

- 7 Sabouchi-Schütt F, Aström J, Hellman U, *et al.* Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: a proteomics approach. *Eur Respir J* 2003; 21: 414–420.
- 8 Chang WC, Hsu PI, Chen YY, *et al.* Observation of peptide differences between cancer and control in gastric juice. *Proteomics Clin Appl* 2008; 2: 55–62.
- 9 Petricoin EF, Belluco C, Araujo RP, *et al.* The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat Rev Cancer* 2006; 6: 961–967.

DOI: 10.1183/09031936.00169509

# An evaluation of a videobronchoscopy-based autofluorescence system in lung cancer

To the Editors:

Autofluorescence (AF) has improved the sensitivity of bronchoscopy for pre-invasive cancer but has reduced specificity. Whether this applies to videobronchoscopy systems or depends on operator experience is unknown. We used a novel “slide-bank” model in order to evaluate the diagnostic performance of a video/AF bronchoscopy (AFI) system in lung cancer (refer to online supplementary material).

AF bronchoscopy systems are designed to enhance the detection of pre-invasive lesions. Abnormal bronchial epithelium fluoresces at an altered wavelength when illuminated by blue light. AF systems process and display these changes visually, revealing occult pre-invasive lesions.

Previously, the accuracy of AF systems in detecting pre-invasive lesions and cancer has been compared with white-light (WL) bronchoscopy [1, 2]. However, increases in sensitivity have been offset by reduced specificity, *i.e.* false positives. Furthermore, bronchoscopy and AF may be subject to significant, unreported, inter-operator variation. We evaluated an AFI system with experienced and inexperienced bronchoscopists to assess sensitivity/specificity, negative predictive value/positive predictive value (NPV/PPV) and inter-observer agreement.

Patients undergoing investigation for haemoptysis/lung cancer underwent bronchoscopy. Those with an established diagnosis of lung cancer involving the central airways were excluded.

The study was approved by the local research ethics committee (Royal Brompton and Harefield Hospital, London, UK). All eligible patients who provided written informed consent underwent bronchoscopy performed by a single experienced operator (P.L. Shah) using the AFI system, under local anaesthetic with or without conscious sedation. The endobronchial tree was systematically examined under WL and then AF. The appearances were classified by the operator as normal, inflammatory, suspicious or definite malignancy. The still images of the endobronchial tree under both modes were captured and saved.

We tested the AFI videobronchoscopy system (Olympus Medical Systems Corporation, Tokyo, Japan). This displays a high-quality WL image and can switch to AF mode by pressing a button. In AF mode the mucosa is illuminated by green and blue light. The combination of fluorescence and reflected green light is detected and processed into a composite image [3]. The on-screen image shows normal bronchial mucosa as green and neoplastic changes as magenta.

Histology from endobronchial biopsies were used as the gold standard for assessment. After completing an examination of the endobronchial tree, P.L. Shah took endobronchial biopsies from areas that appeared suspicious under either mode. Control areas that appeared normal under both modes were also biopsied in each patient. The pathologist (A.G. Nicholson) remained blinded to the bronchoscopic appearances and graded the biopsies according to the World Health Organization classification [4]. Intra-epithelial neoplasia was defined as moderate/severe dysplasia or carcinoma *in situ* [5].

Still images of biopsy sites were stored, creating a “slide-bank” library, and presented to individual observers as a slide show. They graded the appearance of the mucosa in each slide as either: normal, nonspecific abnormality/inflammatory, suspicious of cancer, or definite tumour. Five of the observers were experienced bronchoscopists who had performed >500 bronchoscopies and five were trainees with less experience (<100 bronchoscopies).

The histological findings were defined for statistical purposes as being positive (moderate dysplasia or worse) or negative. The real-time use of the AFI system by P.L. Shah was assessed by using the biopsy results to compare these to his grading of the bronchial mucosa under each AFI mode.

The slide show comprised both AF and WL images from each biopsied area, and paired testing was employed. The McNemar Chi-squared test for paired data (which confines analysis to divergent paired observations) was used to assess the sensitivity, specificity, PPV and NPV.

The difference between WL and AF modes and the effect of observer experience was quantified by the paired t-test for each value.

Inter-observer agreement was quantified using the weighted  $\kappa$ -coefficient of agreement (with quadratic weighting, such that  $\kappa$ -values are exactly synonymous with the intra-class correlation

This article has supplementary material accessible from [www.erj.ersjournals.com](http://www.erj.ersjournals.com)

coefficient). Weighted  $\kappa$ -values require computation between paired observers. Thus, group totals of 10 and 10 weighted  $\kappa$ -values were obtained, corresponding to all possible paired observer combinations for the five experienced and five inexperienced observers. Comparisons of  $\kappa$ -values between the WL and AF modes were made using paired t-tests of all 20 paired observer combinations. Comparisons of weighted  $\kappa$ -values between experienced and inexperienced observer pairings were made using the Wilcoxon rank sum test.

49 subjects (33 males; mean age 61 yrs) had 81 endobronchial areas biopsied. 128 still images of biopsied areas were arranged into the slide show in a random order.

Of the 81 areas biopsied, 11 (14%) were invasive carcinoma and five (6%) showed intra-epithelial neoplasia.

The sensitivity, specificity, PPV and NPV for the video-bronchoscopy mode/AF mode were 93.8/93.3%, 92.3/81.8%, 75.0/53.8% and 98.4/98.2%, respectively. There were a total of nine areas that appeared normal under WL but abnormal under AF. No biopsies from these nine areas showed intra-epithelial neoplasia; AF detected no more significant lesions than the WL mode. Of the 16 biopsies that showed at least moderate dysplasia, 15 (94%) were detected by both WL and AF and one (6%) was missed by both.

There was no difference between experienced and inexperienced observers in any measure of diagnostic accuracy using either WL or AF. When all 10 observers were grouped together there was no significant difference between the two bronchoscopic modes in any of the measures of diagnostic accuracy. However, there was a trend to reduced specificity using the AF mode ( $p=0.06$ ).

To clarify this trend the results were also analysed by considering the two groups of observers separately. The effect for the two groups in using the AF mode compared with WL is shown in table 1. The mean overall inter-observer agreement was  $\kappa=0.47$ .

The effect of experience and the bronchoscopic mode was seen by calculating separate  $\kappa$ -values. For both groups the inter-observer agreement was significantly better for the video-bronchoscopy mode *versus* AF mode:  $\kappa$  0.7–0.59 ( $p<0.01$ ) and  $\kappa$  0.71–0.63 ( $p=0.04$ ) for experienced and inexperienced, respectively.

Previous studies have compared real-time AF bronchoscopy with WL bronchoscopy for detecting pre-invasive lesions.

These studies showed that there was increased sensitivity for intra-epithelial neoplasia, at the expense of reduced specificity. They also showed a step-up in the sensitivity of the WL system when videobronchoscopes replaced fibreoptic scopes in 2005 of 20% to >60% [2, 3, 5–7]. The sensitivity of the WL mode in our analysis is high and we suggest this is a function of the clarity of the on-screen images. We are wary of concluding too much regarding the sensitivity/specificity of the AF mode from this small sample size. The incidence of pre-invasive lesions in our patient group is lower than the other previous studies; LAM *et al.* [5] had an incidence of 15% and CHIYO *et al.* [3] had an incidence of 39% compared to our 6%. This low incidence would tend to disadvantage the performance of AF. A fair conclusion from this data would be that the performance of AF depends on the pre-test probability of intra-epithelial neoplasia and, as such, AF bronchoscopy should not yet be a routine diagnostic instrument. It certainly does have a role in selected patient groups and as a research tool.

To our knowledge no previous assessment has been published of the agreement between multiple bronchoscopists independently looking at the same images. Our slide-bank model is a novel method of making this assessment. It could be argued that real-time moving images recorded at bronchoscopy would be more “real life” than static images. However, the decision by a bronchoscopist regarding the mucosal appearance at any one point is a “snap shot” and we believe a static image is a valid assessment tool. The degree of inter-observer agreement for both bronchoscopic modes and for both levels of experience in our study was reassuringly high. It is widely agreed that a substantial level of inter-observer agreement for a diagnostic test is a  $\kappa$ -value of >0.6, such as radiological assessment of high-resolution computed tomography scans [8]. It is not surprising that in our study the level of agreement was lower for the novel AF mode than for the familiar WL mode.

Our slide-bank assessment method presents previously unreported results from inexperienced bronchoscopists. First, their performance with the AF mode was better than the experienced bronchoscopists. This is curious and we can only speculate as to the reason. Perhaps they learnt faster. Secondly, their use of the AF mode did not lead to a drop in any of the measures of diagnostic performance, and in fact the NPV significantly increased. We suggest that improving the NPV in inexperienced bronchoscopists is clinically useful and important. At bronchoscopy the NPV is a measure of how often the bronchoscopists correctly assess an area as normal epithelium and, in reality, therefore choosing not to biopsy.

**TABLE 1** Assessment of the diagnostic value of videobronchoscopy (VB) and autofluorescence (AF) bronchoscopy according to the level of experience

Measure	Experienced VB	Experienced AF	p-value	Inexperienced VB	Inexperienced AF	p-value
<b>Sensitivity</b>	84.4±10.2	85.6±11.4	NS	82.2±8.1	88.9±8.8	NS
<b>Specificity</b>	77.8±12.1	64.8±8.4	<0.001	71.7±17.0	74.8±9.4	NS
<b>PPV</b>	59.8±14.5	48.7±8.6	0.038	53.2±12.1	58.0±7.9	NS
<b>NPV</b>	92.7±4.2	92.0±4.7	NS	91.2±1.9	94.5±3.6	<0.001

Data are presented as mean ± sd. PPV: positive predicted value; NPV: negative predicted value; NS: nonsignificant.

These data suggest that the extra sensitivity provided by AF bronchoscopy improves the ability of trainees to make the correct diagnosis at bronchoscopy and suggest a role for AFI in teaching hospitals in training inexperienced bronchoscopists.

Our preliminary study shows how a “slide bank” of bronchoscopic images can be used to assess the diagnostic performance of a bronchoscopic system by multiple operators. We believe it is valid and eminently transferable to other studies. This work needs to be repeated with a larger sample of patients and in a population with a higher incidence of intra-epithelial neoplasia.

**E.J. Cetti\***, **A.G. Nicholson<sup>#</sup>**, **S. Singh<sup>¶</sup>**, **A.U. Wells\*** and **P.L. Shah\*<sup>¶</sup>**

Depts of \*Respiratory Medicine and <sup>#</sup>Histopathology, Royal Brompton Hospital, and <sup>¶</sup>Dept of Respiratory Medicine, Chelsea and Westminster Hospital, London, UK.

**Correspondence:** P.L. Shah, Dept of Respiratory Medicine, Royal Brompton Hospital, Sydney Street, London, SW3 6NP, UK. E-mail: pallav.shah@imperial.ac.uk

**Statement of Interest:** Statements of interest for E.J. Cetti, A.U. Wells and P.L. Shah can be found at [www.erj.ersjournals.com/misc/statements.dtl](http://www.erj.ersjournals.com/misc/statements.dtl)

## REFERENCES

- 1 Lam S, Kennedy T, Unger M, *et al.* Localization of bronchial intraepithelial neoplastic lesions by fluorescence bronchoscopy. *Chest* 1998; 113: 696–702.
- 2 Ikeda N, Honda H, Hayashi A, *et al.* Early detection of bronchial lesions using newly developed videoendoscopy-based autofluorescence bronchoscopy. *Lung Cancer* 2006; 52: 21–27.
- 3 Chiyo M, Shibuya K, Hoshino H, *et al.* Effective detection of bronchial preinvasive lesions by a new autofluorescence imaging bronchovideoscope system. *Lung Cancer* 2005; 48: 307–313.
- 4 Travis WD, Brambilla E, Muller-Hermelink HK, eds. *In: World Health Organization Classification of Tumours. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus And Heart.* Lyon, IARC Press, 2004.
- 5 Lam S, Kennedy T, Unger M, *et al.* Localization of bronchial intraepithelial neoplastic lesions by fluorescence bronchoscopy. *Chest* 1998; 113: 696–702.
- 6 Vermeylen P, Pierard P, Roufosse C, *et al.* Detection of bronchial preneoplastic lesions and early lung cancer with fluorescence bronchoscopy: a study about its ambulatory feasibility under local anaesthesia. *Lung Cancer* 1999; 25: 161–168.
- 7 Ernst A, Simoff MJ, Mathur PN, *et al.* D-light autofluorescence in the detection of premalignant airway changes. a multicenter trial. *J Bronchol* 2005; 12: 133–138.
- 8 Aziz ZA, Wells AU, Hansell DM, *et al.* HRCT diagnosis of diffuse parenchymal lung disease: interobserver variation. *Thorax* 2004; 59: 506–511.

DOI: 10.1183/09031936.00148809

# Assaying for circulating osteopontin in practice: a technical note

*To the Editors:*

Plasma osteopontin is a very interesting marker for pulmonary diseases, being proposed for evaluating prognosis in patients with nonsmall cell lung cancer [1] or tuberculosis [2]. Osteopontin has also been implicated in the pathogenesis of idiopathic pulmonary fibrosis [3] and chronic obstructive pulmonary disease [4]. A more debated use of this marker has been its diagnostic utility in malignant pleural mesothelioma [5, 6]. Two major commercial kits are available, supplied by R&D systems, Minneapolis, MN, USA (Quantikine Human Osteopontin Immunoassay) and by Immuno-Biological Laboratories Co., Ltd, Gunma, Japan (Human osteopontin ELISA). It has been suggested in the literature that the results of osteopontin dosage are highly dependent on the choice of the ELISA kit. Indeed a previous publication reported important differences between available assays with little correlation between the results of the two commercial kits [7]. However, correlation is not the only analysis that should be performed when comparing two tests that measure the same variable [8].

We performed a comparative evaluation of the two commercial ELISA kits in a series of 75 samples of plasma retrieved

from 13 patients treated for tuberculosis. The samples were obtained at diagnosis, after 1, 2, 4 and 8 weeks, and 4 and 6 months of treatment. A standard operation procedure was in place for retrieving the samples and all patients gave informed consent. After centrifugation, plasma samples were immediately transferred in a -80°C freezer until assayed. A mean of six samples per patient (interquartile range 5–6) were available.

A correlation analysis between the two measurements yielded a Spearman  $\rho$  of only 0.4924 but this was statistically significant ( $p < 0.001$ ; fig. 1a). If we used the Pearson correlation coefficient (and thus assuming a normal distribution of data) this correlation was also low but statistically significant ( $r = 0.4515$ ;  $p < 0.001$ ). However, when we used a Bland–Altman plot to analyse the results, we obtained an almost straight line (fig. 1b). Thus, the lack of strong correlation between the two methods is, in fact, the result of a systematic bias between the two methods, which is proportional with the measured value. This type of result, called “proportional error” is, in fact, the result of the nonlinearity of the two measures and probably is a consequence of the amplification generated by different enzymes in the ELISA assay. Since we do not have a “gold standard”, we cannot say exactly which method is biased and