

MULTIPLE REGRESSION PREDICTING LUNG CANCER BASED ON RISK FACTORS – A CASE STUDY FOR THE INDUSTRY

DUMITRU MIHAELA¹, PRAISLER MIRELA^{*1,2}, REBEGEA LAURA¹,
FIRESCU DOREL³

¹"Sf. Ap. Andrei" Emergency Clinical Hospital, Galati, Department of Radiotherapy,
Brailei St.177, 800830 Galati, Romania.

²"Dunarea de Jos" University of Galati, Department of Chemistry, Physics and
Environment, Domneasca St. 47, 800008 Galati, Romania.

³"Sf. Ap. Andrei" Emergency Clinical Hospital, Galati, Surgery Clinic II, Brailei St. 177,
800830 Galati Romania

Abstract: The objective of this study is to determine a multivariate model able to perform the identification of the tumor stage and of the histopathological type for lung cancer patients, based on predictive environmental and behavioral factors. The database was built by using clinical and personal information about 106 patients with stage III or IV lung cancer and a mean age of 58 years, who were subjected to radiotherapy at the Radiotherapy Department of „Sf. Ap. Andrei” Emergency Clinical Hospital between January and December 2010. The following factors were taken into consideration: work conditions, smoking habits and duration of the exposure to risk factors. Most of the patients (43.33%) have worked in metallurgy for a mean of 27.54 years and 86.67% have smoked for a mean period of 26.92 years. The duration of smoking and the habit of smoking itself were identified by multiple linear regression as the most important predictive factors for tumor stage and for the histopathological type, respectively. Artificial intelligence techniques were used and the results have indicated that tobacco use and the environmental factors related to the work place (e.g. metallurgical industry) are the best predictive risk factors for the incidence of lung cancer.

Keywords: lung cancer, risk factors, artificial intelligence system, metallurgical industry

1. INTRODUCTION

Lung cancer is a public health problem in world, being on first place as frequency in Romania and in entire world, representing the main cause of death for men. Unfortunately, lung cancer is usually diagnosed in Romania in advanced stages, when the medical treatment is no longer very efficient. The great majority of epidemiological studies show that lung cancer incidence increased in time in the same manner for all age groups, regardless of the gender type. The aim of this study was to detect which of the know risk factors for lung cancer are the most predictive in the particular conditions in the region served by the „Sf. Ap. Andrei” Emergency Clinical Hospital, Galati. The main industries in this region belong to metallurgy and metal processing industries (e.g. shipyard).

*Corresponding author, e-mail: Mirela.Praisler@ugal.ro

The main relevant factors for lung cancer to be studied have been identified by a literature research: as the factors contributing to the development of cancer are so diverse, correlations are usually determined by processing data about very large populations of patients. The literature research has indicated that smoking is the greatest risk factor for lung cancer, this habit being the most probable cause for the great majority of lung cancers for both men and women [1]. A smoker would die from lung cancer with a probability 15 times higher than a non-smoking patient [2]. The duration of smoking and the number of cigarettes smoked per day are the main predicting risk indicators for developing lung cancer [2-4]. Another risk related to lung damage is starting smoking at a young age [5]. The environment in which the patient has been working is also an important factor influencing the chances of developing lung cancer. The literature mentions among the most important carcinogen agents from industry compounds such as: asbestos, arsenic, tar, chromium, nickel, and polycyclic hydrocarbons. The production of non-ferrous metal and painting are also usually toxic environments, which have been associated to lung cancer [6 - 9]. Others documented risk factors linked to the occupational exposure are diesel exhaust, [10, 11], the exposure to herbicides, insecticides, silica or asbestos [12-15].

2. EXPERIMENTAL RESULTS

Our study involved 106 patients with lung cancer with a mean age of 58 years old (range 49 - 80 years). The group consisted of 81 men (76.42% of cases) and 25 women (23.58% of cases). The following predictive factors were analyzed for each patient: smoking, work environment, and the duration of exposure at risk factors, according to the tumor stage and to the histopathological type in lung cancer cases. As much as 86.67% of patients were smokers with a mean duration of the smoking habit of 26.92 years (15 - 45 years). Regarding occupational exposure, 43.33% of patients worked in metallurgy, 23.33% patients in agriculture, 10% in building construction, 6.67% of in ship building, and 6.67% in furniture industry; 10% of the patients have had no occupational exposure. The patients were diagnosed with lung cancer based on histopathological and cytological investigations. Within the patient group 40% were in stage III and 60% in stage IV of lung cancer.

Analyzing the histopathological types, we have found a predominance of epidermoid carcinoma at 50% of patients (all smokers), 13.33% had adenocarcinoma, and 23.33% of patients didn't have histopathological confirmation, neoplasia being put in evidence by thoracic CT scan or bronchoscopy associated with positive cytology. The comorbid illnesses were obstructive chronic bronchopneumopathy for 63.33% of all the patients and 73.07% of the smoking patients. Chronic ischemic cardiopathy was found for 23.33% of all patients, myocardial infarction 6.67% and diabetes mellitus 6.67% of all cases. The treatment was performed with radiotherapy equipment Theratron Elite 100, in isocentric technique, with 2 fields. The fractionation schedules used for patients included in our study were 50Gy in 25 fractions/5 weeks for radiotherapy with curative intent. The fractionation schedules palliative intent, for local symptoms relief, thoracic pain syndrome and superior cava vena syndrome were 30Gy in 10fractions/10 days.

Our objective was to find the main predictive factors for the disease stage and for the histopathological type. The independent factors that were studied were the occupational exposure, the duration of exposure, the smoking habit and the duration of the smoking habit.

The data was processed by using the STATISTISXL version of SPSS software. Table 1 shows that the R squared and the adjusted R squared of the regression results for predicting the tumor stage (0.473 and 0.389) shows a medium/week correlation between the predictive factors and the tumor stage. However, we must take into account that the values of these indices are lowered by the risk factors included in the analysis and which were found to be predictive, but without statistical significance, i.e. the type of occupational exposure ($p = 0.079$) and smoking ($p = 0.452$).

The results of the multiple linear regression analysis have shown that the main predictive factors for the tumor stage are related to the time during which the patient has been exposed to risk factors: the period of time that the patient has been a smoker (and not the smoking itself, see Figure 1), and the period of time in which the patient has worked under the influence of environmental risk factors such as coal gasification, asbestos, polycyclic hydrocarbons (see Table 1).

The results presented in Table 2 and Figure 2 show that only one predictive factor was found for the histopathological type and this is the habit of smoking itself (and not the duration of smoking).

Table 1. Linear regression results for predicting the tumor stage.

Factor	Coef	Std Error	t stat	p Value
Constant	3.238	0.229	14.135	0.000
Occupational exposure	0.084	0.046	1.831	0.079
Work duration in risk medium (years)	-0.021	0.009	-2.378	0.025
Smoke	0.283	0.370	0.763	0.452
Duration of smoking	0.02	0.01	2.31	0.03
R Squared	0.473		Std Error	0.389
Adj R Sq	0.389		SS reg	3.408
F Value	5.616		SS resid	3.792

The type of occupational exposure and the duration of working in an unhealthy environment were also not found to be predictive factors for the histopathological type of the tumor. As a result, the same situation as in the case of the models predicting the tumor stage is encountered for the level of the R squared and the adjusted R squared obtained for the linear regression predicting the histopathological type of the tumor. The values of these indices (0.224 and 0.052) show a weak correlation between the predictive factors and the histopathological type of the tumor. However, the habit of smoking itself has been proved to be an important predictive factor for the histopathological type of the tumor.

Table 2. Linear regression results for predicting the histopathological type of the tumor.

Factor	Coef.	Std Error	t stat	p Value
Constant	6.465	0.566	11.414	0.000
Occupational exposure	0.041	0.108	0.380	0.71
Work duration in risk medium (years)	0.002	0.021	0.095	0.92
Smoke	-1.829	0.892	-2.052	0.05
Duration of smoking	0.025	0.02	1.23	0.23
R Squared	0.224		Std Error	0.77
Adj R Sq	0.052		SS reg	3.08
F Value	1.301		SS resid	10.66

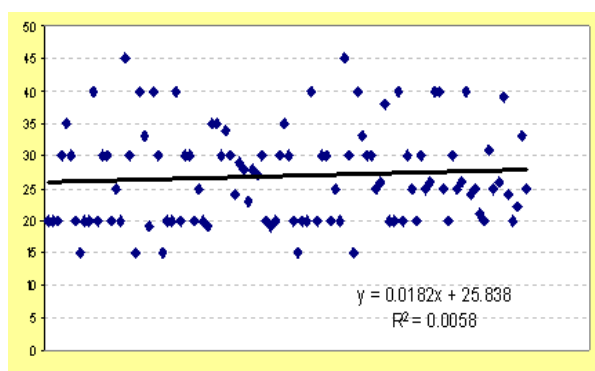


Fig. 1. Scatterplot illustrating the linear correlation between smoking and the tumor stage.

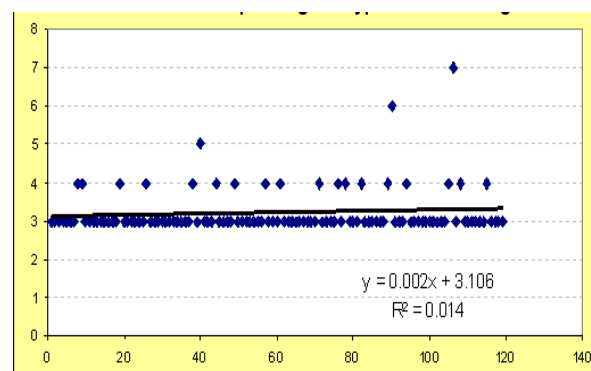


Fig. 2. Scatterplot illustrating the linear correlation between smoking and the histopathological type of the tumor.

3. CONCLUSIONS

The correlations between the main predictive factors presented in the scientific literature and lung cancer pathology were confirmed by our study, and the specific intensity of their effect has been evaluated for the particular conditions faced by the inhabitants of the South-East region of Romania. We have determined that the duration of smoking is a strong predictive factor for lung tumors evolving from early to advanced stage ($p < 0.05$). The duration of work with risk exposure, e.g. in metallurgy, has also been found to be an important predictive factor for the tumor stage for the patients of the studied geographical area. In addition, we have identified which factors are the best predictors of the tumor stage and of the histopathological type of the tumor. The correlations are especially true for the employees of the metallurgical plants and associated industries, in which the occupational exposure to carcinogen factors such as coal gasification and asbestos increase the risk of developing lung cancer, especially if the patient is a smoker.

REFERENCES

- [1] Peto, R., Lopez, A.D., Boreham, J., Thun, M., Heath, Jr. C., Mortality from Smoking in Developed Countries 1950-2000. Oxford University Press, Oxford, 1994.
- [2] Doll, R., Peto, R., Boreham, J., Sutherland I., Mortality from cancer in relation to smoking: 50 years observations on British doctors. *Br J Cancer*, vol. 92, no. 3, 2005, p. 426-429.
- [3] Lubin, J.H., Alavanja, M.C.R., Caporaso, N., Brown, L.M., Brownson, R.C., Field, R.W., Montserrat, G.C., Hartge, P., Hauptmann, M., Cigarette smoking and cancer risk: modeling total exposure and intensity *Am J Epidemiol*, vol. 166, no. 4, 2007, p. 479-489.
- [4] Lubin, J.H., Caporaso, N.E., Cigarette smoking and lung cancer: modeling total exposure and intensity *Cancer Epidemiol Biomarkers Prev*, vol. 15, nr. 3, 2000, p. 517-523.
- [5] Wiencke, J.K., Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. *J Natl Cancer Inst*. vol. 91, no. 7, 1999, p. 614 - 619.
- [6] Alberg, A.J., Samet J.M., Epidemiology of lung cancer. *Chest*, vol. 123 (1 Suppl), 2003, p. 21S-49S.
- [7] Armstrong, B., Hutchinson, E., Unwin, J., Fletcher, T., Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: a review and meta-analysis. *Environ Health Perspect*, vol. 112, no. 9, 2004, p. 970-978.
- [8] Richiardi, L., Boffetta P., Simonato L., Forastiere F., Zambon P., Fortes C., Occupational risk factors for lung cancer in men and women: a population-based case-control study in Italy *Cancer Causes Control*, vol 15, no. 3, 2004, p. 285-294.
- [9] Zeka, A., Mannetje, A., Zaridze, D., Szeszenia-Dabrowska, N., R., Peter, L. Jolanta, Fabiánová, E., Lung cancer and occupation in nonsmokers: a multicenter case-control study in Europe. *Epidemiology*, vol. 17, issue 6, 2006, p. 615-623.
- [10] Lipsett, M., Campleman, S., Occupational exposure to diesel exhaust and lung cancer: a meta-analysis. *Am J Public Health*, vol 89 no. 7, 1999, p. 1009-1017.
- [11] Bhatia, R., Lopipero, P., Smith, A.H., Diesel exhaust exposure and lung cancer. *Epidemiology*, vol. 9, no. 1, 1998, p. 84-91.
- [12] Alavanja, M. C., Dosemeci, M., Samanic, C., Lubin, J., Lynch, C. F., Knott, C., Barker, J., Hoppin, J. A., Sandler, D. P., Coble, J., Thomas, K., Blair, A. (2004) Pesticides and lung cancer risk in the agricultural health study cohort, *Am. J. Epidemiol.*, vol.160, 2004, p. 876-885.
- [13] Purdue, M.P., Hoppin, J.A., Blair, A., Dosemeci, M., Alavanja, M.C., Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. *Int J Cancer*, vol. 120, no. 3, 2007, p. 642-649.
- [14] Pelucchi, C., Pira, E., Piolatto, G., Coggiola, M., Carta, P., La Vecchia, C., Occupational silica exposure and lung cancer risk: a review of epidemiological studies 1996-2005. *Ann Oncol*, vol. 17, no. 7, 2006, p. 1039-1050.
- [15] Lee, P.N., Relation between exposure to asbestos and smoking jointly and the risk of lung cancer. *Occup Environ Med*, vol. 58, no. 3, 2001, p. 145-153.