Early report

Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study

Michael Phillips, Kevin Gleeson, J Michael B Hughes, Joel Greenberg, Renee N Cataneo, Leigh Baker, W Patrick McVay

Summary

Background Many volatile organic compounds (VOCs), principally alkanes and benzene derivatives, have been identified in breath from patients with lung cancer. We investigated whether a combination of VOCs could identify such patients.

Methods We collected breath samples from 108 patients with an abnormal chest radiograph who were scheduled for bronchoscopy. The samples were collected with a portable apparatus, then assayed by gas chromatography and mass spectroscopy. The alveolar gradient of each breath VOC, the difference between the amount in breath and in air, was calculated. Forward stepwise discriminant analysis was used to identify VOCs that discriminated between patients with and without lung cancer.

Findings Lung cancer was confirmed histologically in 60 patients. A combination of 22 breath VOCs, predominantly alkanes, alkane derivatives, and benzene derivatives, discriminated between patients with and without lung cancer, regardless of stage (all $p<0{\cdot}0003$). For stage 1 lung cancer, the 22 VOCs had 100% sensitivity and 81·3% specificity. Cross-validation of the combination correctly predicted the diagnosis in 71·7% patients with lung cancer and 66·7% of those without lung cancer.

Interpretation In patients with an abnormal chest radiograph, a combination of 22 VOCs in breath samples distinguished between patients with and without lung cancer. Prospective studies are needed to confirm the usefulness of breath VOCs for detecting lung cancer in the general population.

Lancet 1999; **353:** 1930–33
See Commentary page 1897

Menssana Research Inc, Fort Lee, New Jersey; and Department of Medicine, St Vincent's Medical Center, Staten Island, New York (Prof M Phillips MD, J Greenberg BS, R N Cataneo MA)**, Department of Medicine, New York Medical College, Valhalla** (M Phillips, J Greenberg, R N Cataneo)**; Pulmonary/Critical Care Division, Department of Medicine, Penn State-Geisenger Health System, MS Hershey Medical Center, Hershey, Pennsylvania, USA** (K Gleeson MD)**; Department of Respiratory Medicine, Imperial College School of Medicine at Hammersmith Hospital, London, UK** (Prof J M B Hughes DM, L Baker BSc)**; and McVay Consulting Associates, Doylestown, Pennsylvania, USA** (W P McVay BS)

Correspondence to: Dr Michael Phillips, Department of Medicine, St Vincent's Medical Center, Staten Island, NY 10310, USA (e-mail: menssana@bellatlantic.net)

Introduction

Every year, in the USA, 99 000 men and 78 000 women develop lung cancer. 5 years after diagnosis, only 14% of these people are alive. If, however, the lung cancer is localised at the time of diagnosis and treated promptly, 5-year survival increases to 48%.[1] This fact has stimulated the search for screening tests to detect lung cancer at an early stage when it is probably localised.

Breath may contain clinically useful markers of lung cancer.[2] In 1971, Pauling and co-workers[3] reported that normal human breath contains a complex mixture of several hundred volatile organic compounds (VOCs). Since most VOCs are exhaled in picomolar concentrations, special methods are needed to collect and concentrate VOCs before assay. O'Neill and colleagues[4,5] identified 28 breath VOCs as candidate markers of lung cancer—principally alkanes such as hexane and methylpentane, and benzene derivatives. *o*-toluidine, aniline, and altered lipid-peroxidation activity have also been found in the breath of patients with lung cancer.[6,7]

In this study, we studied VOCs in the breath of patients with and without lung cancer.

Methods

In a cross-sectional study, eligible patients were those scheduled for bronchoscopy to investigate a localised chest-radiograph abnormality. Other inclusion criteria were: aged 18 or older, comprehension of the breath-collection procedure, and signed informed consent. Patients with known neoplasms of any kind were excluded. The study was approved by the institutional review boards of Penn State Medical Center, Hammersmith Hospital, and St Vincent's Medical Center.

Bronchoscopy was done by standard procedures.[8] After premedication with intramuscular meperidine and atropine, the patient's nose, nasopharynx, and oropharynx were sprayed with 1% lidocaine. Intraluminal lesions were washed or brushed for samples for cytology and a biopsy specimen was cut with standard alligator forceps. Parenchymal lesions had lavage of the appropriate airway segment for cytological washings and transbronchial biopsy under direct fluoroscopic guidance. Biopsy samples were preserved in formalin for histology. Patients with negative findings at bronchoscopy had additional investigations including computed tomography scans of the chest until the diagnosis of cancer was established or excluded. The tumour was staged by the tumour, node, metastasis (TNM) system for lung cancer.[1]

Alveolar breath samples were collected from patients after they had fasted overnight within 24 h before bronchoscopy: the breath-collection apparatus is a portable electrical device.[9] Patients wore a nose clip while breathing in and out of the device, via a disposable mouthpiece, for 5 min. A 10 L sample of breath was pumped through a sorbent trap which contained activated carbon and captured the VOCs for analysis. Ambient room air was collected on another sorbent trap. Each trap was stored in a hermetically sealed container for shipping to the laboratory.

Histology	
Small cell	10
Non-small cell	50
Epidermoid	24
Adenocarcinoma	23
Large cell	1
Mesothelioma	1
Melanoma	1
Stage	
X	3
I	9
II	3
IIIa	11
IIIb	7
IV	27

Table 1: **Histology and stage of lung cancers**

VOCs were thermally desorbed from each sorbent trap by automated instrumentation, concentrated by two-stage cryofocusing, separated by gas chromatography, then quantified and identified by mass spectroscopy.[9]

In a study of 50 healthy volunteers, the mean number of VOCs in a breath sample was 204 (SD 20). 3481 different VOC were found of which 1753 had a positive alveolar gradient. 27 VOCs were found in all 50 participants.[10]

All VOCs were tentatively identified and quantified. The chemical structure was identified from a computer-based library of mass-spectra. Each VOC was quantified by the ratio of the area under the curve (AUC) of the chromatographic peak to the AUC of a standard. The alveolar gradient,[11] the difference between the amounts in breath and in room air, was calculated as

$$(AUC_{breath} \div AUC_{standard}) - (AUC_{air} \div AUC_{standard})$$

The technicians analysing the breath samples were masked to the results of the bronchoscopy and biopsy findings. Similarly, the physicians who did the bronchoscopies and the pathologists who analysed the biopsy samples were masked to the results of the breath test.

Forward-stepwise discriminant analysis was used to identify VOCs that could discriminate between patients with and without lung cancer. The independent variable was the clinical stage of lung cancer, and the dependent variables were the alveolar gradients of a breath VOC found in more than 50% of all patients. The relative contribution of each VOC in the model was ranked by partial Wilks' lambda. The final model was then used to calculate the posterior probability of lung cancer in each subject from the breath sample. Discriminant analysis was also used to compare predictive values from the model with those based on demographic factors (age, tobacco smoking, and sex). A cross-validation of the patients classification was done by the SPSS "leave one out" discriminant analysis procedure which predicted whether a patient belonged to the group with or without lung cancer, based on the breath VOC model derived from all the other patients in the study.

Styrene (ethenylbenzene)
Heptane, 2,2,4,6,6-pentamethyl
Heptane, 2-methyl
Decane
Benzene, propyl-
Undecane
Cyclopentane, methyl-
Cyclopropane, 1-methyl-2-pentyl-
Methane, trichlorofluoro-
Benzene
Benzene, 1,2,4-trimethyl-
1,3-butadiene, 2-methyl- (isoprene)
Octane, 3-methyl-
1-hexene
Nonane, 3-methyl-
1-heptene
Benzene, 1,4-dimethyl
Heptane, 2,4-dimethyl
Hexanal
Cyclohexane
Benzene, 1-methylethenyl-
Hepatanal

*Chemical identification was tentative. Listed in descending order of contribution to model.

Table 2: **22 breath VOC picked out by discriminant analysis***

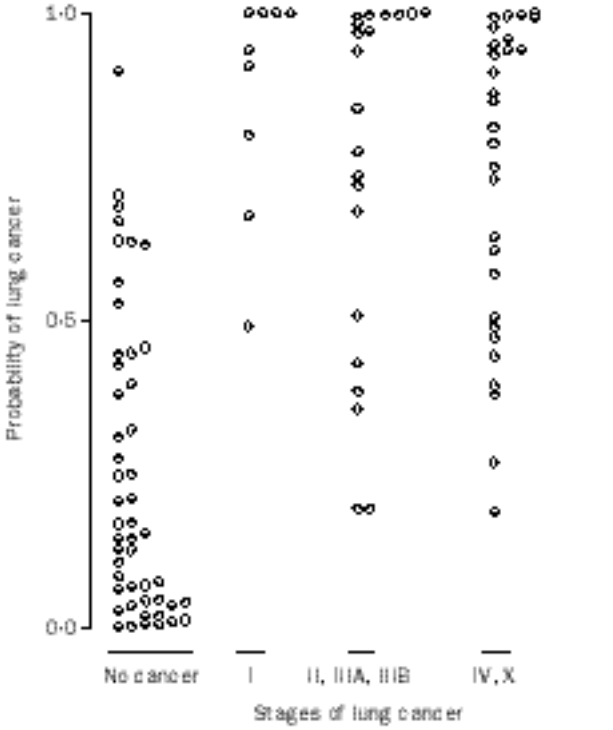

Figure 1: **Post-test probability of lung cancer by breath VOC assay**

Results

Between August, 1995, and October, 1996, 108 eligible patients agreed to participate. The collection of breath samples was not associated with any adverse effects. Lung cancer was confirmed histologically in 60 patients (34 men) and excluded in 48 patients (29 men). The mean (SD) age of patients was 66·9 years (12·5) in patients with lung cancer and 61 years (13·4) in patients without. Five patients with lung cancer had never smoked compared with 12 in the group without lung cancer. The histological diagnoses are shown in table 1.

67 VOCs were common to the breath samples of 62 (57·4%) patients; of these VOCs, 22 were selected by discriminant analysis (table 2). The mean post-test probability of lung cancer was significantly higher in patients with lung cancer than in those without lung cancer (all stages $p<0·0003$, figure 1). In patients with stage I lung cancer, a post-test probability of 0·46 had 100% sensitivity and 81·3% specificity; a post-test probability over 0·90 had 66·7% sensitivity and 100% specificity. The diagnosis was correctly predicted by the combination of age, tobacco smoking, and sex in 65·7% of cases, compared with 81·5% by the breath VOCs. Thus the breath VOCs provided diagnostic information independent of the demographic data. Cross-validation correctly predicted the diagnosis in 71·7% of patients with lung cancer and 66·7% of those without lung cancer.

Discussion

The 22 breath VOCs that discriminated between the patients with and without lung cancer were similar to those reported by O'Neill and co-workers[4] to be markers of lung cancer. There were some minor differences in

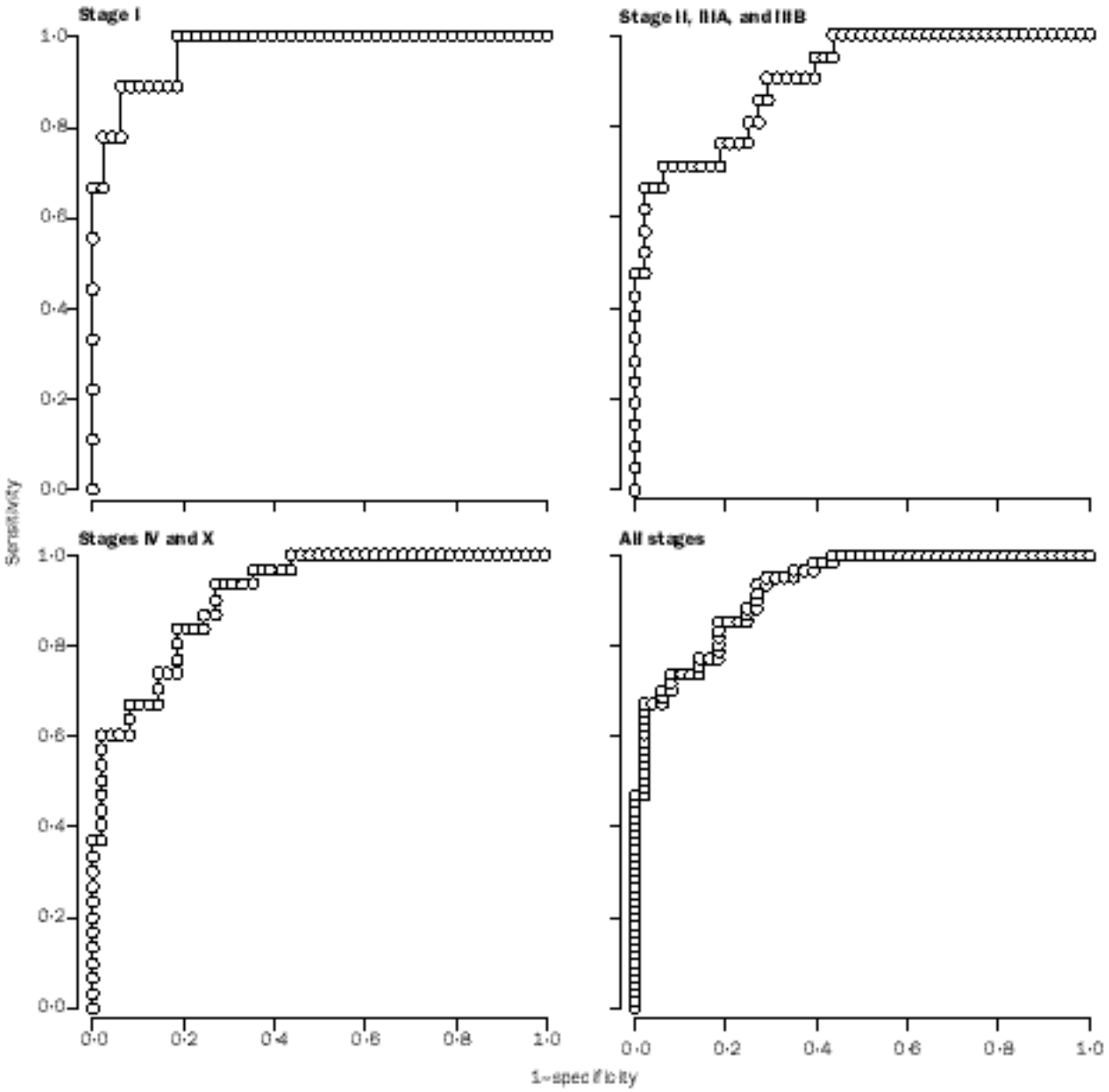

Figure 2: **Receiver-operating-characteristic curves of breath VOCs for different stages of lung cancer**

chemical structure which might be due to the use of different libraries of mass spectra. Structurally similar breath VOCs were observed in patients with and without lung cancer, but there were significant quantitative differences between the two groups.

The pathophysiology to explain our finding is not known. Part of the explanation may involve increased oxygen free-radical activity in cancerous cells.[12–15] Oxygen free radicals degrade cell membranes by lipid peroxidation and convert these polyunsaturated fatty acids to volatile alkanes that are excreted in the breath.[16,17] High concentrations of pentane in breath samples have been reported in breast cancer,[18] acute myocardial infarction,[19] heart transplant rejection,[20] rheumatoid arthritis,[21] and acute bronchial asthma.[22] Alkanes are cleared from the body mainly by excretion through the lungs or by oxidation to alkyl alcohols via the cytochrome P450 mixed-oxidase system.[23,24] 15 of the 22 VOCs were either alkanes or alkane derivatives and this structural similarity, particularly the five heptane derivatives, suggests an altered production of closely related compounds in the same metabolic pathway. In addition to the alkanes and alkane derivatives, six other VOCs were identical to those reported by O'Neill and colleagues;[4,5] isoprene, benzene, and four benzene derivatives. The source of the benzene and its derivatives is unknown.

Tobacco smoking cannot account for the benzene derivatives since these VOCs were found in the breath of non-smokers and ex-smokers (data not shown). Nor did smoking affect the VOC markers of lung cancer since smoking, age, and sex were not indirect mediators of the predictive value of the breath VOC model. Also, the most common breath VOC from tobacco smoking, 2,5-dimethylfuran,[25] was not among our 22 discriminatory VOCs.

There were no significant differences in sensitivity and specificity of breath VOCs between early and advanced lung cancer. This finding was unexpected, since the predictive power of most tumour markers generally

increases with tumour mass. However, this observation may have been skewed by the small number of patients with stage I lung cancer.

Bronchoscopy and biopsy are the "gold standards" for the diagnosis of lung cancer, but can occasionally miss a tumour. Some patients classified as "cancer-free" in this high-risk group may have had an occult neoplasm. It is less probable that there was a false-positive diagnosis from bronchoscopy and biopsy, although the specificity of the test is not well defined.[26–29]

The chemical identification of each VOC was tentative, based on the similarity of its mass spectrum to that in a computer-based library. The fit between the breath and library spectrum was generally high, but definite identification will need an analytical procedure, such as establishing the chromatographic elution time of the pure reagent.

For each patient, the outcome of the breath VOC assay was expressed as a probability of disease. Although a high probability by breath VOC analysis may indicate lung cancer, it is not a definitive diagnosis. Since this was a cross-sectional study of a high-risk group, the predictive value of the breath test for screening an unselected population is not yet known. In practical terms, this would require an optimum combination of sensitivity and specificity. Further studies to investigate the use of breath VOCs in the general population, in whom the prevalence of lung cancer is relatively low, may indicate that high specificity is more desirable than high sensitivity to avoid an excess of false-positive findings.[30]

Finally, this study detected a combination of 22 VOCs in the breath that were the "fingerprint" of lung cancer. As the number of variables in a statistical model increases, so too does the risk of observing significant differences arising from chance associations. There are three reasons why random statistical associations are unlikely to account for our findings. First, the VOCs were similar to those described in other reports of breath VOCs in lung cancer. Second, alkanes in the breath are consistent with a possible mechanism via oxygen free-radical activity in cancer. Third, cross-validation tests of the predictive model correctly classified the majority of patients with and without lung cancer. Nonetheless, these findings should be regarded as tentative, and validation studies in large numbers of patients and the general population will be required before widespread use can be recommended.

Contributors

Michael Phillips co-ordinated the study: he is chief executive of Menssana Research Inc; holds a patent on the breath-collection apparatus; and has applied for a patent on the findings reported in this manuscript. Kevin Gleeson and Michael Hughes supervised the collection of samples and data. Joel Greenberg and Renee Cataneo analysed the breath samples and these data. Leigh Baker collected breath samples and patient's data in the UK. Patrick McVay built the breath-collection apparatus and maintained these instruments. All investigators were involved in the writing of the paper.

Acknowledgments

This research was supported by a grant from the Ben Franklin Technology Center of Central and Northern Pennsylvania and Sunrise DeVilbiss Health Care, Inc. We thank Donald A Brand for assistance with database management, Eugene A Sersen and Sheldon Blackman for assistance with statistical analysis, and Supelco Inc, for the sorbent traps.

References

1 Minna JD. Neoplasms of the lung. In: Fauci AS, Braunwald E, Isselbacher KJ, et al. eds. Harrison's principles of internal medicine, 14th edn. New York: McGraw Hill, 1998.

2 Phillips M. Breath tests in medicine. *Sci Am* 1992; **267:** 74–79.

3 Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Nat Acad Sci USA* 1971; **68:** 2374–76.

4 O'Neill HJ, Gordon SM, O'Neill MH, Gibbons RD, Szidon JP. A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. *Clin Chem* 1988; **34:** 1613–18.

5 Gordon SM, Szidon JP, Krotoszynski BK, Gibbons RD, O'Neill HJ. Volatile organic compounds in exhaled air from patients with lung cancer. *Clin Chem* 1985; **31:** 1278–82.

6 Preti G, Labows JN, Kostelc JG, Aldinger S, Daniele R. Analysis of lung air from patients with bronchogenic carcinoma and controls using gas chromatography mass spectrometry. *J Chromatogr* 1988; **432:** 1–11.

7 Khyshiktyev BS, Khyshiktueva NA, Ivanov VN, Darenskaia SD, Novikov SV. Diagnostic value of investigating exhaled air condensate in lung cancer (in Russian). *Vopr Onkol* 1994; **40:** 161–64.

8 Prakash UBS. Optimal bronchoscopy. *J Bronchology* 1994; **1:** 4–62.

9 Phillips M. Method for the collection and assay of volatile organic compounds in breath. *Anal Biochem* 1997; **247:** 272–78.

10 Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, Cataneo RN. Variation in volatile organic compounds in the breath of normal humans. *J Chromatograph B Biomed Sci Appl* (in press).

11 Phillips M, Greenberg J, Sabas M. Alveolar gradient of pentane in normal human breath. *Free Radic Res* 1994; **20:** 333–37.

12 Dreher D, Junod AF. Role of oxygen free radicals in cancer development. *Eur J Cancer* 1996; **32A:** 30–38.

13 Feig DI, Reid TM, Loeb LA. Reactive oxygen species in tumorigenesis. *Cancer Res* 1994; **43** (7 suppl): 1890s–94s.

14 Knight JA. Diseases related to oxygen-derived free radicals. *Ann Clin Lab Sci* 1995; **25:** 111–21.

15 Borek C. Molecular mechanisms in cancer induction and prevention. *Environ Health Perspect* 1993; **101** (suppl 3): 237–45.

16 Kneepkens CMF, Ferreira C, Lepage G, Roy CC. The hydrocarbon breath test in the study of lipid peroxidation: principles and practice. *Clin Invest Med* 1992; **15:** 163–86.

17 Kneepkens CMF, Lepage G, Roy CC. The potential of the hydrocarbon breath test as a measure of lipid peroxidation. *Free Radic Biol Med* 1994; **17:** 127–60.

18 Hietanen E, Bartsch H, Bereziat JC, et al. Diet and oxidative stress in breast, colon and prostate cancer patients: a case control study. *Eur J Clin Nutr* 1994; **48:** 575–86.

19 Weitz ZW, Birnbaum AJ, Sobotka PA, Zarling EJ, Skosey JL. High breath pentane concentrations during acute myocardial infarction. *Lancet* 1991; **337:** 933–35.

20 Sobotka PA, Gupta DK, Lansky DM, Costanzo MR, Zarling EJ. Breath pentane is a marker of acute cardiac allograft rejection. *J Heart Lung Transplant* 1994; **13:** 224–29.

21 Humad S, Zarling E, Clapper M, Skosey JL. Breath pentane excretion as a marker of disease activity in rheumatoid arthritis. *Free Rad Res Commun* 1988; **5:** 101–06.

22 Olopade CO, Zakkar M, Swedler WI, Rubinstein I. Exhaled pentane levels in acute asthma. *Chest* 1997; **11:** 862–65.

23 Crosbie SJ, Blain PG, Williams FM. Metabolism of n-hexane by rat liver and extrahepatic tissues and the effect of cytochrome P-450 inducers. *Hum Exp Toxicol* 1997; **16:** 131–37.

24 Scheller U, Zimmer T, Kargel E, Schunck WH. Characterization of the n-alkane and fatty acid hydroxylating cytochrome P450 forms 52A3 and 52A4. *Arch Biochem Biophys* 1996; **328:** 245–54.

25 Gordon SM. Identification of exposure markers in smokers' breath. *J Chromatogr* 1990; **511:** 291–302.

26 Govert JA, Kopita JM, Matchar D, Kussin PS, Samuelson WM. Cost-effectiveness of collecting routine cytologic specimens during fiberoptic bronchoscopy for endoscopically visible lung tumour. *Chest* 1996; **109:** 451–56.

27 Mak VH, Johnston ID, Hetzel MR, Grubb C. Value of washings and brushings at fibreoptic bronchoscopy in the diagnosis of lung cancer. *Thorax* 1990; **45:** 373–76.

28 Popp W, Merkle M, Schreiber B, Rauscher H, Ritschka L, Zwick H. How much brushing is enough for the diagnosis of lung tumours? *Cancer* 1992; **70:** 2278–80.

29 Popp W, Rauscher H, Ritschka L, Redtenbacher S, Zwick H, Dutz W. Diagnostic sensitivity of different techniques in the diagnosis of lung tumors with the flexible fiberoptic bronchoscope. *Cancer* 1991; **67:** 72–75.

30 Brawley OW, Kramer BS. Prevention and early detection of cancer. In: Fauci AS, Braunwald E, Isselbacher KJ, et al. eds. Harrison's principles of internal medicine, 14th edn. New York: McGraw Hill, 1998.