RISK ASSESSMENT PRACTICE IN HYGIENIC AND EPIDEMIOLOGICAL STUDIES

UDC 614.7
DOI: 10.21668/health.risk/2017.1.06.eng

MULTIVARIATE ANALYSIS OF STRUCTURE AND CONTRIBUTION PER SHARES MADE BY POTENTIAL RISK FACTORS AT MALIGNANT NEOPLASMS IN TRACHEA, BRONCHIAL TUBES AND LUNG

G.T. Aydinov [1,2], B.I. Marchenko [1,3], Yu.A. Sinelnikova[1]

[1]Center of Hygiene and Epidemiology in the Rostov region, 67 7-th Line Str., Rostov-on-Don, 344019, Russian Federation
[2]Rostov State Medical University, 29 Nakhichevanskiy pereulok, Rostov-on-Don, 344022, Russian Federation
[3]Southern Federal University, 105/42 Bolshaya Sadovaya Str., Rostov-on-Don, 344006, Russian Federation

The article gives the results of multivariate analysis of structure and contribution per shares made by potential risk factors at malignant neoplasms in trachea, bronchial tubes and lung. The authors used specialized databases comprising personified records on oncologic diseases in Taganrog, Rostov region, over 1986-2015 (30,684 registered cases of malignant neoplasms, including 3,480 cases of trachea cancer, bronchial tubes cancer, and lung cancer). When carrying out analytical research we applied both multivariate statistical techniques (factor analysis and hierarchical cluster correlation analysis) and conventional techniques of epidemiologic analysis including etiologic fraction calculation (EF), as well as an original technique of assessing actual (epidemiologic) risk. Average long-term morbidity with trachea, bronchial tubes and lung cancer over 2011-2015 amounts to 46.64 $^{o}/_{oooo}$. Over the last 15 years a stable decreasing trend has formed, annual average growth being –1.22 %. This localization holds the 3rd rank place in oncologic morbidity structure, its specific weight being 10.02 %. We determined etiological fraction (EF) for smoking as a priority risk factor causing trachea, bronchial tubes and lung cancer; this fraction amounts to 76.19 % for people aged 40 and older, and to 81.99 % for those aged 60 and older. Application of multivariate statistical techniques (factor analysis and cluster correlation analysis) in this research enabled us to make factor structure more simple; namely, to highlight, interpret, give a quantitative estimate of self-descriptiveness and rank four group (latent) potential risk factors causing lung cancer.

Key words: *social and hygienic monitoring, risk assessment, risk factors, malignant neoplasms, carcinogenic risk, factor analysis, hierarchical cluster correlation analysis*

Currently, the priority direction in addressing problems associated with innovative methodological database for hygienic research is to improve national methods for risk assessment and analysis. This is a modern toolkit that significantly expands analytical and prognostic capabilities of researchers, including elements of situational and simulation modeling [6, 8, 9]. The methodology for risk assessment and analysis harmoniously complements conventional techniques of epidemiologic analysis, adapted for socio-hygienic monitoring at regional and municipal levels. Thus, in the Rostov region, when studying cause-and-effect relationships

Gennadiy T. Aydinov – Doctor of Medical Sciences, Professor, Corresponding member of the International Academy of Science, ecology, safety of human and nature, Chief Doctor, Head of the Hygiene Department (e-mail: master@donses.ru; tel.: +7 (863) 251-04-92).

Boris I. Marchenko – Doctor of Medical Sciences, professor, research consultant (e-mail: hie_sfu@mail.ru; tel.: +7 (863) 464-29-62).

Juliya A. Sinel'nikova – Head of the Social and Epidemiological Monitoring Department (e-mail: sgm@donses.ru; tel.: +7 (863) 251-04-74).

in the "environment - population health" system, parallel with the categorized data, presented in formal state and industry statistical reports, the specialized databases are in use. They are formed based on personified records of malignant tumors events, which conditions using the techniques of multivariate statistical analysis [10].

A wide spectrum of heterogeneous carcinogenic risk factors of genetic (hereditary), environmental (carcinogens, chemical carcinogenesis modifiers and others), occupational-production and of individual nature is included in the number of potential causes of cancer development. Currently, active and passive tobacco smoking is considered as a priority risk factor for lung cancer (LC). Its role is assessed as more significant than such an individual spectrum risk factor, as alcohol abuse [2, 4, 14]. Thus, according to epidemiological studies, the etiological fraction (EF) of tobacco smoking in the cause structure of malignant tumors of trachea, bronchial tubes and lung for men varies within the range of 85-95%. For women, this indicator is somewhat lower, and makes 65-80% [2, 16, 17, 19]. It is shown that in men who smoke more than 30 cigarettes a day, the relative risk (RR) of lung cancer is 4.3-4.5 times higher than that of smokers with fewer than 10 cigarettes [15]. At the same time, data obtained in a number of studies show that risk of developing lung cancer depends more on the length of employment than on the intensity of tobacco smoking [4]. The combination of active and passive smoking with the effects of hereditary background that potentiates its effect is considered as an individual oncologic predisposition to lung cancer [1]. According to a number of authors, both overweight, and low physical activity, and lack of fruits and vegetables - the main source of antioxidants, are also referred to significant individual risk factors, along with tobacco smoking and alcohol abuse [13, 14]. Among the risk factors of the population level, an important role is played not only by anthropogenic air pollution with carcinogenic xenobiotics, but also by the effects of ionizing radiation, primarily due to radon. Thus, according to meta-analysis results, about 10% of all lung cancer deaths are caused by radon, more than 30% of deaths are among non-smokers [5, 11, 12, 18, 20].

From our point of view, taking into account the hypothesis on multi-casual nature of malignant neoplasms (MNs), a compulsory component of risk factor analysis within the framework of socio-hygienic monitoring is studying their structure. It is important the quantitative characteristics and qualitative interpretation of the most typical combinations: group (latent, larvate) risk factors, which combine heterogeneous, but closely correlated together, initial parameters [7, 10].

The purpose of the study was to analyze the structure and the contribution of potential factors of carcinogenic risk to priority localizations and forms of malignant neoplasms.

Materials and methods. Analytical studies were carried out following a specialized database of personified records on oncologic pathology in the city of Taganrog, the Rostov region, with a population of about 255,000 people. During the period covered (1986-2015), 30,684 cases of malignant tumors were recorded. We’ve applied two multidimensional mathematical-statistic methods: factor analysis and cluster hierarchical correlations analysis [7, 10]. Correlation matrices in factor analysis were calculated based upon the information of 107 different potential risk factors and other significant parameters registered in special "Investigation charts on the case of newly diagnosed malignant neoplasm". The group (latent) factors were taken by principal components method. To determine the number of group (latent) factors, we used Kaiser and Cattell criteria. Factor rotation was carried out according to Varimax. The initial data applicability for factor analysis was evaluated by Kaiser-Meyer-Olkin measure of sampling and Barlett sphericity. Each of the group (latent) factors taken combines the closely correlated together original (recorded in the primary database) risk factors and the relatively high values of factor loads. The significance (contribution per shares) of individual initial potential risk factors within each taken group (latent) factor based on factor analysis, was estimated from the calculated values of their factor loads; which, in turn, quantitatively characterize the level of relationship between the initial and the group (latent) factors. The initial potential risk factors were regarded as highly informative and significant at the values of the corresponding factor loads of 0.500 and more. A qualitative interpretation of the taken group (latent) factors consisted in the semantic identification thereof through initial potential risk factors. Applying the method of cluster hierarchical correlations analysis made it possible to classify initial potential risk factors with their grouping into hierarchically organized clusters and to represent the results graphically in visual dendrograms. The oncologic morbidity analysis was carried out using specialized software package "Turbo oncologist", version 2.01. It ensures databases creation of both the categorical form, upon statistical reports, as well as personified databases. It implements algorithms of epidemiological analysis of intensity (level), structure, dynamics and spatial characteristic, as well as the original method of assessing actual (epidemiological) risk [3]. When carrying out the factor analysis and hierarchical cluster analysis of correlations, a professional software package “Statistical Package for Social Science” (SPSS), version 13.0 was used [7].

Results and discussion. In doing this research, we went on with studying the structure and contribution per shares of potential risk factors for malignant neoplasms of priority forms and process localizations [10]. We used data on the lung cancer incidence, the long-term average annual occurrence of which among the population of Taganrog for the period 2001-2015 made 46.64 ± 3.26 o/oooo. And over the past 15 years, there has been a steady decline in the disease incidence with an average annual growth rate of 1.35%. The long-term morbidity dynamics model, described by the exponential curve with the equation: $\mathbf{P(X) = 45.101 \cdot 0.987^{X}}$, where **P(X)** is the incidence rate for the year with the ordinal number X, was statistically valid ($p<0.05$). It was used to calculate the mid-term extrapolation prognosis for 2016 and 2017: 36.30 ±3.71o/oooo, and 35.81 ± 3.71o/oooo, respectively (Fig. 1).

In the structure of total cancer incidence, this localization of malignant neoplasms over the past 15 years stably ranks third with a specific weight of 10.02%. The epidemiological risk assessment done, taking into account regional criteria, based on the calculation of the standardized background risk for the population of regional subordination cities over a fifteen-year period (27.72 o/ooo), allows MN of trachea, bronchial tubes and lung to considered as the priority for the population of Taganrog. For these neoplasms, an increased actual risk level was diagnosed with its individual norm-based score (Wi) equal to 1.069 (fourth ranked among the cities of the Rostov region).

According to an estimate of the main potential risk factors prevalence for the period of 1986-2015, the percentage of smokers among 3,480 patients with MN of trachea, bronchus and lungs is 82.18% (with a city-average of 38.14%). According to the data for the last thirty years, the etiological fraction

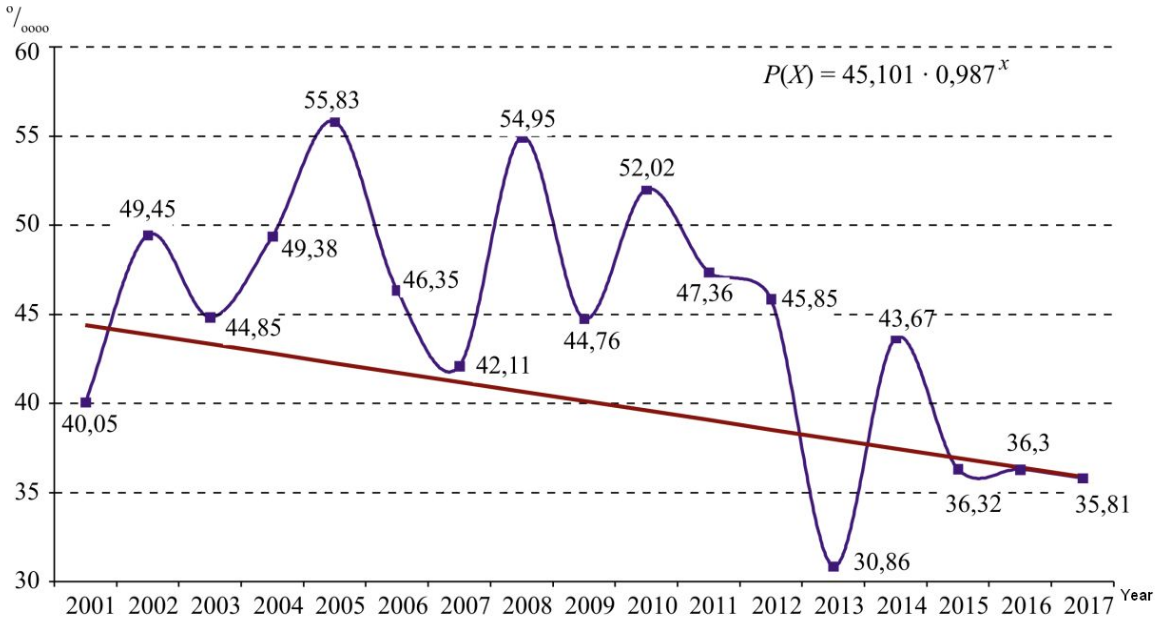

Figure 1. Dynamics of trachea, bronchial tubes and lung MNs incidence in the city of Taganrog over 1986-2015 and medium-term extrapolation prognosis for 2016 and 2017

(EF) of smoking in the occurrence of malignant neoplasms of the given localization makes 76.19% for the residents of Taganrog at the age of 40 years and older, 81.99% for the people aged 60 years and over. 36.38% of patients with lung cancer have chronic respiratory diseases, 13.31% - malignant neoplasms of similar localization among blood relatives, 27.97% have specific adverse occupational factors, 23.97% - heterogeneous risk factors in everyday life, including passive smoking.

For eight iterations, we obtained the results of factor analysis fulfilled. It was possible to determine and interpret meaningfully the four group (latent) factors, explaining total variance of 86.525%.

The group factor was ranked first with a variance share of 37.082%, which combines four primary registered potential risk factors. The highest factor load (0.982) in the structure of the first group (latent) factor accounts for active smoking. It should be noted that only cases of habitual and partisan tobacco smoking of 5 years smoking period or longer were taken into account. The second rank place among the initial potential risk factors belongs to chronic respiratory diseases. The corresponding factor load thereto is 0.885. The factor of alcohol abuse, with a factor load of 0.678 ranks third.

The fourth: predominant strong alcoholic beverages drinking (0.601). Thus, the first group (latent) factor, taking into account the specifics of initial potential risk factors it includes, and interrelations between them, can be meaningfully interpreted as "individual habitual intoxications and concomitant chronic diseases of respiratory system".

A group (latent) factor of the second rank with a variance share of 19.003% combines three initial potential risk factors, which on the whole should be interpreted as "unfavorable parameters of anthropogenic load and industrial-occupational environment." The first two ranks by the values of factorial loads are taken by the initial risk factors: 1) living in residential areas with relatively high levels of anthropogenic impact on atmospheric air due to motor transport emissions (0.777); 2) living in a zone of intensive emissions effect from stationary sources of industrial

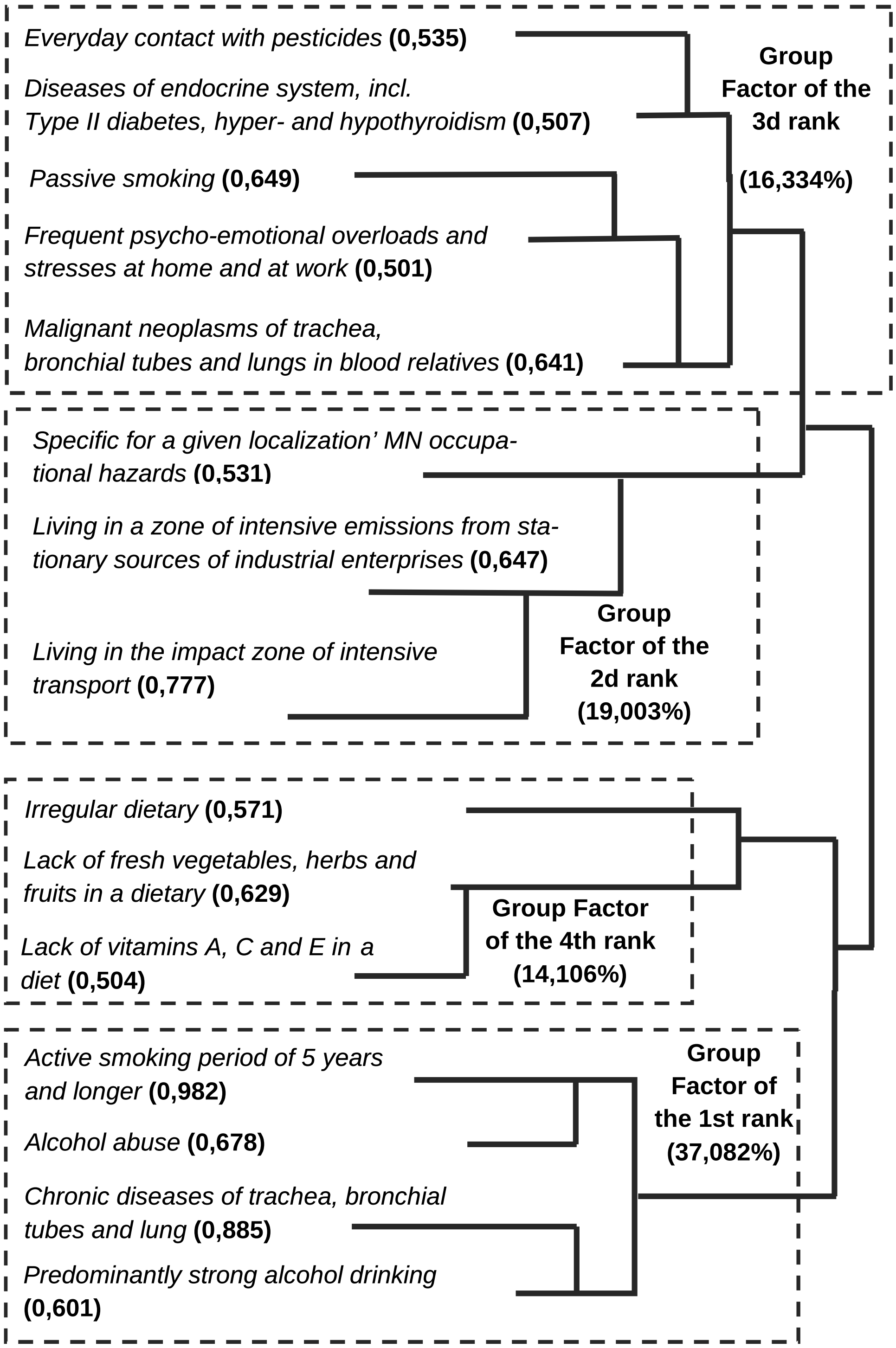

Figure 2. Dendrogram of the factors structure at MN of trachea, bronchus, and lung in the city of Taganrog for the period of 1986-2015.

enterprises (0.647). The third initial potential risk factor is a specific occupational hazard (0.531).

This refers to the contact mentioned in the patient's history with various occupational hazards specific for a given malignancies localization: chromium, nickel, arsenic and their compounds; silica, soot, asbestos, benzene, toluene, wood dust, ionizing radiation and others.

A group factor of the third rank with a 16.334% variance share is interpreted as "hereditary predisposition and potential risk factors of the individual spectrum". It combines five initial factors: 1) MN of trachea, bronchus and lung in blood relatives (factor load 0.641); 2) frequent psycho-emotional overloads and stresses in family and at work (0.501); 3) passive smoking (0.649); 4) diseases of endocrine system, including type II diabetes, hyperthyroidism, hypothyroidism (0.507); 5) everyday life contact with pesticides (0.535).

A group factor of the fourth rank with a variance share of 14.106% is interpreted as "individual features of food ration and diet". In this group, three initial risk factors for the individual spectrum were identified: 1) lack of fresh vegetables herbs and fruits in a dietary (factor load 0.629); 2) irregular dietary (0.571); 3) a lack of foods high in vitamins A, C and E in a diet (0.504). According to hierarchical cluster analysis of correlations, it’s been established that in the structure of the first rank group factor, the most closely pair-wise interact: a) active habitual and partisan smoking of 5 years period and longer, and alcohol abuse; b) chronic diseases of trachea, bronchial tubes, lung, and predominantly strong alcohol drinking (Figure 2).

Within the second rank group factor, the most closely related factors are residence in the areas of relatively high levels of air pollution with automobile transport and in a zone of emissions from stationary sources of industrial enterprises. In the structure of the third rank group factor ("hereditary predisposition and potential risk factors of the individual spectrum"), in the cluster correlations analysis, the relationships in the two groups of the initial factors were found. The first group includes contact with pesticides in everyday life and diseases of endocrine system, the second: passive smoking, frequent psycho-emotional overloads and stress at home and at work, as well as hereditary predisposition – malignant neoplasms of respiratory organs in blood relatives. In the group factor of the fourth rank ("individual features of dietary"), the commonality of the two primary factors is determined: a deficiency in fresh vegetables, herbs and fruits in combination with a deficit of vitamins A, C and E (Figure 2).

Conclusion. Thus, using factor analysis for the purpose of studying an optionally arranged system "population and individual factors of cancer risk - malignant neoplasms of trachea, bronchus and lung" makes it possible to determine and quantify its structural organization. This allows for assessment of practical value for optimizing management decisions and justifying the priority of preventive and health-improving measures being developed. Cluster analysis of correlations adds to the results of factor analysis in detecting the groups of closely related variables (of the primary potential risk factors) and provides their visual representation in the form of informative dendrograms.

Applying these multidimensional statistical techniques in this study made it possible to simplify the factors structure: to isolate, interpret, quantify the informational content (by the contribution share to the total variance). We’ve managed to range, according to the confidence level, and to study the hierarchical structure of four group (latent) potential risk factors for occurrence of trachea, bronchus and lungs MNs.

Based on the results of factorial and cluster analysis of correlations, the option of further optimization for the monitoring procedure is a significant reduction in the volume of primary information being recorded during transition from scientific research to actual practice. For this purpose, for each identified group (latent) risk factor, among the closely correlated together primarily recorded risk factors, a marker-factor is distinguished – of the highest factor load, which is subject to further registration at socio-hygienic monitoring.

References

1. Brenishter S.I. Geneticheskie faktory predraspolozhennosti k raku legkogo [Genetic factors to lung cancer predisposition]. *Medline.ru*, 2013, vol. 14, no. 2, pp. 362–389. Available at: www.medline.ru/public/pdf/14_032.pdf (18.09.2016) (in Russian).

2. Volkotrub L.P. Tabakokurenie kak faktor kantserogennogo riska [Tobacco smoking as a cancerogenic risk factor] *Zdravookhranenie Rossiiskoi Federatsii*, 2010, no. 2, pp. 10–15 (in Russian).

3. Egorova I.P., Marchenko B.I. Otsenka epidemiologicheskogo riska zdorov'yu na populyatsionnom urovne pri mediko-gigienicheskom ranzhirovanii territorii: posobie dlya vrachei [Evaluation of epidemiological risk to health on the population level at the medical and hygienic ranging of the areas: guide for physicians]. Utverzhdeno sekciej po gigiene uchenogo soveta Minzdrava RF 24.12.1999, protokol № 9. Moscow, 1999, 48 p. (in Russian).

4. Mikhailov E.A., Levshin V.F., Goryacheva A.N., Tsybulina L.P. Issledovanie anamneza kureniya u bol'nykh so zlokachestvennymi novoobrazovaniyami [Study of smoking history in cancer patients]. *Vestnik RONTs im. N.N. Blokhina RAMN*, 2009, vol. 20, no. 1, pp. 36–42 (in Russian).

5. Kononenko D.V. Otsenka radiatsionnogo riska dlya naseleniya Sankt-Peterburga pri obluchenii radonom [Risk assessment for the population of Saint-Petersburg from residential exposure to radon]. *Radiatsionnaya gigiena*, 2013, vol. 6, no. 1, pp. 31–37 (in Russian).

6. Zaitseva N.V., Popova A.Yu., May I.V., Shur P.Z. Metody i tekhnologii analiza riska zdorov'yu v sisteme gosudarstvennogo upravleniya pri obespechenii sanitarno-epidemiologicheskogo blagopoluchiya naseleniya [Methods and technologies of health risk analysis in the system of state management under assurance of the sanitation and epidemiological welfare of population]. *Gigiena i sanitariya*, 2015, vol. 94, no. 2, pp. 93–98 (in Russian).

7. Nasledov A.D. SPSS: Komp'yuternyi analiz dannykh v psikhologii i sotsial'nykh naukakh [SPSS: Data computer analysis in psychology and social sciences]. St. Petersburg, Piter Publ., 2005, 416 p. (in Russian).

8. Onishchenko G.G. Aktual'nye zadachi gigienicheskoi nauki i praktiki v sokhranenii zdorov'ya naseleniya [Actual problems of hygiene science and practice in the preservation of Public health]. *Gigiena i sanitariya*, 2015, vol. 94, no. 3, pp. 5–9 (in Russian).

9. Popova A.Yu. Strategicheskie prioritety Rossiiskoi Federatsii v oblasti ekologii s pozitsii sokhraneniya zdorov'ya natsii [Strategic priorities of the Russian Federation in the field of ecology from the position of preservation of health of the nation]. *Zdorov'e naseleniya i sreda obitaniya*, 2014, vol. 251, no. 2, pp. 4–7 (in Russian).

10. Aidinov G.T., Marchenko B.I., Sofyanikova L.V., Sinel'nikova Yu.A. Primenenie

mnogomernykh statisticheskikh metodov pri vypolnenii zadach sovershenstvovaniya informatsionno-analiticheskogo obespecheniya sistemy sotsial'no-gigienicheskogo monitoringa [The application of multidimensional statistical methods in the tasks of improving of information and analytical providing of the system socio-hygienic monitoring]. *Zdorov'e naseleniya i sreda obitaniya*, 2015, vol. 268, no. 7, pp. 4–8 (in Russian).

11. Risk vozniknoveniya raka legkogo pri obluchenii radonom i produktami ego raspada: zayavlenie po radonu [Risk of lung cancer evolvement under exposure to rays of radon and products of its decay: statement on radon]. In: M.V. Zhukovskogo, S.M. Kiseleva, A.T. Gubina eds. Moscow, FGBU GNTs FMBTs im. A.I. Burnazyana FMBA Rossii Publ., 2013, 92 p. (in Russian).

12. Krewski D., Lubin J.H., Zielinski J.M. [et al.]. A combined analysis of North American case-control studies of residential radon and lung cancer. *Journal Toxicology and Environ Health, Part A*, 2006, vol.69, no.7, pp. 533–597.

13. Alberg A.J., Brock M.V., Samet J.M. Epidemiology of lung cancer: looking to the future. *Journal Clin. Oncol*, 2005, vol. 23, pp. 3175–3185.

14. Alberg A.J., Ford J.G., Samet J.M. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*, 2007, no. 132, suppl. 3, pp. 29–55.

15. Brennan P., Fortes C., Butler J. A multicenter case-control study of diet and lung cancer among non-smokers. *Cancer Causes Control*, 2000, no. 11, pp. 49–58.

16. Freedman N.D., Leitzmann M.F., Hollenbeck A.R., Schatzkin A., Abnet C. Cigarette smoking and subsequent risk of lung cancer in men and women. *Lancet Oncol*, 2008, vol. 9, no. 7, pp. 649–656.

17. Ezzati M., Lopez A.D. Regional, disease specific patterns of smoking-attributable mortality in 2000. *Tobacco Control*, 2004, vol. 13, no. 4, pp. 388–395.

18. Krewski D., Lubin J.H., Zielinski J.M. Residential radon and risk of lung cancer: a combined analysis of 7 North American case-control studies. *Epidemiology*, 2015, no. 16, pp. 137–145.

19. Parkin D.M., Boyd L., Walker L.C. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British Journal of Cancer*, 2011, no. 105, pp.77–81.

20. Darby S., Hill D., Auvinen A.et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *Br. Med. J.*, 2015, vol. 330, pp. 223–227.

Aydinov G.T., Marchenko B.I., Sinelnikova Yu.A. Multivariate analysis of structure and contribution per shares made by potential risk factors at malignant neoplasms in trachea, bronchial tubes and lung. Health Risk Analysis, 2017, no. 1, pp. 47–55. DOI: 10.21668/health.risk/2017.1.06.eng