

## Discriminant analysis on small cell lung cancer and non-small cell lung cancer by means of NSE and CYFRA-21.1

G. Paone\*, G. De Angelis\*\*, R. Munno+, G. Pallotta\*\*, D. Bigioni\*\*,  
C. Saltini++, A. Bisetti\*, F. Ameglio#

*Discriminant analysis on small cell lung cancer and non-small cell lung cancer by means of NSE and CYFRA-21.1. G. Paone, G. De Angelis, R. Munno, G. Pallotta, D. Bigioni, C. Saltini, A. Bisetti, F. Ameglio. ©ERS Journals Ltd 1995.*

**ABSTRACT:** A correct diagnosis of small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) is essential both for prognostic and therapeutic reasons. We used discriminant analysis as a method to optimize the discriminant power of serum tumour marker levels for differentiation between SCLC and NSCLC.

A panel of serum markers, including neurone specific enolase (NSE), cytokeratin fragment antigen 21.1 (CYFRA-21.1), tissue polypeptide antigen (TPA) and carcinoembryonic antigen (CEA) was obtained in 50 consecutive NSCLC and 17 SCLC. Data were analysed by the BMDP statistical program after logarithmic transformation of marker levels.

The variables selected were NSE and CYFRA-21.1. Considered together, they were able to give a 97% rate of correct classification. The formula generated (canonic variable, CV) was validated on a group of seven SCLC and 22 NSCLC patients. Only two errors occurred.

We therefore conclude that the canonic variable tested, based on NSE and CYFRA-21.1, provides a good discrimination between the two types of lung cancer. The method is rapid, relatively inexpensive, and based on simple serum tests.

*Eur Respir J., 1995, 8, 1136–1140.*

\*Dept of Cardiovascular and Respiratory Sciences, University "La Sapienza", Rome, Italy. \*\*III Division of Pneumology, and +Centre of Nuclear Medicine, C. Forlanini Hospital, Rome, Italy. ++Dept of Medical, Oncologic and Radiological Sciences, University of Modena, Italy. #Laboratory of Clinical Pathology, S. Galliciano Institute, Rome, Italy.

Correspondence: G. Paone  
I Clinica Tisiologica  
Ospedale C. Forlanini  
Piazza C. Forlanini 1  
00151, Roma  
Italy

Keywords: Carcinoembryonic antigen, CYFRA 21.1, discriminant analysis, lung cancer, neuron-specific enolase, tissue polypeptide antigen

Received: June 2 1994

Accepted after revision March 25 1995

Lung cancer represents the most frequent malignant neoplasm in man [1], with an incidence of 33,000 new cases annually in Italy [2], and 600,000 deaths worldwide [3]. The latter figure may be an underestimation due to the lack of reliable statistics in many countries.

Lung cancers are classified into four major cell types by histology: small cell lung cancer (SCLC); lung adenocarcinoma (LADC); squamous cell lung cancer (SQCLC); and large cell lung cancer (LCLC) [4]; the last three types being grouped together as non-small cell lung cancer (NSCLC). Other known types of lung cancer are not represented in our study.

Differentiation between SCLC and NSCLC is very important for prognostic and therapeutic reasons, due to their different behaviour [5, 6].

In addition to histology, an alternative diagnostic methodology may be useful, especially if the system is based on simple laboratory tests, performed on serum. Until now, by using single tumour markers, it was not possible to classify lung cancer as SCLC or NSCLC. Some authors tried to improve the discrimination rate by combining several markers, but with limited success [7–13].

Discriminant analysis is a mathematical method, which may be applied to a set of markers and which may improve the classification power of any single variable, increasing the final discrimination rate [14–17].

By means of a suitable computer program [18, 19], working with four tumour markers, we investigated the possibility of enhancing differentiation between SCLC and NSCLC by discriminant analysis.

### Patients and methods

#### *Patient population*

From January 1993 to December 1993, a first group of 67 consecutive unselected and untreated patients with newly diagnosed lung cancer was evaluated in a division of the Forlanini Hospital (Rome, Italy) to generate a canonic variable (CV) for discrimination between SCLC and NSCLC using serum tumour marker levels.

A second group of 29 patients was then enrolled to validate the discriminant power of the canonic variable in patients not previously used to generate the algorithm [20].

Histological types and disease stages of both study groups are displayed in table 1.

#### *Lung cancer diagnosis*

Histological diagnosis was made by at least two pathologists, following World Health Organisation (WHO)

Table 1. – Characteristics of the groups of patients used: a) to generate the canonic variable (n=67); and b) to validate the canonic variable (n=29)

Histology	Patients	Sex M/F	Age yrs*	Stages							
				I	II	IIIa	IIIb	IV	LD	ED	
<b>a) To generate the canonic variable</b>											
LADC	16	12/4	63±10	1	0	2	6	7			
SQCLC	32	29/3	70±8	6	3	6	3	14			
LCLC	2	2/0	63±8	0	1	0	1	0			
SCLC	17	13/4	61±10						6	11	
<b>b) To validate the canonic variable</b>											
LADC	7	5/2	57±11	1	1	0	2	3			
SQCLC	12	13/2	68±10	3	0	2	2	8			
LCLC	3	2/1	62±7	1	0	1	0	1			
SCLC	7	5/2	56±7						2	5	

M: male; F: female; LD: limited disease; ED: extensive disease; LADC: lung adenocarcinoma; SQCLC: squamous cell lung cancer; LCLC: large cell lung cancer; SCLC: small cell lung cancer. \*:data for age are presented as mean±SD.

criteria [4]. For each patient, three sputum samples and bronchoscopic biopsies were routinely evaluated. Needle transthoracic aspiration was needed for seven patients and resected tissues for three. Serum marker data were not known by the pathologists.

#### Tumour markers

Sera were obtained by venipuncture and stored at -20°C until assayed for the following tumour markers: 1) carcinoembryonic antigen (CEA) (Sorin Biomedica, Saluggia, Italy; normal values less than 5.5 ng·mL<sup>-1</sup>) [9]; 2) tissue polypeptide antigen, (TPA) (Byk Sangtec, Cormano, Italy; normal values less than 95 U·mL<sup>-1</sup>) [11]; 3) a cytokeratin fragment antigen CYFRA-21.1, (CIS Diagnostici, Vercelli, Italy; normal values less than 3.3 ng·mL<sup>-1</sup>) [8, 10]; and 4) neurone-specific enolase (NSE), (CIS Diagnostici, Vercelli, Italy; normal values less than 12.5 ng·mL<sup>-1</sup>) [9]. All tests (radioimmunoassays) were performed in duplicate, following the manufacturers' instructions.

The laboratory personnel did not know the histological diagnosis.

#### Statistical analysis

Due to non-normal distribution of the raw data, a logarithmic transformation was needed. To permit comparisons with other studies, results are presented as medians and ranges of the individual non-transformed data and as mean±SD of numbers obtained after the Ln transformation.

Comparisons or correlations were evaluated by using non-parametric tests (Kruskall Wallis one-way variance analysis or Spearman Rank test, respectively) on the raw data, and by means of Student's t-test on normalized data.

Discriminant analysis, a multiparametric test, was performed by using a computer furnished with the BMDP program (BMDP, statistical software, University of California, USA, P7M module) [18, 19].

This method generates an index, negative or positive, able to separate two groups. This index is named canonic variable (CV). CV=0 serves as the cut-off point. Patients with CV >0 are classified as SCLC, patients with CV<0 are classified as NSCLC. Inclusion of the variables into the CV formula was obtained by the criterion of corrected means by F statistics. The analysis was performed using the option of the "equal prior probability" to assign the subjects to groups.

The jack-knifed approach was used to discriminate the patients. With this method, each patient is evaluated by a canonic variable generated after exclusion of the same patient data. Furthermore, the canonic variable formula was also validated on another group of patients enrolled consecutively after the first [20].

#### Sample size

As suggested by LACHENBRUCH [21], the number of subjects considered must be at least five times greater than that of the variables selected. In our case, this recommendation was largely satisfied. The significance level of the discrimination rate obtained in the study, indicated that the sample size was sufficient.

## Results

As shown in table 2, only CYFRA-21.1 and NSE were statistically different between the two types of lung cancer.

Figure 1 compares the distributions of Ln CYFRA-21.1 and Ln NSE in SCLC and NSCLC patients. NSE has the highest discrimination power (21 patients misclassified by CYFRA-21.1 and eight by NSE).

Correlation matrices of the four variables in the groups are shown in table 3a (SCLC) and 3b (NSCLC). Serum levels of CEA and TPA, TPA and CYFRA-21.1, and TPA and NSE were significantly correlated in both cancer types. In NSCLC the levels of CEA and NSE, and in SCLC the levels of CYFRA-21.1 and NSE, were also correlated.

Table 2. – Medians and ranges of the variables analyzed and means and standard deviations of Ln-transformed levels of tumour markers

a)		CEA ng·mL <sup>-1</sup>	TPA U·mL <sup>-1</sup>	CYFRA-21.1 ng·mL <sup>-1</sup>	NSE ng·mL <sup>-1</sup>
SCLC	Median	4.3	82	2.5	19
(n=17)	range	1–70	25–330	0.6–7.6	10–97
NSCLC	Median	4.8	139	8.7	8.2
(n=50)	range	0.3–100	7.2–1374	0.3–150	4.5–15.9
p-value		0.32	0.13	0.003	<0.0001
b)		Ln CEA ng·mL <sup>-1</sup>	Ln TPA U·mL <sup>-1</sup>	Ln CYFRA-21.1 ng·mL <sup>-1</sup>	Ln NSE ng·mL <sup>-1</sup>
SCLC	Mean	1.51	4.48	0.72	3.28
(n=17)	SD	0.99	0.77	0.84	0.68
NSCLC	Mean	1.68	4.90	1.89	2.07
(n=50)	SD	1.07	1.05	1.52	0.29
p-value		0.59	0.14	0.004	<0.0001

CEA: carcinoembryonic antigen; TPA: tissue polypeptide antigen; CYFRA-21.1: cytokeratin fragment antigen; NSE: neurone-specific enolase; SCLC: small cell lung cancer; NSCLC: non-small cell lung cancer

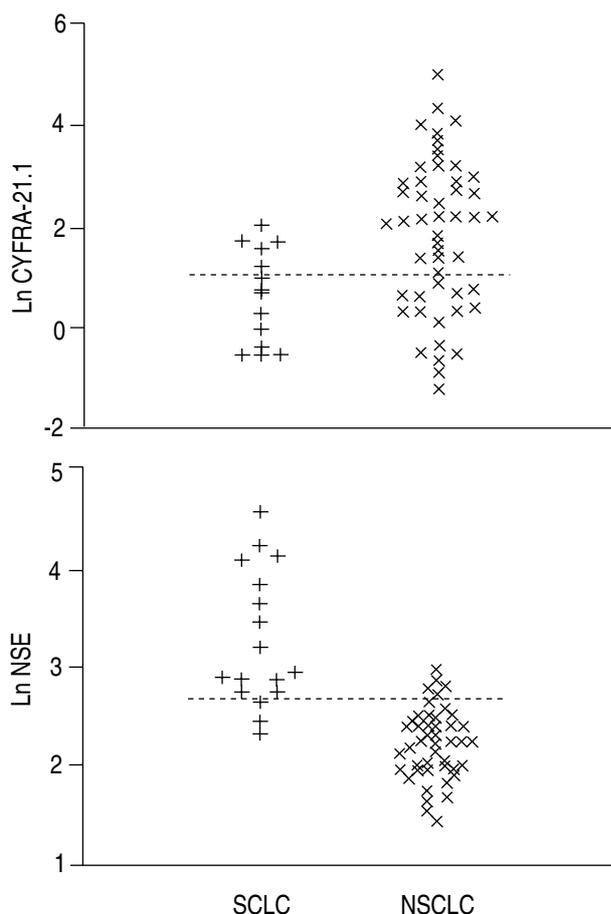


Fig. 1. – Serum level distributions of Ln NSE and Ln CYFRA-21.1 in 17 SCLC (+) and 50 NSCLC (x) patients. -----: discriminant cut off; SCLC: small cell lung cancer; NSCLC: non-small cell lung cancer; NSE: neurone specific enolase; CYFRA: cytokeratin fragment antigen

Table 3. – Correlation coefficients (r) matrix for: SCLC groups and NSCLC group

Variables	CEA	TPA	CYFRA-21.1	NSE
<b>SCLC group</b>				
CEA	1.00			
TPA	0.56*	1.00		
CYFRA-21.1	0.01	0.72***	1.00	
NSE	0.26	0.71**	0.75***	1.00
<b>NSCLC group</b>				
CEA	1.00			
TPA	0.64***	1.00		
CYFRA-21.1	0.23	0.46***	1.00	
NSE	0.33*	0.37**	0.23	1.00

For abbreviations see legend to table 2. \*: p<0.05; \*\*: p<0.01; \*\*\*: p<0.001.

Applying the discriminant analysis to SCLC and NSCLC, a good classification was obtained by means of Ln NSE and Ln CYFRA-21.1. The results are reported in figure 2; the calculation formula is reported in the legend to this figure. CV showed significantly different values between the two groups: SCLC (mean±SD 1.71±1.42 *versus* NSCLC -1.62±0.81, (p<0.0001).

The statistical evaluation of the correlation found between the two classification systems was obtained by canonic correlation (r coefficient = 0.83; p<0.0001).

As reported in the legend to figure 2, the Ln NSE variable coefficient is positive in contrast to that of Ln CYFRA-21.1, indicating that these two variables have an opposite effect in determining the final CV value, as reported previously [7–10, 12]. The most important discriminant variable is NSE (see fig. 1).

TPA and CEA were not selected, possibly due to non-significant differences found between the two groups of subjects and strong correlations with CYFRA-21.1 and NSE markers (tables 2 and 3).

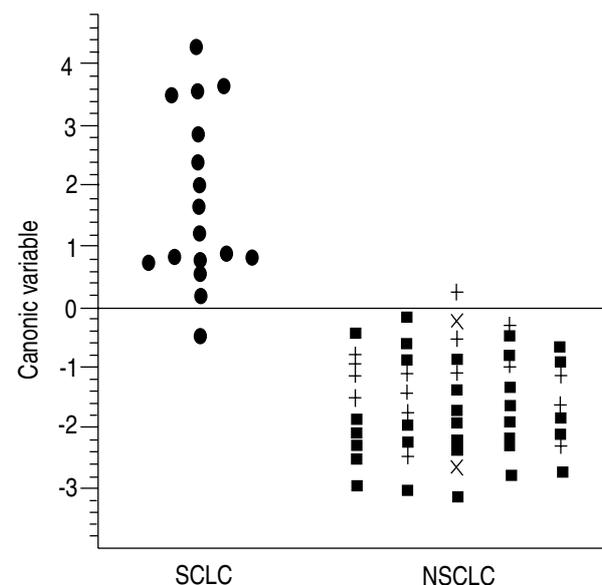


Fig. 2. – Discriminant analysis generated on 50 NSCLC and 17 SCLC patients by means of Ln NSE and Ln CYFRA-21.1. Formula: CV = Ln NSE × 2.37032 - Ln CYFRA-21.1 × 0.37699 - 5.55988. ●: small cell lung cancer (SCLC); +: lung adenocarcinoma; ■: squamous cell lung carcinoma; x: large cell lung carcinoma. NSCLC: non-small cell lung cancer.

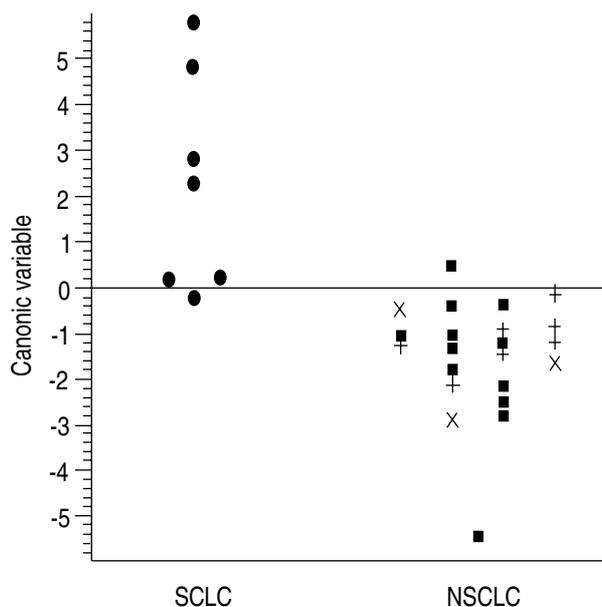


Fig. 3. – Discriminant analysis validation on 22 NSCLC and 7 SCLC patients by means of the formula reported in figure 2. ●: small cell lung cancer (SCLC); +: lung adenocarcinoma; ■: squamous cell lung carcinoma; ×: large cell lung cancer. NSCLC: non-small cell lung carcinoma.

To validate the formula previously generated, a control group (table 1b) was used. The rate of correct classification obtained was 93%, with only two errors: one patient affected with NSCLC and one with SCLC (fig. 3). Although misclassified, these two patients presented CV values between -0.5 and 0.5.

Early stages (I or limited disease (LD)) were misclassified in two cases, 1 of 13 belonging to the first group (table 1a) and 1 of 7 to the validation group (table 1b).

A further attempt to improve the correct classification rate by adding the squamous cell carcinoma antigen (SCC-Ag) and carbohydrate antigenic determinant 19-9 (Ca 19-9) [22] did not improve the results (data not shown).

### Discussion

Due to the different biology, prognosis and sensitivity to therapy of SCLC and NSCLC, their differentiation is very important. Generally, this aim is obtained using histological techniques. The recognition that lung cancer is often associated with changes in the levels of various plasma markers, suggests their possible employment as diagnostic and discriminant indices. For this purpose, several studies have been performed to evaluate the ability of tumour markers to diagnose and to differentiate the various histological types of lung cancer [7–13].

Discriminant analysis represents one of the best methods to associate the discriminant power of more variables to obtain the maximum classification between two or more groups. This methodology is currently applied in taxonomy. Several examples of the use of this method are also reported in the literature concerning different fields [14–17, 20]. Recently, serum tumour marker levels were reliably used to distinguish between primary and metastatic malignant bone tumours [16].

Our study attempted to optimize the use of some common markers of lung cancer (TPA, CEA, NSE and CYFRA-21.1), to differentiate SCLC and NSCLC by means of discriminant analysis. To our knowledge, no other reports exist in the literature for this topic.

The canonic variable generated was able to separate SCLC from NSCLC with an overall 97% (98% on NSCLC and 94% on SCLC) rate of correct classification, whilst no acceptable classification was obtained among LADC, SQCLC and LCLC.

One of the critical points for the correct use of discriminant analysis is the validation of the canonic variable generated. In fact, overestimation cannot be ruled out, since the formula directly derives from the data of the group selected. To overcome this problem, a second group of patients, not previously employed to generate the algorithm, was evaluated confirming the validity of the formula generated.

Generally, histology furnishes both cancer diagnosis and histological typing; whilst, due to the type of groups used, the formula obtained cannot be employed to recognize lung cancer from other diseases, but only to distinguish histological types. Therefore, a previous clinical examination was necessary to indicate the presence of a lung malignancy.

In addition to the theoretical importance of CV in recognizing histological types of cancer, a clinical role for this method may be represented by those cases where histology cannot be obtained. For instance, there are subsets of patients with poor cardiorespiratory function and negative sputum cytology in which fibrobronchoscopy, needle transthoracic aspiration and thoracotomy may not be performed. Although histology remains the reference method, a relatively simple serological test may be helpful to obtain a presumptive diagnosis of lung cancer type in these patients.

Consistent with the notion that tumour marker levels in the serum are more readily detectable within advanced stages, one might expect that this multimarker serological test may not be employed with success in early stage cancers. However, although the relative frequency of misclassification was higher in early stage cancers, the rate of correct classification (90%) appears promising.

In conclusion, the results indicate that, the use of discriminant analysis with a small panel of markers may be useful to differentiate SCLC and NSCLC, especially when applied to an appropriate subset of lung cancer patients.

### References

1. Samet JM. The epidemiology of lung cancer. *Chest* 1993; 103: 20S–29S.
2. Capocaccia R, Valente F, Verdecchia A. Incidence and prevalence of cancers estimated from the official mortality data in Italy. In: Zanetti R, Crosignani P, eds. *Cancer in Italy*. Torino, Tipografia Silvestrelli e Cappelletto, 1992.
3. Aisner J, Belani CP. Lung cancer: recent changes and expectations of improvements. *Semin Oncol* 1993; 20: 383–393.
4. World Health Organization. The World Health Organization

- histological typing of the lung tumors. *Am J Clin Pathol* 1982; 77: 123–136.
5. Sandler AB, Buzaid AC. Lung cancer: a review of current therapeutic modalities. *Lung* 1992; 170: 249–265.
  6. Ihde DC, Minna JD. Non-small cell lung cancer. Part I. Biology, diagnosis and staging. *Curr Probl Cancer* 1991; 15: 61–104.
  7. Lombardi C, Tassi GF, Pizzocolo G, Donato F. Clinical significance of a multiple biomarker assay in patients with lung cancer: a study with logistic regression analysis. *Chest* 1990; 97: 639–644.
  8. Stieber P, Hasholzenr U, Bodenmuller H, *et al.* CYFRA 21-1: a new marker in lung cancer. *Cancer* 1993; 72: 707–713.
  9. Jarvisalo J, Hakama M, Knekt P, *et al.* Serum tumor markers CEA, CA 50, TATI and NSE in lung cancer screening. *Cancer* 1993; 71: 1982–1988.
  10. Pujol JL, Grenier J, Daurès JP, Daver A, Pujol H, Michel FB. Serum Fragment of cytokeratin subunit 19 measured by CYFRA 21-1: immunoradiometric assay as a marker of lung cancer. *Cancer Research* 1993; 53: 61–66.
  11. Buccheri GF, Ferrigno D, Sartoris AM, Violante B, Vola F, Curcio A. Tumor markers in bronchogenic carcinoma: superiority of tissue polypeptide antigen to carcinoembryonic antigen and carbohydrate antigenic determinant 19-9. *Cancer* 1987; 60: 42–50.
  12. Anastasiades KD, Mullins RE, Conn RB. Neuron-specific enolase: assessment by ELISA in patients with small cell carcinoma of the lung. *Am J Clin Pathol* 1987; 87: 245–249.
  13. Buccheri G, Violante B, Sartoris AM, Ferrigno D, Curcio A, Vola F. Clinical value of a multiple biomarker assay in patients with bronchogenic carcinoma. *Cancer* 1986; 57: 2389–2396.
  14. Ameglio F, Abbolito MR, Giannarelli D, *et al.* Detection of *Helicobacter pylori* carriers by discriminant analysis of urea and pH levels in gastric juices. *J Clin Pathol* 1991; 44: 697–698.
  15. Merucci P, Taggi F, Marolla A, *et al.* Discriminant analysis of Hodgkin's and non-Hodgkin's lymphomas by age and serum proteins. *Eur J Cancer Clin Oncol* 1984; 20: 1243–1247.
  16. Shinozaki T, Chigira M, Kato K. Multivariate analysis of serum tumor markers for diagnosis of skeletal metastases. *Cancer* 1992; 69: 108–112.
  17. Ameglio F, Giannarelli D, Cordiali-Fei P, *et al.* Use of discriminant analysis to assess disease activity in pulmonary tuberculosis with a panel of specific and non-specific serum markers. *Am J Clin Pathol* 1994; 101: 719–725.
  18. BMDP statistical software. Department of Biomathematics. University of California, Los Angeles, University of California Press.
  19. Overall JE, Kleitt JC. Applied multivariate analysis. Chapter 14. Use of canonical variate for classification. New York, McGraw Hill Inc, 1978.
  20. Thompson EM, Proce AB, Altman DG, Sowter C, Slavin G. Quantitation in inflammatory bowel disease using computerised interactive image analysis. *J Clin Pathol* 1985; 38: 631–638.
  21. Lachenbruch PA. Some misuses of discriminant analysis. *Methods Inf Med* 1977; 16: 255–258.
  22. Niklinski J, Furman M, Laudanski J, Kozlowki M. Prognostic value of pretreatment CEA, SCC-Ag and CA 19-9 levels in sera of patients with non-small cell lung cancer. *Eur J Cancer Prev* 1992; 1: 401–406.