and higher diagnostic yields, but this would require additional study. High-volume hospitals that used moderate sedation and sampled more LNs (eg, hospital E) had higher diagnostic yields than did hospitals that sampled fewer nodal stations, but the yields were still lower than those in high-volume hospitals that used anesthesia. It is important to note that all the hospitals in this study probably have higher EBUS-TBNA volumes than the national average. As with all forms of regression analysis, caution should be exercised when

Table 5—Multivariate Hierarchical Logistic Regression Model for Diagnostic Yield Per Patient

Variable	All Patients (n = 891)			Patients With PET Scans (n = 379)		
	OR	95% CI	P Value	OR	95% CI	P Value
Hospital volume	1.003	1.000-1.006	.037	1.003	0.999-1.007	.14
Sex: male vs female	0.72	0.5-1.05	.08	0.75	0.42-1.35	.26
Smoking status: smoker vs never smoker	1.55	1.02-2.34	.042	1.59	0.79-3.18	.15
Number of LN sites: > 2 vs ≤ 1	0.57	0.38-0.85	.015	0.55	0.3-1.04	.06
On-site cytologic evaluation: yes vs no	1.15	0.62-2.12	.56	1.39	0.26-7.36	.48
Anesthesia: deep/general vs moderate	1.11	0.62-2.01	.66	1.06	0.19-5.95	.90
Largest LN, cm						
≤ 1 vs < 1-2 (reference)	0.51	0.34-0.77	.003	0.41	0.23-0.75	.006
< 2-3 vs < 1-2	2.49	1.61-3.85	< .001	2.27	1.07-4.78	.034
> 3 vs < 1-2	3.61	2.17-6.00	< .001	3.37	1.26-8.98	.019
Linear EBUS-TBNA biopsy: yes vs no	2.42	0.91-6.44	.07	0.67	0.04-10.22	.67
PET scan: positive vs negative				3.12	1.39-7.01	.018

See Table 1 and 3 legends for expansion of abbreviations.

Table 6—*Multivariate Hierarchical Logistic Regression Model for Diagnostic Yield Per LN*

Variable	All LNs (N = 1,819)			LNs From Patients With PET Scans (n = 839)		
	OR	95% CI	*P* Value	OR	95% CI	*P* Value
Hospital volume	1.002	1.000-1.004	.013	1.000	0.997-1.003	.90
Sex: male vs female	0.73	0.53-0.99	.045	0.84	0.53-1.32	.45
Smoking status: smoker vs never	1.45	1.02-2.06	.037	1.30	0.76-2.22	.34
Number of LN sites: >2 vs 1 or 2	0.38	0.05-3.08	.11	0.40	0.25-0.65	<.001
Deep/general vs moderate anesthesia	0.91	0.57-1.46	.70	1.23	0.52-2.9	.63
On-site cytologic evaluation: yes vs no	0.90	0.56-1.44	.63	2.55	0.68-9.55	.11
Size of LN, cm						
≤1 vs <1-2 (reference)	0.29	0.21-0.39	<.001	0.33	0.21-0.53	<.001
<2-3 vs <1-2	2.10	1.35-3.28	.001	1.52	0.7-3.31	.29
>3 vs <1-2	2.05	1.08-3.91	.029	2.49	0.78-8.00	.12
Location of LN: 4L vs other	0.98	0.65-1.47	.91	1.17	0.65-2.11	.59
Linear EBUS-TBNA biopsy: yes vs no	2.05	0.80-5.27	.13	4.19	0.39-44.42	.19
PET scan: positive vs negative				4.30	2.57-7.18	<.001

On-site cytologic evaluation was treated as an LN-level variable. See Table 1 and 3 legends for expansion of abbreviations.

applying the results, since the relationship between hospital volume and diagnostic yield is valid only over the range of observed values. The effect of hospital volume on diagnostic yield may be different at either lower or higher volumes than those in this study. One might hypothesize that once a certain upper volume limit is reached there may be a plateau in terms of diagnostic yield. Similarly, there may be a lower volume limit beyond which diagnostic yield plateaus and becomes relatively uniform (albeit low).

The relationship between hospital volume and diagnostic yield could also be explained by other hospital-level variables; the quality of cytopathologic processing and interpretation and cytopathologists' willingness to render specific diagnoses may be associated with the volume of cases seen. The amount of intraobserver and interobserver variability among cytopathologists

has been evaluated in other areas of diagnosis, such as of cervical smears,[17] but the evidence on TBNA biopsy evaluation is limited. Interobserver agreement for endoscopic ultrasonography and EBUS-TBNA biopsy, as measured by κ values, has been reported to range from 0.52 to 0.89 in other studies.[18,19] EBUS-TBNA biopsy studies in patients with suspected sarcoidosis[20] demonstrated a >20% increase in diagnostic yield when an experienced lung pathologist evaluated the slides. Therefore, much of the variation in bronchoscopy diagnosis may be explained by factors outside of the bronchoscopy laboratory.

The sources of between-hospital variations are important from a policy standpoint because of the increasing emphasis on pay for performance and quality improvement. Centralization of services may be an option if variations in yield are driven by hospital

Table 7—*Trend Analysis by Anesthesia Among All Patients*

LN Characteristic	Moderate Anesthesia (n = 310)	Deep or General Anesthesia (n = 581)	χ² *P* Value	Linear Trend *P* Value (One-Sided)
No. of LN sites undergoing biopsy			.0003	<.0001
1	131 (42.4)	178 (57.6)		
2	102 (36.4)	178 (63.6)		
3	54 (28.3)	137 (71.7)		
4	15 (23.1)	50 (76.9)		
5	8 (21.6)	29 (78.4)		
6	0 (0)	9 (100)		
Largest LN size, cm			<.0001	<.0001
≤1	35 (18.3)	156 (81.7)		
>1-2	137 (32.5)	284 (67.5)		
>2-3	73 (46.2)	85 (53.8)		
>3	65 (53.7)	56 (46.3)		
On-site cytologic evaluation			<.0001	<.0001
No	167 (79.1)	44 (20.9)		
Yes	143 (21.0)	537 (79.0)		

Data presented as No. (%) unless otherwise noted. See Table 3 for expansion of abbreviation.

Table 8—*Trend Analysis by Anesthesia Among Patients With Lung Cancer*

LN Finding	Moderate Anesthesia (n = 105)	Deep or General Anesthesia (n = 200)	χ^2 P Value	Linear Trend P Value (One-Sided)
LN sites, No.			.28[a]	.024
1	48 (37.5)	80 (62.5)		
2	35 (40.2)	52 (59.8)		
3	16 (24.6)	49 (75.4)		
4	4 (28.6)	10 (71.4)		
5	2 (25.0)	6 (75.0)		
6	0 (0)	3 (100)		
Largest LN, cm			<.0001	<.0001
≤1	11 (19.0)	47 (81.0)		
>1-2	34 (28.3)	86 (71.7)		
>2-3	28 (43.1)	37 (56.9)		
>3	32 (51.6)	30 (48.4)		
Largest LN			.006	.004
≤1	11 (19.0)	47 (81.0)		
>1	94 (38.1)	153 (61.9)		

Data presented as No. (%) unless otherwise noted. See Table 3 for expansion of abbreviation.
[a]Fisher exact test.

volume. Changes in bronchoscopy practices may also affect outcome. However, if variation in cytology interpretation is the key driver, that lends itself to a different and more technical set of solutions. For example, one solution might involve education to decrease interobserver variation, which has been demonstrated to be effective in cervical cytologic evaluation.[18] Alternatively, standardized specimen processing coupled with remote electronic slide reading could be used to provide all hospitals with access to high-volume and high-quality cytologic evaluation.

This study adds to the existing body of medical literature on EBUS-TBNA biopsy by emphasizing the importance of hospital-level variables in addition to patient-level variables. On the basis of our findings, between-hospital variation may be as important or more important than many patient-level variables from a quality improvement standpoint. Importantly, many variables typically considered patient level are often affected by hospital-level decisions. For example, anesthesia availability may be considered a hospital-level decision because at many institutions it is not available for routine bronchoscopy. In our cohort, no low-volume hospitals used anesthesia, but greater anesthesia availability may alter practice patterns at low-volume hospitals, leading to increased LN sampling and improved diagnostic yields.

The finding of significant differences in diagnostic yield between hospitals and the importance of hospital-level variables should be considered when designing future studies of EBUS-TBNA. We recommend that future multicenter studies of EBUS-TBNA be assessed for between-hospital heterogeneity and, when appropriate, that hierarchical models be used for analysis. The use of ordinary regression in such instances may lead to significant errors if between-hospital variation is not accounted for (Simpson's paradox).

The diagnostic yields found in this study are similar to those in previously published studies.[20,21] It is important to recognize that diagnostic yield is not the same as sensitivity. Different authors have used the

Table 9—*Trend Analysis by Annual TBNA Volume Among All Patients*

Measure	Low Volume Hospitals			High Volume Hospitals			χ^2 P Value	Linear Trend P Value
	A (V = 97)	B (V = 166)	C (V = 169)	D (V = 276)	E (V = 325)	F (V = 435)		
No. of LN sites undergoing biopsy							<.0001	<.0001
1	30 (49.2)	36 (41.4)	48 (50.0)	57 (28.5)	25 (25.8)	113 (32.3)		
2	22 (36.1)	34 (39.1)	27 (28.1)	85 (42.5)	28 (28.9)	84 (24.0)		
3	9 (14.8)	16 (18.4)	17 (17.7)	45 (22.5)	25 (25.8)	79 (22.6)		
4	0 (0)	1 (1.1)	2 (2.1)	10 (5.0)	13 (13.4)	39 (11.1)		
5	0 (0)	0 (0)	2 (2.1)	3 (1.5)	6 (6.2)	26 (7.4)		
6	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	9 (2.6)		
Mean (SD)[a]	1.65 (0.73)	1.79 (0.78)	1.78 (0.95)	2.09 (0.92)	2.45 (1.19)	2.45 (1.37)	…	…
Largest LN, cm							<.0001	<.0001
≤1	8 (13.1)	11 (12.6)	11 (11.5)	50 (25.0)	7 (7.2)	104 (29.7)		
>1-2	20 (32.8)	34 (39.1)	44 (45.8)	90 (45.0)	49 (50.5)	184 (52.6)		
>2-3	17 (27.9)	24 (27.6)	15 (15.6)	36 (18.0)	30 (30.9)	36 (10.3)		
>3	16 (26.2)	18 (20.7)	26 (27.1)	24 (12.0)	11 (11.3)	26 (7.4)		
Mean (SD)[a]	2.67 (1.01)	2.56 (0.96)	2.58 (1.01)	2.17 (0.94)	2.46 (0.79)	1.95 (0.83)	…	…

Data presented as No. (%) unless otherwise noted. Letters A-F correspond to participating hospitals. V = annual TBNA volume. See Table 1 and 3 legends for expansion of other abbreviations.
[a]Linear regression analysis of mean number and size of LN sites undergoing biopsy based on volume revealed that per one hospital volume increase, the mean number of LN sites undergoing biopsy increased 0.0027 (P = .003), and the largest size of LN undergoing biopsy decreased 0.002 cm (P = .018).

Table 10—Relationship Between Adjusted Diagnostic Yield, Hospital Volume, and Use of Deep or General Anesthesia

Hospital	Annual TBNA Volume, Cases/y	Percentage of TBNA Cases Using Deep or General Anesthesia	Percentage of TBNA Cases Using On-site Cytologic Evaluation	Risk-Adjusted Diagnostic Yield[a] (95% CI)
F	435	99	100	0.58 (0.50-0.67)
D	276	91	93	0.53 (0.43-0.64)
E	325	1	2	0.48 (0.36-0.64)
C	169	39	9	0.44 (0.33-0.59)
B	166	8	89	0.35 (0.24-0.49)
A	97	8	92	0.38 (0.25-0.56)

See Table 1 and 9 legends for expansion of abbreviations.
[a]Risk-adjusted diagnostic yield was based on indirect standardization using LN size.

term diagnostic yield to mean different things; some have used the term loosely and have included cases with specific diagnoses or adequate lymphocytes. We believe that our definition is more stringent and clinically relevant because physicians need specific diagnoses on which to base treatment decisions. In two systematic reviews of EBUS-TBNA biopsy in patients with suspected lung cancer (19 studies with a total of 1,955 patients),[21,22] diagnostic yield ranged from 8% to 93% by our definition (mean, 49% ± 29%).[7,21,23-37] Many of these studies were clinical trials designed to determine sensitivity or test new technologies and are thus not necessarily representative of routine practice. We reported consecutive cases, not all of whom had a high suspicion of cancer or mediastinal disease. As such, our data are more generalizable and serve to illustrate that registries can offer quality improvement, benchmarking, and clinical effectiveness insights that controlled clinical trials cannot. All approaches have strengths and weaknesses, so it is best to view registry data as complementing the knowledge gained from clinical trials.

These observations are useful, but it is important to recognize the limitations of the current study. The AQuIRE Registry is new and still growing. At the time of this report, it included relatively few hospitals and only 11 participating physicians. Therefore, we could not systematically evaluate the effect of between-physician variability on outcome. In addition, as noted here, no low-volume hospitals used anesthesia, which limited our ability to draw inferences about the effect of anesthesia on outcome. Few conventional TBNA biopsy procedures were performed; thus, these results apply predominantly to EBUS-TBNA. Case selection may also impact diagnostic yield. Some centers may be more aggressive in doing EBUS for CT scan- and PET scan-negative lymph nodes in patients with cancer.[29,35] Diagnostic yield would of course be expected to be lower in CT/PET scan-negative cases than in those with larger PET scan-positive nodes. We did control for this in the analysis. There were no differences between centers in terms of the overall proportion of cases result-

ing in a cancer diagnosis, so it is likely that EBUS is used to investigate a similar spectrum of problems across centers. However, as with any observational study, there is always the possibility that residual confounding exists. In addition, the registry does not incorporate cost data. As such, cost-effectiveness is beyond the scope of this study, although the data generated can inform cost-effectiveness analyses.

Finally, because this is a report of registry data rather than a research trial, we can only comment on diagnostic yield and not diagnostic sensitivity, because not all patients underwent surgery. The registry was designed to provide a quality improvement tool and to study clinical effectiveness; thus, this was a necessary trade-off because it would be impractical to obtain surgical biopsy specimens in everyday practice. Given these limitations, we believe that the ability to track risk-adjusted yields within institutions is a useful quality improvement tool.

In as much as quality improvement (QI) is a local process involving measurement, evaluation, and intervention in an iterative and ongoing fashion, there is no single intervention that can improve outcomes universally given the variety of practice settings. Indeed, all QI interventions have been shown to work in some situations, but no one intervention is demonstrably superior in every situation.[38] However, one critical element necessary for all QI programs is measurement of outcomes. The Donabedian[39] model of quality includes three dimensions: structure, process, and outcomes. Outcomes represent the combined effects of structure and process. Monitoring of outcomes is to monitor performance, which is conditional upon structure and process. Structure and process are easier to measure and are often used as surrogates. Unfortunately, a change in structure or process does not necessarily impact outcomes in the expected manner. It is continuous performance monitoring that keeps quality of care high.

As applied to EBUS-TBNA, the primary outcome of interest is diagnostic yield. We therefore recommend that clinicians monitor their own risk-adjusted outcomes and compare them to those of other

institutions. AQuIRE provides a convenient way to do this. Individual centers can then look critically at their own structures (equipment, space, services available, staff training, and so forth) and processes (patient selection, sedation methods, cytology quality, and so forth). Interventions to alter structures or processes can then be implemented that are feasible at the local level and their end results measured. This must be an ongoing and iterative process. Monitoring of outcomes is just the first step of QI, but it is an absolutely essential step.

In conclusion, this report from the AQuIRE registry represents the first multicenter prospective observational study to identify important patient-level and hospital-level determinants of EBUS-TBNA diagnostic yield. Patient-level variables that affected diagnostic yield included LN size, smoking status, and number of LN stations undergoing biopsy. A high hospital TBNA biopsy volume had a significant effect on diagnostic yield that likely reflects some combination of team experience, anesthesia use, and cytopathologic results interpretation. The AQuIRE registry generated risk-adjusted diagnostic yields at the hospital level that should be useful for quality improvement. Future studies should expand this registry to other hospitals of varying sizes and more diverse practice patterns so that these preliminary findings can be verified and to more accurately assess the effect of anesthesia and between-physician variation on diagnostic yield.

REFERENCES

1. Alberts WM. Updated ACCP Guidelines for the diagnosis and management of lung cancer: what are the important changes in recommendations? *Pol Arch Med Wewn*. 2008;118(1-2):4-5.
2. Detterbeck FC, Jantz MA, Wallace M, Vansteenkiste J, Silvestri GA; American College of Chest Physicians. Invasive mediastinal staging of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*. 2007;132(suppl 3):202S-220S.
3. Silvestri GA, Gould MK, Margolis ML, et al; American College of Chest Physicians. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest*. 2007;132(suppl 3):178S-201S.
4. The American Thoracic Society and The European Respiratory Society. Pretreatment evaluation of non-small-cell lung cancer. *Am J Respir Crit Care Med*. 1997;156(1):320-332.
5. De Leyn P, Lardinois D, Van Schil PE, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. *Eur J Cardiothorac Surg*. 2007;32(1):1-8.
6. Dasgupta A, Mehta AC. Transbronchial needle aspiration. An underused diagnostic technique. *Clin Chest Med*. 1999;20(1):39-51.
7. Ernst A, Anantham D, Eberhardt R, Krasnik M, Herth FJ. Diagnosis of mediastinal adenopathy-real-time endobronchial ultrasound guided needle aspiration versus mediastinoscopy. *J Thorac Oncol*. 2008;3(6):577-582.
8. Eberhardt R, Anantham D, Ernst A, Feller-Kopman D, Herth F. Multimodality bronchoscopic diagnosis of peripheral lung lesions: a randomized controlled trial. *Am J Respir Crit Care Med*. 2007;176(1):36-41.
9. Herth F, Becker HD, Ernst A. Conventional vs endobronchial ultrasound-guided transbronchial needle aspiration: a randomized trial. *Chest*. 2004;125(1):322-325.
10. Agency for Healthcare Research and Quality. Registries for evaluating patient outcomes. A user's guide. Rockville, MD: Agency for Healthcare Research and Quality; 2010. AHRQ publication 10-EHC049.
11. You YN, Jacobs L, Martinez ED, et al; American College of Surgeons Oncology Group. Improved surgeon performance in clinical trials: an analysis of quality assurance audits from the American College of Surgeons Oncology Group. *J Am Coll Surg*. 2006;203(3):269-276.
12. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138(5):837-843.
13. Mongan JJ, Ferris TG, Lee TH. Options for slowing the growth of health care costs. *N Engl J Med*. 2008;358(14):1509-1514.

14. Ernst A, Simoff M, Ost D, Michaud G, Chandra D, Herth FJ. A multicenter, prospective, advanced diagnostic bronchoscopy outcomes registry. *Chest*. 2010;138(1):165-170.

15. Ernst A, Simoff M, Ost D, Goldman Y, Herth FJ. Prospective risk-adjusted morbidity and mortality outcome analysis after therapeutic bronchoscopic procedures: results of a multi-institutional outcomes database. *Chest*. 2008;134(3):514-519.

16. *ACCP Quality Improvement Registry Education and Evaluation (AQuIRE)*. Northbrook, IL: American College of Chest Physicians; 2009. http://www.chestnet.org/accp/quality-improvement/aquire. Updated July 24, 2009. Accessed February 28, 2010.

17. Stoler MH, Schiffman M; Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*. 2001;285(11):1500-1505.

18. 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(4):234-241.

19. Annema JT, Versteegh MI, Veseliç M, et al. Endoscopic ultrasound added to mediastinoscopy for preoperative staging of patients with lung cancer. *JAMA*. 2005;294(8):931-936.

20. Tremblay A, Stather DR, Maceachern P, Khalil M, Field SK. A randomized controlled trial of standard vs endobronchial ultrasonography-guided transbronchial needle aspiration in patients with suspected sarcoidosis. *Chest*. 2009;136(2):340-346.

21. Adams K, Shah PL, Edmonds L, Lim E. 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(9):757-762.

22. Gu P, Zhao YZ, Jiang LY, Zhang W, Xin Y, Han BH. Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. *Eur J Cancer*. 2009;45(8)1389-1396.

23. Okamoto H, Watanabe K, Nagatomo A, et al. Endobronchial ultrasonography for mediastinal and hilar lymph node metastases of lung cancer. *Chest*. 2002;121(5):1498-1506.

24. Yasufuku K, Chiyo M, Koh E, et al. Endobronchial ultrasound guided transbronchial needle aspiration for staging of lung cancer. *Lung Cancer*. 2005;50(3):347-354.

25. Rintoul RC, Skwarski KM, Murchison JT, Wallace WA, Walker WS, Penman ID. Endobronchial and endoscopic ultrasound-guided real-time fine-needle aspiration for mediastinal staging. *Eur Respir J*. 2005;25(3):416-421.

26. Yasufuku K, Nakajima T, Motoori K, et al. Comparison of endobronchial ultrasound, positron emission tomography, and CT for lymph node staging of lung cancer. *Chest*. 2006;130(3):710-718.

27. Plat G, Pierard P, Haller A, et al. Endobronchial ultrasound and positron emission tomography positive mediastinal lymph nodes. *Eur Respir J*. 2006;27(2):276-281.

28. Pierard PE, Faber J, Bauwens O, et al. Evaluation of a new tool in mediastinal staging of lung cancer: Real-time guided needle aspiration using endobronchial ultrasound (EBUS-NA). *Chest*. 2006;130(4):146s.

29. Herth FJ, Ernst A, Eberhardt R, Vilmann P, Dienemann H, Krasnik M. Endobronchial ultrasound-guided transbronchial needle aspiration of lymph nodes in the radiologically normal mediastinum. *Eur Respir J*. 2006;28(5):910-914.

30. Krasnik M, Vilmann P, Larsen SS, Jacobsen GK. Preliminary experience with a new method of endoscopic transbronchial real time ultrasound guided biopsy for diagnosis of mediastinal and hilar lesions. *Thorax*. 2003;58(12):1083-1086.

31. Vilmann P, Krasnik M, Larsen SS, Jacobsen GK, Clementsen P. Transesophageal endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA) and endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) biopsy: a combined approach in the evaluation of mediastinal lesions. *Endoscopy*. 2005;37(9):833-839.

32. Herth FJ, Eberhardt R, Vilmann P, Krasnik M, Ernst A. Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. *Thorax*. 2006;61(9):795-798.

33. Vincent BD, El-Bayoumi E, Hoffman B, et al. Real-time endobronchial ultrasound-guided transbronchial lymph node aspiration. *Ann Thorac Surg*. 2008;85(1):224-230.

34. Wallace MB, Pascual JM, Raimondo M, et al. Minimally invasive endoscopic staging of suspected lung cancer. *JAMA*. 2008;299(5):540-546.

35. Herth FJ, Eberhardt R, Krasnik M, Ernst A. Endobronchial ultrasound-guided transbronchial needle aspiration of lymph nodes in the radiologically and positron emission tomography-normal mediastinum in patients with lung cancer. *Chest*. 2008;133(4):887-891.

36. Lee HS, Lee GK, Lee HS, et al. Real-time endobronchial ultrasound-guided transbronchial needle aspiration in mediastinal staging of non-small cell lung cancer: how many aspirations per target lymph node station? *Chest*. 2008;134(2):368-374.

37. Bauwens O, Dusart M, Pierard P, et al. Endobronchial ultrasound and value of PET for prediction of pathological results of mediastinal hot spots in lung cancer patients. *Lung Cancer*. 2008;61(3):356-361.

38. Donabedian A. The effectiveness of quality assurance. *Int J Qual Health Care*. 1996;8(4):401-407.

39. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260(12):1743-1748.