

Centre-independent Detection of Non-small Cell Lung Cancer (NSCLC) by Means of Classification with Receiver Operating Characteristic (ROC)-based Data Transformation

NORMAN BITTERLICH¹, THOMAS MULEY² and JOACHIM SCHNEIDER³

¹Medizin und Service GmbH, Chemnitz, Germany;

²Translational Research Unit, Thoraxklinik am Universitätsklinikum Heidelberg, Heidelberg, Germany;

³Institut und Poliklinik für Arbeits- und Sozialmedizin der Justus-Liebig-Universität, Giessen, Germany

Abstract. *The diagnosis of non-small cell lung cancer (NSCLC) by using tumour markers needs to be improved and standardised in order to compare marker profiles from different centres. A centre-independent tool based on receiver operating characteristics (ROC) curves instead of cut-off-based approaches for NSCLC diagnosis was established. Carcinoembryonic antigen (CEA) and cytokeratin-19 fragments (CYFRA 21-1) were measured in 326 NSCLC patients and 160 patients with benign lung diseases (Heidelberg, HD) and compared to 158 NSCLC patients and 128 controls from an occupational medicine high-risk cohort (Giessen, GI). The cohorts differed in tumour-stages, marker cut-offs and therefore in sensitivity for NSCLC detection. Sensitivity for CYFRA 21-1 (most sensitive marker) was 65% GI, 35% HD, for CEA: 43% GI and 35% HD. Marker-combination increased sensitivity to 53% HD resp. 73% GI, accompanied by decreasing specificity. A transfer of the cut-off-based classification methods from HD to GI and vice versa led to false classifications. Sensitivity and specificity do not change using classification methods on transformed data such as the described decision guarantee. CEA/CYFRA-combination allows a classification method transferable despite structural differences of the cohorts. Only 0.8% of the datasets showed discordance in classification. The diagnosis of NSCLC based on ROC curves eliminate centre-specific differences. Classification methods lead to an improvement in NSCLC diagnosis.*

Correspondence to: Professor Dr. Joachim Schneider, Institut und Poliklinik für Arbeits- und Sozialmedizin der Justus-Liebig Universität, Aulweg 129/III, D-35385 Giessen, Germany. Tel: +49 06419941300, Fax: +49 06419941309, e-mail: Joachim.Schneider@arbmed.med.uni-giessen.de

Key Words: Lung cancer, CEA, CYFRA 21-1, marker panel, ROC-curves.

Tumour markers currently play only a minor role in the primary diagnosis of lung carcinomas due to their limited sensitivity and specificity, as well as to their low organ specificity. Thus, the diagnosis of malignancy using tumour marker profile evaluation remains a challenge (6).

In numerous studies (1-3) successful efforts have been made to improve these diagnostic parameters in lung cancer detection. The results, however, are always dependent on application-specific conditions such as composition of the study cohort and laboratory techniques. These markedly influence the desired high diagnostic accuracy. Classification methods are therefore rarely directly comparable and can only be applied with difficulty to other laboratories. Opportunities for the transfer of cut-off-based classification methods will be discussed in the context of data obtained from two centres, and an approach to improve the transferability will be presented.

Patients and Methods

The study group consisted of 486 individuals from the Thoraxklinik Heidelberg (HD), including 326 patients (262 males, 64 females; mean age 63.2 ± 9.2 years) with histologically confirmed lung cancer and 160 subjects with benign lung disease (121 males, 39 females; mean age 56.7 ± 14.2 years). They were compared to 286 individuals from the Institute and Outpatient Clinics of Occupational and Social Medicine at the University of Giessen (GI), including 158 patients with newly-diagnosed, histologically confirmed lung cancer (139 males, 19 females; mean age 65.6 ± 9.3 years). A control group of 128 patients (126 males, 2 females; mean age 64.5 ± 7.2 years) without tumours consisted of subjects with silicosis or asbestosis, chronic obstructive pulmonary diseases (COPD) or inflammatory lung diseases and healthy subjects who had been exposed to carcinogens and who were at high risk of lung cancer.

Histological classification of the primary lung tumour cases revealed that they were restricted to non-small cell lung cancer (NSCLC). The tumour stages of the NSCLC were separated into four groups according to the Unio Internationalis Contra Cancrum (UICC) recommendations (4). In the HD group 114 (44.2%) lung tumours were diagnosed as stage I, 50 (19.4%) as stage II, 69 (26.7%) as stage III and 25 (9.7%) as stage IV. In the GI group 30 (19.0%) NSCLC

were diagnosed as stage I, 22 (13.9%) as stage II, 56 (35.4%) as stage III and 50 (31.6%) as stage IV.

In HD almost two-thirds of the NSCLC patients were in a lower stage I or II (mostly resectable patients); in GI over two-thirds of the patients were classified as stage III or IV ($p < 0.001$; χ^2 -test). Exclusion criteria were small cell lung cancers, cancer therapy and relapse, pulmonary metastases of extrapulmonary tumours, mesothelioma, sarcoma and lymphoma.

In GI blood samples were centrifuged (1000×g, 5 min) within 120 min. Sera were kept frozen at -18°C until analysis. Carcinoembryonic antigen (CEA) and cytokeratin-19 fragments (CYFRA 21-1) were measured with an ES 600 ELISA analyzer (Roche, Mannheim, Germany). In HD pretherapeutical samples were centrifuged at 1500×g for 10 minutes and measured immediately using Elecsys 2010 Bioanalyzer (Roche, Mannheim, Germany).

Statistical analysis was performed as previously described by Bitterlich and Schneider (5). The sensitivity-adapted decision guarantee (DG) in lung cancer diagnosis should be determined. In brief: based on a reference study group, for a measured tumour marker value m , the specificity SP_m and sensitivity SE_m , which will produce a cut-off with this value m as threshold, is calculated. Knowing the sensitivity SE^* at 95% specificity in the cohort, DG is calculated as follows:

$$DG(m) = \begin{cases} 0.5 \cdot \left(1 - \frac{SE_m - SE^*}{1 - SE^*}\right) & \text{if } SE_m \geq SE^* \\ 0.5 \cdot \left(1 + \frac{SE^* - SE_m}{SE^*}\right) & \text{if } SE_m < SE^* \end{cases}$$

The normalisation using SE^* (or $1 - SE^*$) assures that the value of the DG lies between 0 and 1, and, predictably, for $SE_m = 100\%$, takes on a value of 0, while for $SE_m = 0\%$, a value of 1. DG and the cut-off-based evaluation are qualitatively equal, because the DG threshold 0.5 corresponds by definition to the cut-off value at 95% specificity. Hence a DG above 0.5 is classified as 'malignant'.

Results

Sensitivity and specificity of CEA and CYFRA 21-1. The cut-off points and the sensitivity data at a specificity of 95% of the controls (patients with benign disease of the same organ) are listed in Table I. There were minor deviations based on the data of the specific cohorts (HD, GI) compared to the manufacturer's recommendations. As expected, CYFRA 21-1 was shown to be the most sensitive marker with a sensitivity of 65.6% in the GI cohort and 35.0% in the HD cohort at a specificity of 94.4%.

The sensitivity and specificity for the 95th percentile levels of the control group were confirmed by receiver-operating characteristics (ROC) curves for both centers. Comparison of the ROC curves for the tumour marker CEA shows a good agreement for both centres (Figure 1). The differences between the centres were markedly higher for the ROC curve of CYFRA 21-1 (Figure 2).

Combination of the markers CEA and CYFRA 21-1 in detection of NSCLC. The sensitivity can be increased when, instead of a single marker analysis, the values of CEA or CYFRA 21-1 are evaluated according to the following

Table I. Sensitivity and specificity of CEA and CYFRA 21-1 in the cohorts of Heidelberg (HD) and Giessen (GI).

| Marker | Recommended cut-off (ng/ml) | Centre HD | | Centre GI | |
|--------|-----------------------------|-----------------|---------------------------|-----------------|---------------------------|
| | | Cut-off (ng/ml) | Sensitivity (specificity) | Cut-off (ng/ml) | Sensitivity (Specificity) |
| CEA | 5.0 | 6.1 | 35.0% (94.4%) | 5.1 | 43.0% (94.5%) |
| CYFRA | 3.3 | 2.9 | 35.0% (94.4%) | 2.6 | 65.6% (94.5%) |

principles: a sample is classified as positive when at least one value lies above the cut-off; if all values of the particular tumour markers are less than their respective cut-offs, then the sample is classified as negative. In Figure 3 a two-marker profile is applied. The two-dimensional areas where data point pairs are classified as benign or malignant are visualized.

Combination of CEA and CYFRA 21-1 increased sensitivity of NSCLC from 35.0% (for either single marker) to 53.1% in the HD cohort; for the GI cohort, the sensitivity increased from 65.6% (best single marker CYFRA 21-1) to 73.4% (combined markers). This was accompanied by a decrease in specificity to 88.8% in the HD cohort, while specificity in the GI cohort remained unaffected at 94.5%.

A decrease in specificity can be counteracted if the cut-off values are evaluated using a multiparametric method so that the specificity remains unchanged at 95% for the marker combination of CEA/CYFRA 21-1. This specificity should be identical in both cohorts due to the similarity of the classification results. This is the case in the HD cohort when the cut-off for CEA is raised from 6.1 to 6.48 ng/ml (107%) and, for CYFRA 21-1, from 2.9 to 3.48 ng/ml (120%); in the GI cohort, the respective cut-off increases are, for CEA, from 5.1 to 7.93 ng/ml (130%) and, for CYFRA 21-1, from 2.6 to 2.78 ng/ml (107%). Thus, at a specificity of 95% for both markers, a sensitivity of 47.9% for the HD cohort and 68.4% for the GI cohort can be achieved.

As expected, the factors necessary to obtain 95% specificity are different and centre specific. Therefore, optimisation of each multiparametric classifier is required. Applying the classifier based on GI data to the datasets from HD without any further adaptation, leads to a loss in specificity (90%) without an appreciable increase in sensitivity (54.6%). This leads to a false classification in 6% of all datasets. The same is true when the HD classification is applied to the GI data. The sensitivity of the marker combination reaches 50% accompanied by a decline in specificity to 93.1%.

Results of the ROC-based classification. Classifiers that are based only on fixed cut-offs describe the specificity selectively, e.g. at 95% specificity. ROC curves contain additional information on data structure (14). For every data point of a

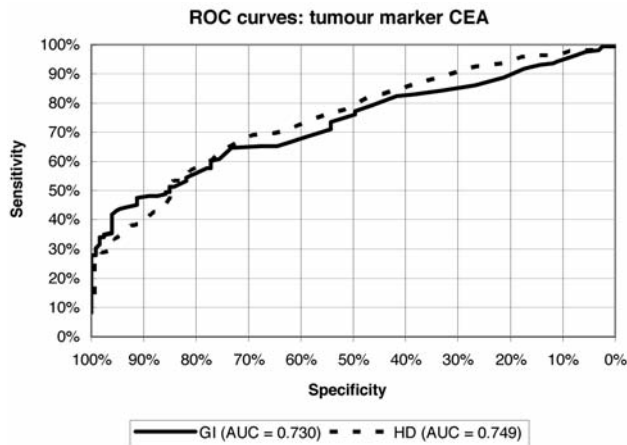


Figure 1. Receiver operating characteristic (ROC) curves comparing the performance of CEA in the HD patient group (326 NSCLC and 160 benign subjects) and in the GI patient group (158 NSCLC and 128 control subjects); AUC, area under curve.

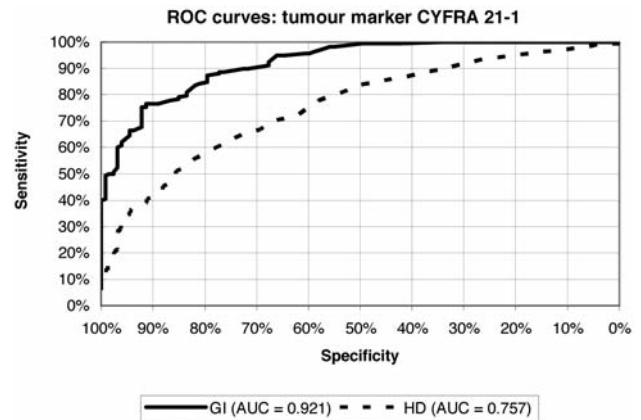


Figure 2. ROC curves comparing the performance of CYFRA 21-1 in the HD and in the GI cohorts (see Figure 1 legend for details).

tumour marker, it is possible to determine specificity and sensitivity from the diagnosis-based decision (DG) at this particular cut-off.

DG is estimated from a ROC curve for a particular dataset. For example, in HD a value of 2.9 ng/ml for CEA corresponds to 62.9% sensitivity at 75.6% specificity and may be assigned to a decision confidence of 28.5% (Figure 4). In the ROC curve from GI, the closest data point shows a sensitivity of 60.8% at 75.6% specificity. The decision confidence for this data point amounts to 34.8% with respect to GI. One can transform GI data to HD values. In order to calculate this transformation, it is not necessary to have knowledge about the datasets from GI – the availability of the ROC curve is sufficient. If one is familiar with the database, however, the method of comparison can even be applied to the data (5). The value of 2.9 ng/ml in HD corresponds to a value of 3.0 ng/ml in GI.

If the ROC curves are different, which is the case for the tumour marker CYFRA 21-1 (see Figure 2), the choice of the ‘closest data point’ can only allow an estimation of the comparability. Regarding specificity, it is advisable to use the data point on the ROC curve that corresponds to the closest point on the other ROC curve with the same specificity. For example, in HD 2.7 ng/ml for CYFRA 21-1, results in a sensitivity of 37.4% at 93.8% specificity (Figure 5). DG is 47.4%. In the GI dataset the corresponding DG is 47.2% at the same specificity (93.7%) and a sensitivity of 67.7%. Therefore the DG value is similar for HD and GI.

In agreement with the cut-offs, a DG value >0.5 ($>50\%$) will lead to a classification of the result as ‘malignant’. In a two-marker profile a sample is classified as positive if at least one DG value is greater than 0.5 ($>50\%$). The two-dimensional area in which data point pairs are classified as benign or malignant can be visualized (see Figure 6).

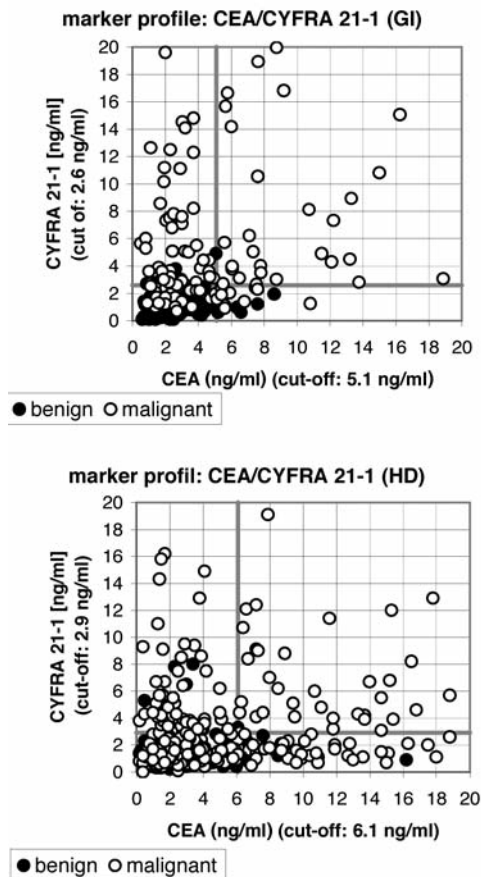


Figure 3. Combination of CEA and CYFRA 21-1 in the diagnosis of NSCLC. Results of the GI (above) and the HD (below) cohorts. The corresponding cut-off values are marked by grey lines.

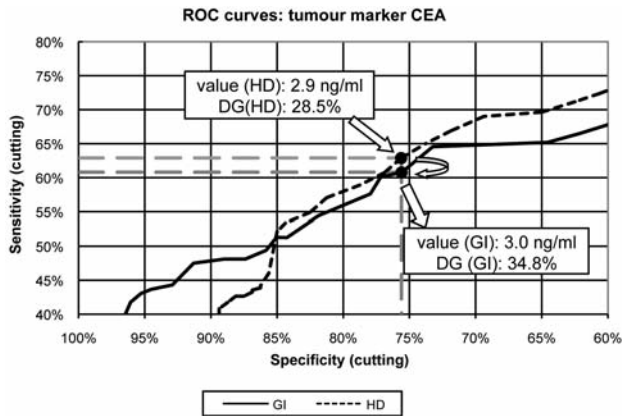


Figure 4. From ROC curves to DG. A CEA value of 2.9 ng/ml in the HD cohort (DG: 28.5%) corresponds to a CEA value of 3.0 ng/ml (DG: 34.8%) in the GI cohort at the same specificity of 75.6%.

The diagnostic parameters (sensitivity and specificity) do not change when the classification method is based on transformed data such as DG instead of the original data values. Multiparametric classification based on DG of the combination of CEA and CYFRA 21-1 leads to comparable sensitivities in the detection of NSCLC with 47.9% in the HD cohort and 68.4% in the GI cohort at a high specificity of at least 94.4%. Therefore, it is important to apply the classifier in GI to the DG of data values from HD that have been transformed in this manner. The quantitative comparison of the combination CEA/CYFRA permits the transfer of the classification method despite differences in the structure. The classification is for the most part still successful. Only 4 (0.8%) of the datasets from HD showed deviations in classification results. Thus, the transfer of multiparametric classifiers does not lead to a significantly different judgement regarding to malignancy.

Discussion

The analysis of tumour markers as diagnostic tools is generally based on the evaluation of individual measurements in relation to defined cut-off values (5). The definition of the cut-offs influences the diagnostic merit of this evaluation considerably. Variations in laboratory methods or reference population may lead to differing results. Therefore, data from different laboratories are often difficult to compare.

In this study, a centre-independent detection of malignancy was evaluated by means of classification with ROC-based data transformation in patients with NSCLC in comparison to benign lung disorders.

The sensitivities reported in the literature for CYFRA 21-1 or CEA are comparable with our results (overview 6, 7). In order to improve the diagnostic accuracy of marker tests for primary lung cancer, various combinations of tumour markers

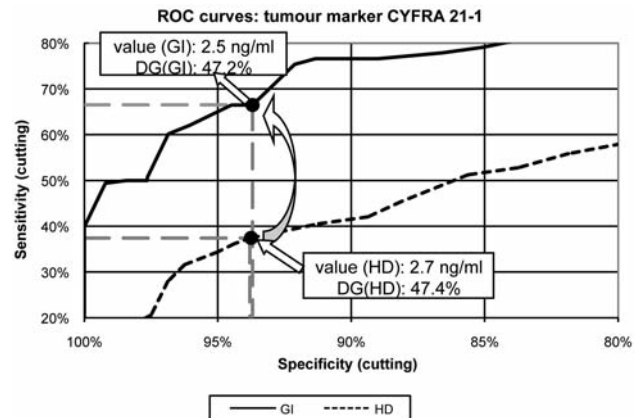


Figure 5. From ROC curves to DG. A CYFRA 21-1 value of 2.7 ng/ml in the HD cohort (DG: 47.4%) corresponds to a CYFRA 21-1 value of 2.5 ng/ml (DG: 47.2%) in the GI cohort at the same specificity of 93.7%.

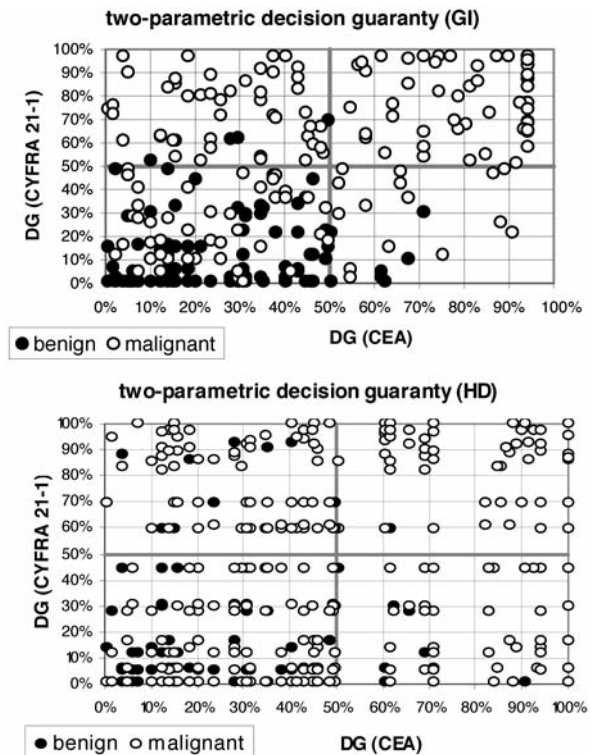


Figure 6. Two-parameter classification in the diagnosis of NSCLC based on DG. Results of the GI (above) and the HD (below) cohort. All DG above 0.5 (marked by lines) were classified as malignant.

have been proposed (3, 8, 9). Multiple marker panels proved to be more sensitive and than any single marker.

By combining CEA, squamous cell carcinoma antigen (SCC) and neuron-specific enolase (NSE) the sensitivity in

NSCLC could be raised to 65% (10). In other reports, tissue polypeptide-specific antigen and CYFRA 21-1 (11), or CEA and CYFRA 21-1 (10, 12), or CEA and cell adhesion molecule CAM 123-6 (13), showed the best performance. Combining CYFRA 21-1 (sensitivity: 57.7%) and CEA (sensitivity: 45.3%) increased sensitivity for NSCLC to a total of 75.4%. Unfortunately this was accompanied by a decrease in specificity, down to 86.5% (12).

In this study, the combination of CYFRA 21-1 and CEA increased sensitivity to 53.1% in the HD cohort consisting of more than 50% early tumour stages and 73.4% in the GI patients with more advanced tumour stages. However, the marker panel was accompanied by a reduced specificity of 88.8% for the HD data set.

A cut-off-independent diagnostic evaluation of tumour markers may avoid laboratory-based and method-derived systematic variation. DG is an appropriate parameter that is determined using a defined reference population and its respective ROC curve (5). Besides this, the relationship of a value to the cut-off permits not only a qualitative characterization of malignant or benign, but also describes the probability with which a diagnosis of NSCLC can be made. The distance between a data point and the cut-off point reflects, to some extent, the decision certainty or guarantee that accompanies this method of analysis (14).

The algorithms for the two-marker combination CEA and CYFRA 21-1 were optimised for the actual data at the high specificity of 95% either for the HD or GI cohorts, and the diagnostic performance of the method was tested by applying it to data of the other centre. Discordance in classification of malignancy was only seen in 0.8% of the HD data sets.

This study was able to demonstrate a method to analyse multi-parametric laboratory data, independent of the method for the diagnosis of NSCLC. The use of the sensitivity-adapted DG ensures that systematic differences in laboratory results influencing the cut-off value are eliminated. Any factors resulting from the specific recruitment of the patient collective influencing the ROC curves should receive attention. For instance, differences in the ROC curves may be observed when the proportion of patients with high to low tumour stages is changed or when the composition of histological subtypes of NSCLC is varied.

Multiparametric classification based on DG of the combination of CEA and CYFRA 21-1 led to comparable sensitivities for NSCLC in the HD cohort (47.9%) and the GI cohort (68.4%) at a high specificity of 94.4%. Therefore the use of the sensitivity-adapted DG opens new avenues for quality assurance.

A diagnosis of malignancy that is based on ROC curves can eliminate laboratory, technical and structural differences among centres. Thus, we recommend users to be provided with ROC curves. With a systematic feedback between laboratory results and the specialist physician's comprehensive clinical findings,

the sensitivity and specificity of the laboratory parameter can continuously be monitored.

References

- Schneider J, Bitterlich N, Velcovsky HG, Morr H, Katz N and Eigenbrodt E: Fuzzy logic-based tumor-marker profiles improved sensitivity in the diagnosis of lung cancer. *Int J Clin Oncol* 7: 145-151, 2002.
- Schneider J, Peltri G, Bitterlich N, Neu K, Velcovsky HG, Morr H, Katz N and Eigenbrodt E: Fuzzy logic-based tumor marker-profiles, including a new marker Tumor M2-PK improved sensitivity to the detection of progression in lung cancer patients. *Anticancer Res* 23: 899-906, 2003.
- Plebani M, Basso D, Navaglia F, De Paoli M, Tommasini A and Cipriani A: Clinical evaluation of seven tumour markers in lung cancer diagnosis: can any combination improve the results? *Br J Cancer* 72: 170-173, 1995.
- Unio Internationalis Contra Cancrum (UICC): TNM/L-Klassifikation maligner Tumoren. 5. Aufl. Deutsche Übersetzung (Hrsg.: Witteind Ch, Wagner G). Springer, Berlin, New York, 1997.
- Bitterlich N and Schneider J: Decision guarantee in tumour marker analysis: a cut-off independent assessment. *Anticancer Res* 27: 1933-1940, 2007.
- Schneider J: Tumor markers in detection of lung cancer. *Advances Clin Chem* 42: 1-41, 2006.
- Ebert W and Muley T: Tumor-Marker – Bedeutung für Diagnostik und Verlaufskontrolle. In: Drings, Dienemann, Wannenmacher (eds.). *Management des Lungenkarzinoms*. Springer Verlag Berlin Heidelberg New York pp. 135-146, 2003.
- Lombardi C, Tassi GF, Pizziccolo G and Donato F: Clinical significance of a multiple biomarker assay in patients with lung cancer. A study with logistic regression analysis. *Chest* 97: 639-644, 1990.
- Kakari S, Stringou E, Toubis M, Ferderigos AS, Poulaki E, Chondros K, Dema A, Kotsoyoulou V and Pavlidis N: Five tumor markers in lung cancer: significance of total and "lipid"-bound sialic acid. *Anticancer Res* 11: 2107-2110, 1991.
- Moro D, Villemain D, Vuillez JP, Delord CA and Brambilla C: CEA, CYFRA 21-1 and SCC in non-small cell lung cancer. *Lung Cancer* 13: 169-176, 1995.
- Giovanella L, Ceriani L, Bandera M, Rimoldi R, Beghe B and Roncari G: Tissue polypeptide specific antigen (tps) and cytokeratin 19 fragment (CYFRA 21-1) immunoradiometric assay in non-small cell lung cancer evaluation. *Q J Nucl Med* 39: 285-289, 1995.
- Koga H, Eguchi K, Shinkai T, Tamura T, Ohe Y, Oshita F, Saijo N, Kondo H, Oki K and Okura H: Preliminary evaluation of the new tumour marker CYFRA 21-1 in lung cancer patients. *Jpn J Clin Oncol* 24: 263-268, 1996.
- Kohno N, Hirasawa Y, Kondo K, Fujioka S, Fujino S, Abe M, Yokoyama A, Hiwada K, Watanabe K and Nishimura J: Comparative studies of CAM 123-6 and carcinoembryonic antigen for the serological detection of pulmonary adenocarcinoma. *Cancer Detect Prev* 21: 124-128, 1997.
- Bitterlich N and Schneider J: Cut-off-independent tumour marker evaluation using roc approximation. *Anticancer Res* 27: 4305-4310, 2007.

Received August 19, 2009

Revised March 12, 2010

Accepted March 15, 2010