

UDC 616.1:57.087

DOI: 10.21668/health.risk/2017.3.04.eng

APPLICATION OF REGRESSION ANALYSIS AND CLASSIFICATION TREES IN CALCULATING ADDITIONAL POPULATION RISK OF ISCHEMIC HEART DISEASE

S.A. Maksimov, D.P. Tsygankova, G.V. Artamonova

Scientific Institution Research Institute for Complex Issues of Cardiovascular Diseases, 6 Sosnoviy blvd, Kemerovo, 650002, Russian Federation

Our research goal was to perform a comparative analysis of regression analysis application and tree classification application in calculating additional population risk on the example of ischemic heart diseases (IHD). Our research object was a random population sample comprising both male and female population aged 25-64 in Kemerovo region (1,628 people) within ESSE-RF multi-centered epidemiologic research. We considered the following IHD risk factors: lipid metabolism parameters, arterial hypertension, lifestyle factors, psychoemotional peculiarities, and social parameters. IHD occurrence was assessed as per sum of 3 epidemiologic criteria: on the basis of ECG changes coding as per Minnesota code, Rose questionnaire, and cardiac infarction in case history. We calculated additional population IHD risk determined by risk factors as per unified original algorithms, but with various statistic analysis techniques: logistic regression analysis and classification trees.

We built up mathematic models for IHD probability as per risk factors, with predictive significance equal to 83.8% for logistic regression analysis and to 71.9% for classification trees. The applied statistical analysis techniques show different contributions made by risk factors into IHD prevalence which results from absence of correlation between them.

IBD risk additional to population one and determined by risk factors as per both statistical analysis techniques in sex-age groups changed from negative values in age groups younger than 45 to positive values in older people. Increase in additional IHD risk in aged groups as per both techniques was practically linear with slight deviations. Difference in additional population risk calculated as per two statistical analysis techniques was insignificant and as a rule it didn't exceed 1.5%. Consequently, both techniques give similar results and can be equally used in calculating IHD population risk.

Key words: regression analysis, risk factor, ischemic heart disease, population risk, predictive models, statistical analysis techniques.

Cardiovascular diseases etiology is determined by a lot of factors and it makes it necessary to consider probability of their involvement and unfavorable outcomes with integral risk assessment models. Such models should include several basic factors. In most cases one respondent can have a combination of two and even more risk factors and a forecast for development and clinical course of cardiovascular diseases is much worse when several, even moderately apparent, risk factors are combined [8, 9,12,]. In particular, PROCAM research revealed that a combination of two or more factors of cardiovascular system disorders risks led to significant in-

Ó Maksimov S.A., Tsygankova D.P., Artamonova G.V., 2017

Sergei A. Maksimov – Candidate of Medical Sciences, Associate Professor; leading researcher at Laboratory for Cardiovascular Diseases Epidemiology (e-mail: m1979sa@yandex.ru; tel.: +7 (3842) 64-42-40).

Dar'ya P. Tsygankova – researcher at Laboratory for Cardiovascular Diseases Epidemiology (e-mail: cigadp@kemcardio.ru; tel.: +7 (3842) 64-34-71).

Galina V. Artamonova – Doctor of Medical Sciences, Professor, Deputy Director for Research, Head of Medical Technologies Optimization Department for cardiovascular diseases (e-mail: artamonova@kemcardio.ru; tel.: +7 (3842) 64-45-73).

crease in number of sudden deaths and cardiac infarctions [11, 13]. It made for a creation of "summary cardiovascular risk" notion [8]. Models for predicting individual risks have been widely used in everyday cardiologists' practices as they help to make decisions on prevention techniques and remedies, and on prescribing, giving up, or intensifying drug therapy [11, 13, 14]. Up to now, a great number of individual risks prediction models have been developed including such well-known ones as Framingham Risk Score, SCORE, PROCAM etc., as well as their numerous modifications [1, 2, 10, 11, 13, 14].

Population cardiovascular diseases risks are usually assessed when epidemiologic research takes place; such assessments are mostly aimed at modeling and predicting risks at population or subpopulation level (regions or population groups) [4]. Population regularities modeling is frequently used in practice to solve tasks set before a public healthcare system, for example, assessing efficiency of various approaches to treatment and prevention [3, 4, 15].

Approaches to population risk assessment which we have created are methodologically in line with individual cardiovascular risks scores [5]. In the latter case individuals are an object of analysis; and risk factors which they have (for example, smoking) and quantitative characteristics describing them (for example, age or blood pressure) are integrated into a final predictive cardiovascular risk value. When a population risk is assessed, research focuses on population groups and cardiovascular risk factors prevalence can differ significantly among them due to their heterogeneity as per medical-demographical, national, geographic, and social-economic, and other peculiarities. In this case population risk integration is based on analyzing both dis-

crepancies in risk factors prevalence and assessing their contribution into examined cardiovascular events.

Application of methodical approaches which we created allowed to derive a complex characteristics for a cardiovascular risk factors burden (17 factors) in 14 occupational groups with different working conditions [6]. Population risk analysis of ischemic heart disease (IHD) based on integral assessment of 12 risk factors enabled detecting sex-age and social-economic regularities of the examined disease [5, 7]. To assess contributions made by various risk factors into the examined diseases prevalence, we applied a decision tree technique in our research. But as this statistical analysis techniques is rather complicated when applied and is not so widely spread in biomedical research, we thought it necessary to assess possibility of other techniques application.

Our research goal was to perform a comparative analysis of regression analysis application and decision trees application to calculate additional population risk on the example of IHD.

Data and methods. This work is completed within the frameworks of multi-centered epidemiologic research "Cardiovascular diseases etiology and their risk factors in the RF" (ESSE-RF) in Kemerovo region. We chose a random population sampling, both male and female, as our research object; people were 25-64 and they all lived in Kemerovo region. Our sampling was made of 1,628 people, 700 males (43.0%) and 928 females (57.0%).

The research was performed in full conformity with Good Clinical Practice standards and Helsinki Declaration principles. The research protocol was approved by the Ethical Committee of the Scientific Research Institute for Complex Issues of Cardiovascular Diseases. All the partici-

pants gave their written informed consent to take part in it.

We studied such IHD risk factors as lipid metabolism parameters, arterial hypertension, lifestyle factors, psychoemotional peculiarities, social parameters (low incomes, poor education, having no family, being unemployed).

According to WHO classification, we considered body mass index being more than 30 kg/m² being a proof of obesity. Arterial hypertension was classified as per WHO/ISH criteria (1999) when systolic blood pressure was not lower than 140 mm Hg, and diastolic one, not lower than 90 mm Hg, as well as when both these parameters were equal to a standard but under hypertensive drugs.

Hypercholesterolemia was classified when crude cholesterol concentration was higher than 5.0 mmol/l; hypertriglyceridemia, when triglycerides were higher than 1.7 mmol/l; high low density lipoproteins level was registered at their values exceeding 3.0 mmol/l; low high density lipoproteins level, lower than 1.0 mmol/l; hyperglycemia on an empty stomach was registered when dextrose level was higher than 5.6 mmol/l.

We applied Hospital Anxiety and Depression Scale (HADS) validated in the RF to assess depression; susceptibility to stress was assessed as per Perceived Stress Scale. We calculated 75-percentile on the basis of the obtained ordinal series, and values higher than 8 scores and more as per depression scale and 5 scores and more as per stress scale were considered to be a risk factor.

IDH occurrence was assessed as per a sum of three epidemiologic criteria: on the basis of coding of ECG changes as per Minnesota code, Rose questionnaire (pectoris) and cardiac infarction in a case history.

Additional population IHD risk caused by risk factors was calculated as per unified algorithms but with application of different statistical analysis techniques.

At the first stage, we calculated a contribution made by the examined risk factors into IHD prevalence values for the whole sampling. To do this, we first applied logical regression analysis, and then, decision trees. To remove possible modifying influence exerted by sex and age in the course of logical regression analysis, we introduced them into it as well. B-coefficient values were used as a value describing a contribution made by a risk factor into IHD prevalence.

When we applied decision trees we chose on a discriminant one-dimensional branching technique for categorical and ordinal predictors. We took equal costs of wrong objects classification and a priori probabilities proportionate to sizes of dependent variable classes as fidelity criteria. Branching was stopped according to a pruning rule as per an error in classification, minimal number of incorrectly classified objects being equal to 12, and a standard error, equal to 1.0. We used values of predictors significant ranks as parameters of contributions made by risk factors into IHD prevalence.

Further on, calculation was made as per the same scheme. We calculated risk factors prevalence in sex-age groups and in the whole sampling. Load with IHD risk factors was calculated as a sum of multiplying risk factors prevalence by their contribution into IHD risks as per the following formula:

$$P = \sum(RC)_n, \quad (1)$$

Where P is IHD risk factors load; R is a risk factor prevalence, %; C is a risk factor contribution into IHD prevalence values.

Then we calculated a difference between risk factors load in age-sex groups in comparison with the overall sampling as per formula:

$$\Delta P = P_{\text{гp.}} - P_{\text{B.}}, \quad (2)$$

Where ΔP is a difference between risk factors load in age-sex groups in comparison with the overall sampling as per formula; $P_{\text{гp.}}$ is a risk factors load in age-sex groups; $P_{\text{B.}}$ is a risk factor load in the overall sampling.

On the next stage we applied linear regression analysis to determine frequency of IHD association with the risk factor loads values in age-sex groups. B-coefficient which was obtained in the course of linear regression analysis was used to recalculate difference in risk factor loads in age-sex groups into population risk values as per formula:

$$P\% = \Delta P B, \quad (3)$$

Where $P\%$ is IHD risk which is additional to population one and caused by risk factors, %; B is B-coefficient showing a correlation between IHD frequency and risk factors load in linear regression analysis.

Critical level of statistic significance and p-level for a choice on a branching variable (for decision trees) was considered equal to 0.05.

Results and discussion. As per logistic analysis results statistically significant (or very close to such, $0.1 > p > 0.05$) associations with IHD allowing for age and sex are observed as per following risk factors: hypertension, hypertriglyceridemia, hyperglycemia, obesity, pancreatic diabetes, smoking, stress, depression, absence of high education, and being unemployed (Table 1). We used only these risk factors in our further analysis.

Table 1 contains B-coefficients as per regression analysis (from 0.046 hyperglycemia to 0.491 pancreatic diabetes) and ranks as per decision trees (from 20 for smoking to 100 for obesity) which describe the selected risk factors and which were further used in risk factors load calculations (formula 1) as per two techniques. We should note that predictive significance of a mathematical model for IHD probability as per risk factors sum was equal to 83.8% when this model was designed via logistic regression analysis, and to 71.9% in case when decision trees were applied.

We didn't detect any statistically significant correlation between B-coefficient values and risk factors significance ranks as correlation coefficient was equal to -0.32 at $p = 0.37$. Therefore, different statistic analysis techniques gave different values of contributions made by the examined risk factors into IHD prevalence parameters.

Tables 2 and 3 contain information on IHD and risk factors prevalence in age-sex groups and in the overall sampling. Risk factors prevalence in the overall sampling varied from 3.9% (pancreatic diabetes) to 60.7% (absence of high education). IHD prevalence in the overall sampling reached 16.8%; 13.5% in males; 19.2%, in females.

Risk factors load (as per formula 1) and difference in loads in sex-age groups in comparison with the overall sampling (as per formula 2) calculated via two techniques is given in tables 2 and 3. Risk factors load calculated with logistic regression analysis amounted to 71.8 for the overall sampling; 72.0, for males; 71.7, for females. This load naturally grew with age and increased from 49.6 to 93.8 for males, and from 45.6 to 97.0 for females.

The same regularities were observed for risk factor load calculated with decision trees. The load amounted to 19,459.6 for the overall sampling; to 18,831.7, for males;

Table 1

Influence exerted by risk factors on IHD probability as per regression analysis data (allowing for age and sex) and rank values as pre decision trees data.

| Risk factors | OR | 95% CI | <i>p</i> -level | B-coeff. | Ranks |
|-----------------------|------|-----------|-----------------|----------|-------|
| Hypertension | 1,28 | 0,97–1,70 | 0,082 | 0,118 | 68 |
| Hypercholesterolemia | 0,89 | 0,67–1,18 | 0,41 | – | – |
| Hypertriglyceridemia | 1,66 | 1,22–2,25 | 0,0013 | 0,355 | 80 |
| High LDL levels | 0,92 | 0,68–1,24 | 0,58 | – | – |
| Low HDL levels | 0,72 | 0,21–2,49 | 0,61 | – | – |
| Hyperglycemia | 1,37 | 0,99–1,90 | 0,060 | 0,046 | 85 |
| Obesity | 1,49 | 1,13–1,97 | 0,0048 | 0,260 | 100 |
| Pancreatic diabetes | 1,96 | 1,13–3,41 | 0,016 | 0,491 | 81 |
| Smoking | 1,59 | 1,16–2,16 | 0,0036 | 0,359 | 20 |
| Low physical activity | 0,97 | 0,70–1,34 | 0,83 | – | – |
| Stress | 1,56 | 1,15–2,11 | 0,0042 | 0,311 | 74 |
| Depression | 1,88 | 1,39–2,55 | 0,000049 | 0,402 | 47 |
| Income | 0,98 | 0,65–1,47 | 0,91 | – | – |
| Education | 1,57 | 1,17–2,11 | 0,0026 | 0,292 | 71 |
| Employment | 1,40 | 1,04–1,88 | 0,028 | 0,170 | 83 |
| Family | 1,02 | 0,76–1,37 | 0,89 | – | – |

Table 2

IHD and risk factors prevalence and risk factors load in male age groups and in the overall sampling

| Risk factors | Age groups, number of observations | | | | | | | | All men (700) | Overall sampling |
|-------------------------|------------------------------------|------------|------------|------------|------------|------------|-------------|------------|---------------|------------------|
| | Younger than 30 (86) | 31–35 (85) | 36–40 (88) | 41–45 (65) | 46–50 (98) | 51–55 (94) | 56–60 (107) | 61–65 (77) | | |
| Hypertension, % | 20,9 | 37,6 | 44,3 | 56,9 | 51,0 | 69,1 | 61,7 | 71,4 | 51,7 | 43,3 |
| Hypertriglyceridemia, % | 14,1 | 17,6 | 21,8 | 24,6 | 25,5 | 25,5 | 27,4 | 23,7 | 22,7 | 20,3 |
| Hyperglycemia, % | 8,2 | 7,1 | 11,5 | 7,7 | 17,3 | 28,7 | 22,6 | 43,4 | 18,5 | 17,2 |
| Obesity, % | 12,9 | 22,3 | 21,6 | 35,4 | 36,7 | 35,5 | 34,6 | 37,7 | 29,7 | 35,2 |
| Pancreatic diabetes, % | 1,2 | 0,0 | 0,0 | 0,0 | 4,1 | 5,3 | 4,7 | 14,7 | 3,7 | 3,9 |
| Smoking, % | 47,7 | 47,1 | 46,6 | 47,7 | 50,0 | 48,9 | 43,0 | 33,8 | 45,7 | 30,5 |
| Stress, % | 10,5 | 15,3 | 12,5 | 9,2 | 13,3 | 11,7 | 14,0 | 16,9 | 13,0 | 22,6 |
| Depression, % | 8,1 | 11,8 | 10,2 | 9,2 | 13,3 | 13,8 | 16,8 | 27,3 | 13,9 | 19,0 |
| Education, % | 43,0 | 51,8 | 61,4 | 55,4 | 70,4 | 71,3 | 63,6 | 63,6 | 60,6 | 60,7 |
| Employment, % | 9,3 | 9,4 | 9,1 | 7,7 | 16,3 | 26,6 | 32,1 | 64,9 | 22,0 | 25,4 |
| IHD, % | 3,5 | 3,5 | 5,9 | 9,2 | 10,3 | 22,3 | 21,5 | 30,3 | 13,5 | 16,8 |
| Load 1 | 49,6 | 59,9 | 63,3 | 66,2 | 78,2 | 82,4 | 79,9 | 93,8 | 72,0 | 71,8 |
| Difference 1 | –22,2 | –11,9 | –8,5 | –5,6 | 6,4 | 10,6 | 8,1 | 22,0 | 0,2 | 0 |
| Load 2 | 10570 | 13885,1 | 15345 | 16671,4 | 19941,2 | 22920,1 | 22014,8 | 28512,9 | 18831,7 | 19459,6 |
| Difference 2 | –8889,6 | –5574,5 | –4114,6 | –2788,6 | 481,6 | 3460,5 | 2555,2 | 9053,3 | –627,9 | 0,0 |

Table 3

IHD and risk factors prevalence and risk factors load in female age groups

| Risk factors | Age groups, number of observations | | | | | | | | All women (928) |
|-------------------------|------------------------------------|------------|------------|------------|-------------|-------------|-------------|-------------|-----------------|
| | Younger than 30 (97) | 31–35 (97) | 36–40 (86) | 41–45 (91) | 46–50 (112) | 51–55 (159) | 56–60 (170) | 61–65 (116) | |
| Hypertension, % | 10,3 | 8,2 | 18,6 | 30,8 | 42,0 | 50,3 | 52,3 | 57,8 | 37,2 |
| Hypertriglyceridemia, % | 5,2 | 9,3 | 9,3 | 17,6 | 26,4 | 23,4 | 20,4 | 28,1 | 18,5 |
| Hyperglycemia, % | 0,0 | 2,1 | 12,8 | 13,2 | 15,5 | 15,8 | 26,9 | 32,5 | 16,2 |
| Obesity, % | 9,3 | 14,4 | 27,9 | 41,8 | 46,4 | 44,9 | 50,6 | 61,2 | 39,4 |
| Pancreatic diabetes, % | 0,0 | 0,0 | 1,2 | 2,2 | 2,7 | 5,1 | 6,5 | 10,6 | 4,0 |
| Smoking, % | 26,8 | 29,9 | 29,1 | 26,4 | 17,9 | 17,0 | 10,0 | 6,9 | 19,0 |
| Stress, % | 26,8 | 26,8 | 29,4 | 27,8 | 36,6 | 28,9 | 31,4 | 29,3 | 29,8 |
| Depression, % | 11,3 | 7,2 | 17,4 | 17,6 | 23,2 | 28,9 | 31,8 | 31,9 | 22,8 |
| Education, % | 48,5 | 33,0 | 59,3 | 51,6 | 71,4 | 63,5 | 69,4 | 76,7 | 60,9 |
| Employment, % | 20,6 | 12,4 | 24,4 | 11,0 | 14,3 | 18,9 | 44,7 | 63,7 | 27,9 |
| IHD, % | 4,1 | 14,6 | 13,2 | 15,4 | 18,0 | 25,2 | 26,5 | 25,2 | 19,2 |
| Load 1 | 45,6 | 41,8 | 62,0 | 64,6 | 78,8 | 77,6 | 85,0 | 97,0 | 71,7 |
| Difference 1 | -26,2 | -30,0 | -9,8 | -7,2 | 7 | 5,8 | 13,2 | 25,2 | -0,1 |
| Load 2 | 10250 | 9211,9 | 15794,9 | 16971,6 | 21557,3 | 21452,6 | 25717,1 | 30457,8 | 19947 |
| Difference 2 | -9209,6 | -10247,7 | -3664,7 | -2488 | 2097,7 | 1993 | 6257,5 | 10998,2 | 487,4 |

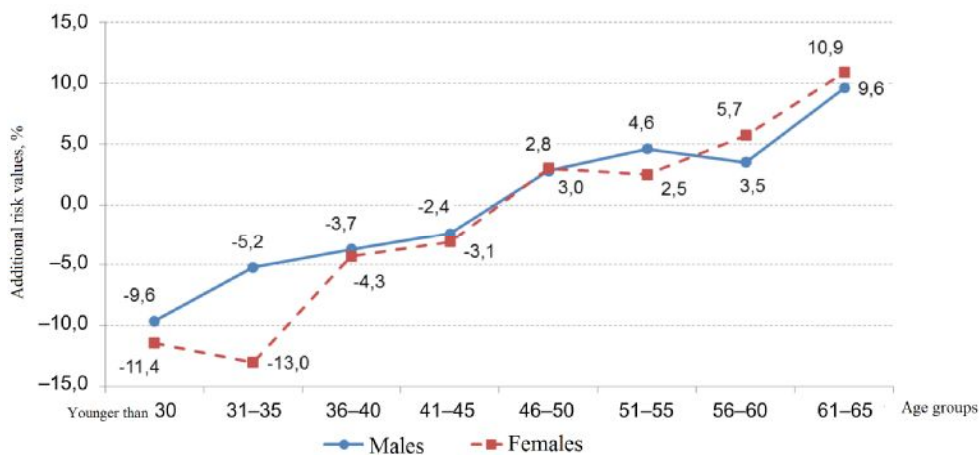
to 19,947.0, for females. The load grew with age from 10,570.0 to 28,512.9 for males and from 10,250.0 to 30,457.8 for females.

Mathematical models created with linear regression analysis showed that risk factor load could account for IHD prevalence in age-sex groups by 62.6% (logistic regression) and by 71.7% (classification trees). Regression B-coefficients calculated as per these models were used to recalculate differences in risk factors loads into population risk values as per formula 3.

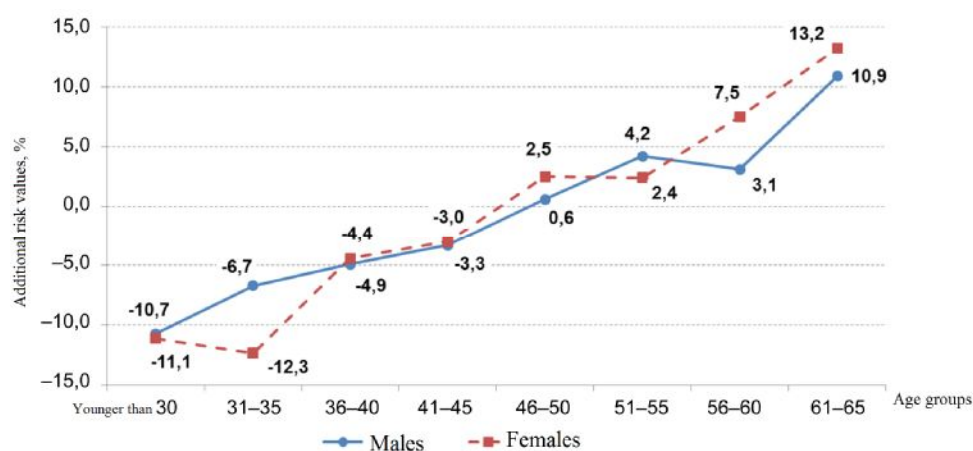
B-coefficient was equal to 0.4345 when logistic regression analysis was applied, and to 0.0012, when decision trees were applied.

IHD risk, additional to population one, caused by risk factors, and calculated

as per formula 3, is shown on Figure 1 (regression analysis) and on Figure 1 (decision trees). Additional risk in younger age groups (younger than 46) calculated as per both techniques was lower than population one, from -2 % to -11 %; but it reached values higher than population one, from 0.5% to 13%, in age groups older than 45. IHD additional risk increase in age groups was practically linear with slight deviations: lower risk for females aged 31-35 and 51-55; and for males aged 56-60 in comparison with the previous age group (Figure 2). Here deviations from linearity of IHD additional risk growth with age are observed as per both techniques, logistic regression analysis and decision trees.

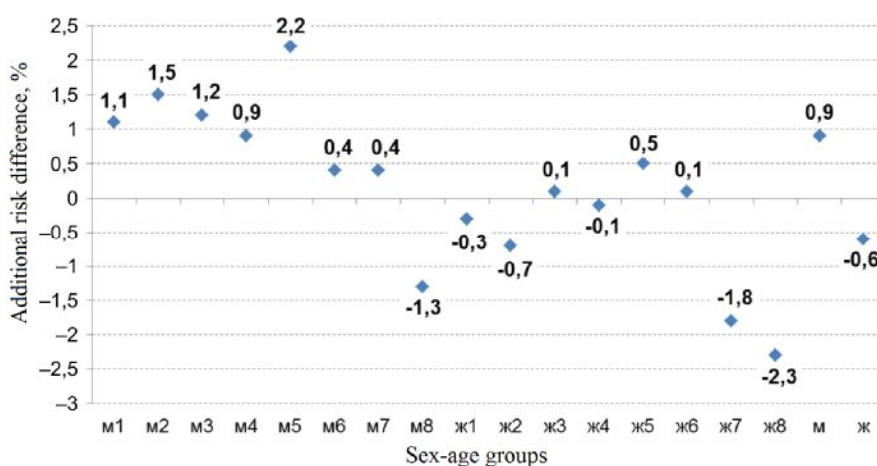


a



b

Figure 1. IHD risk additional to population one, caused by risk factors, %:
a – regression analysis; b – decision trees



Note: м – are males, ж – are females, 1 – younger than 30, 2 – 31-35, 3 – 36-40, 4 – 41-45, 5 – 46-50, 6 – 51-55, 7 – 56-60, 8 – 61-65.

Figure 2. Difference between additional population risk calculated as per two statistical analysis techniques

Data which are obtained via these techniques allow calculating load with risk factors. And here, as per both techniques (regression analysis and decision trees) difference in load in age-sex groups in comparison with the overall sampling changes from negative values in age groups up to 45 years (inclusive), to positive values in people older than 45. IHD risk additional to population one and caused by risk factors changes similarly. Increase in additional IHD risk in age groups as per both techniques is practically linear, with slight deviations. Additional IHD risk in age groups younger than 30, both sexes, is lower than population one by 9.6–11.4 %; it becomes higher than population one in age groups

closer to 50 (by 0.6–2.8 %); and it reaches its maximum values by 65 (by 9.6–13.2 %).

Despite different statistical analysis techniques giving different values of contributions made by the examined risk factors into IHD prevalence parameters, the correlation between values of additional population risk calculated with regression analysis and with decision trees is strong and statistically significant. Difference of additional population risk calculated as per different statistical analysis techniques is slight and, as a rule, doesn't exceed 1.5%. Therefore, both techniques give similar results and can be equally applied to calculate population IHD risks.

References

1. Batyushin M.M. Modernizatsiya shkaly SCORE otsenki desyatiletnego riska serdechno-sosudistoi smertnosti [Modernization of 10-year cardiovascular death risk scale SCORE]. *Rossiiskii kardiologicheskii zhurnal*, 2005, no. 6, pp. 40–44. (in Russian).
2. Zykov M.V., Zyкова D.S., Kashtalap V.V., Pecherina T.B., Barbarash O.L. Znachimost' mul'tifokal'nogo ateroskleroza dlya modifikatsii shkaly otdalennogo riska smertnosti GRACE u bol'nykh ostrym koronarnym sindromom s pod'emom segmenta ST [The prognostic value of peripheral arteries diseases in patients with st-segment elevation myocardial infarction]. *Ateroskleroz*, 2012, vol. 8, no. 1, pp. 14–20 (in Russian).
3. Kontsevaya A.V., Suvorova E.I., Khudyakov M.B. Ekonomicheskaya effektivnost' renal'noy denervatsii u patsientov s rezistentnoy arterial'noy gipertoniey: rezul'taty markovskogo modelirovaniya [Economic efficiency of renal denervation in patients with resistant hypertension: results of markov modeling]. *Kardiologiya*, 2014, vol. 54, no. 1, pp. 41–47 (in Russian).
4. Kontsevaya A.V., Shal'nova S.A. Populyatsionnye modeli prognozirovaniya serdechno-sosudistogo riska: tselesoobraznost' modelirovaniya i analiticheskii obzor sushchestvuyushchikh modeley [Population models of cardiovascular risk prediction: expedience of modeling and analytic review of current models]. *Kardiovaskulyarnaya terapiya i profilaktika*, 2015, vol. 14, no. 6, pp. 54–58 (in Russian).
5. Maksimov S.A., Indukaeva E.V., Artamonova G.V. Integral'naya otsenka riska ishemicheskoy bolezni serdtsa v epidemiologicheskikh issledovaniyakh (ESSE-RF v Kemerovskoy oblasti). Soobshchenie I: vozrastno-polovye determinant [Integral assessment of coronary heart disease risk in the epidemiological studies (ESSE-RF in the Kemerovo Region). Communication 1: Age and sex determinants]. *Profilakticheskaya meditsina*, 2015, vol.18, no. 6, pp. 34–39 (in Russian).
6. Maksimov S.A., Skripchenko A.E., Artamonova G.V. Integral'naya otsenka faktorov riska professional'noy obuslovlennosti arterial'noy gipertenzii [Integral assessment of work-

related arterial hypertension risk factors]. *Rossiyskiy kardiologicheskiy zhurnal*, 2015, vol.120, no. 4, pp. 38–42 (in Russian).

7. Maksimov S.A., Tabakaev M.V., Artamonova G.V. Integral'naya otsenka riska ishemicheskoy bolezni serdtsa v epidemiologicheskikh issledovaniyakh (ESSE-RF v Kemerovskoy oblasti). Soobshchenie II: sotsial'no-ekonomicheskie determinant [Integral assessment of coronary heart disease risk in the epidemiological studies (ESSE-RF in the Kemerovo Region). Communication II: Socioeconomic determinants]. *Profilakticheskaya meditsina*, 2016, vol.19, no. 1, pp. 24–29 (in Russian).

8. Mamedov M.N., Chepurina N.A. Summarnyi serdechno-sosudisty risk: ot teorii k praktike: posobie dlya vrachei [Total cardio-vascular risk: from theory to practice: a manual for physicians]. In: R.G. Oganov, ed. Moscow, 2007, 23 p. (in Russian).

9. Boytsov S.A., Vatolina M.A., Samorodskaya I.V., Barbarash O.L., Ovcharenko O.A., Kondrikova N.V. Mnenie vrachey o roli otdel'nykh faktorov smernosti ot bolezney sistemy krovoobrashcheniya v regionakh Rossiyskoy Federatsii [Medical care practitioners' opinion on the role of specific factors contributing to the mortality from circulatory system disease in the regions of the Russian Federation]. *Kompleksnyye problemy serdechno-sosudistyykh zabolevaniy*, 2015, no. 4, pp. 53–60 (in Russian).

10. Oganov R.G., Shal'nova S.A., Kalinina A.M., Deev A.D., Glazachev O.S., Gusev E.I., Belyaeva I.A., Sudarev A.M. Novyy sposob otsenki individual'nogo serdechno-sosudistogo summarnogo riska dlya naseleniya Rossii [The novel method of assessment of individual total cardiovascular risk for the population of Russia]. *Kardiologiya*, 2008, vol. 48, no. 5, pp. 87–91 (in Russian).

11. Assmann G. Assessment of cardiovascular risk – PROCAM and new algorithms. *Bio-med. Tech. (Berl)*, 2005, vol. 50, no. 7–8, pp. 227–232.

12. Vanuzzo D., Pilotto L., Mirolo R., Pirelli S. Cardiovascular risk and cardiometabolic risk: an epidemiological evaluation. *G. Ital. Cardiol. (Rome)*, 2008, vol. 9, no.4, pp. 6S–17S.

13. Gorenou V., Hagen A. Overview of risk – estimation tools for primary prevention of cardiovascular diseases in European populations. *Cent. Eur. J. Public. Health.*, 2015, vol. 23, no.2, pp. 91–99.

14. Berger J.S., Jordan C.O., Lloyd-Jones D., Blumenthal R.S. Screening for cardiovascular risk in asymptomatic patients. *J. Am. Coll. Cardiol.*, 2010, vol. 55, no.12, pp. 1169–1177.

15. Kahn R., Robertson R.M., Smith R., Eddy D. The impact of prevention on reducing the burden of cardiovascular disease. *Circulation*, 2008, vol. 118, pp. 576–585.

Maksimov S.A., Tsygankova D.P., Artamonova G.V. Application of regression analysis and classification trees in calculating additional population risk of ischemic eart disease. Health Risk Analysis, 2017, no. 3, pp. 31–39. DOI: 10.21668/health.risk/2017.3.04.eng

Received: 05.06.2017

Accepted: 20.09.2017

Published: 30.09.2017