

УДК 616.1: 57.087

DOI: 10.21668/health.risk/2017.3.04

## ПРИМЕНЕНИЕ РЕГРЕССИОННОГО АНАЛИЗА И ДЕРЕВЬЕВ КЛАССИФИКАЦИИ ДЛЯ РАСЧЕТА ДОПОЛНИТЕЛЬНОГО ПОПУЛЯЦИОННОГО РИСКА ИШЕМИЧЕСКОЙ БОЛЕЗНИ СЕРДЦА

С.А. Максимов, Д.П. Цыганкова, Г.В. Артамонова

Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний, Россия, 650002, г. Кемерово, Сосновый бульвар, 6

*Цель исследования состояла в проведении сравнительного анализа применения регрессионного анализа и деревьев классификации для расчета дополнительного популяционного риска на примере ишемической болезни сердца (ИБС). Объектом исследования являлась случайная популяционная выборка мужского и женского взрослого населения в возрасте 25–64 лет, проживающего в Кемеровской области (1628 человек), в рамках многоцентрового эпидемиологического исследования ЭССЕ-РФ. В качестве факторов риска ИБС рассматривались показатели липидного обмена, артериальная гипертензия, факторы образа жизни, психоэмоциональные особенности, социальные показатели. Оценка наличия ИБС проводилась по сумме трех эпидемиологических критериев: на основе кодирования ЭКГ-изменений по Миннесотскому коду, опросника Rose и инфаркта миокарда в анамнезе. Расчет дополнительного популяционного риска ИБС, обусловленного факторами риска, проводился по единым оригинальным алгоритмам, но с помощью разных методов статистического анализа: логистического регрессионного анализа и деревьев классификации.*

*По факторам риска построены математические модели вероятности ИБС с прогностической значимостью для логистического регрессионного анализа 83,8 %, деревьев классификации – 71,9 %. Используемые методы статистического анализа показывают разные значения вклада факторов риска в показатели распространенности ИБС, что следует из отсутствия корреляционной связи между ними.*

*По обоим методам статистического анализа дополнительный к популяционному риск ИБС, обусловленный нагрузкой факторами риска, в половозрастных группах изменяется с отрицательных значений в возрастных группах до 45 лет включительно до положительных значений у лиц более старшего возраста. Увеличение дополнительного риска ИБС в возрастных группах по обоим методам практически линейное, с небольшими отклонениями. Разница дополнительного популяционного риска, рассчитанного по двум методам статистического анализа, незначительная, как правило, не превышает 1,5 %. Следовательно, оба метода показывают схожие результаты и могут в равной степени использоваться при расчете популяционного риска ИБС.*

**Ключевые слова:** регрессионный анализ, фактор риска, ишемическая болезнь сердца, популяционный риск, прогностические модели, методы статистического анализа.

Многофакторность этиологии сердечно-сосудистых заболеваний определяет необходимость рассмотрения вероятности их развития и неблагоприятных исходов с помощью интегральных моделей оценки рисков, включающих в себя несколько основных факторов. В большинстве случаев у одного респондента возможно сочетание двух факторов риска и более, а прогноз развития и течения сердечно-сосудистых заболеваний значительно хуже при сочетании нескольких, даже умеренно выраженных факторов риска [8, 9, 12]. В частности, ис-

следование PROCAM показало, что сочетание двух и более факторов риска развития нарушений функций сердечно-сосудистой системы приводит к значительному увеличению количества инцидентов внезапной смерти и инфаркта миокарда [11, 13]. Это послужило появлению понятия «суммарный сердечно-сосудистый риск» [8]. Модели прогнозирования индивидуального риска давно вошли в рутинную практику кардиологов, помогая принимать решения о средствах и методах профилактики, назначении, отказе или интенсификации меди-

© Максимов С.А., Цыганкова Д.П., Артамонова Г.В., 2017

**Максимов Сергей Алексеевич** – кандидат медицинских наук, доцент, ведущий научный сотрудник лаборатории эпидемиологии сердечно-сосудистых заболеваний (e-mail: m1979sa@yandex.ru; тел. 8 (3842) 64-42-40).

**Цыганкова Дарья Павловна** – научный сотрудник лаборатории эпидемиологии сердечно-сосудистых заболеваний, очный аспирант кафедры кардиологии и сердечно-сосудистой хирургии (e-mail: cigadp@kemcardio.ru; тел. 8 (3842) 64-34-71).

**Артамонова Галина Владимировна** – доктор медицинских наук, профессор, заместитель директора по научным вопросам, заведующий отделом оптимизации медицинских технологий при сердечно-сосудистых заболеваниях (e-mail: artamonova@kemcardio.ru; тел. 8 (3842) 64-45-73).

каментозной терапии [11, 13, 14]. К настоящему времени разработано большое количество моделей прогнозирования индивидуального риска, включая такие широко известные, как Фремингемская шкала, SCORE, PROCAM и другие, а также их многочисленные модификации [1, 2, 10, 11, 13, 14].

Оценка популяционного риска сердечно-сосудистых заболеваний применяется, как правило, в эпидемиологических исследованиях и направлена на моделирование и прогнозирование на уровне популяции или субпопуляции (регионы, группы населения) [4]. На практике моделирование популяционных закономерностей достаточно часто используется для решения задач системы здравоохранения, например, оценки эффективности разных подходов лечения и профилактики [3, 4, 15].

Разработанные нами подходы к оценке популяционного риска методологически перекликаются со шкалами индивидуального сердечно-сосудистого риска [5]. В последнем случае объектом анализа являются индивиды, а наличие у них факторов риска (например, курение) и их количественные характеристики (например, возраст, уровень артериального давления) интегрируются в итоговое прогностическое значение сердечно-сосудистого риска. При оценке популяционного риска объектом исследования являются группы населения, распространенность факторов сердечно-сосудистого риска у которых может в значительной степени отличаться вследствие гетерогенности по медико-демографическим, национальным, географическим, социально-экономическим и другим особенностям. В этом случае интегрирование популяционного риска основывается на анализе как различий распространенности факторов риска, так и оценке их вклада в изучаемые сердечно-сосудистые события.

Применение разработанных нами методических подходов позволило комплексно охарактеризовать бремя факторов сердечно-сосудистого риска (17 факторов) в 14 профессиональных группах с различными условиями труда [6]. Анализ популяционного риска ишемической болезни сердца (ИБС) на основе интегральной оценки 12 факторов риска позволил выявить половозрастные и социально-экономические закономерности изучаемого заболевания [5, 7]. В этих исследованиях для оценки вклада факторов риска в распространенность изучаемых заболеваний применялись деревья классификации. Вследствие того, что данный метод

статистического анализа достаточно сложен в применении и не слишком распространен в биомедицинских исследованиях, необходимо оценить возможность использования других методов.

**Цель исследования** – сравнительный анализ применения регрессионного анализа и деревьев классификации для расчета дополнительной популяционного риска на примере ИБС.

**Материалы и методы.** Работа выполнена в рамках многоцентрового эпидемиологического исследования «Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в Российской Федерации» (ЭССЕ-РФ) в Кемеровской области. Объектом исследования явилась случайная популяционная выборка мужского и женского взрослого населения в возрасте 25–64 лет, проживающего в Кемеровской области. Объем выборки составил 1628 человек, из них 700 мужчин (43,0 %) и 928 женщин (57,0 %).

Исследование выполнено в соответствии со стандартами надлежущей клинической практики (Good Clinical Practice) и принципами Хельсинкской декларации. Протокол исследования был одобрен этическим комитетом НИИ комплексных проблем сердечно-сосудистых заболеваний. До включения в исследование у всех участников было получено письменное информированное согласие.

В качестве факторов риска ИБС рассматривались показатели липидного обмена, артериальная гипертензия, факторы образа жизни, психоэмоциональные особенности, социальные показатели (низкий уровень дохода, отсутствие высшего образования, семьи, работы).

В соответствии с классификацией ВОЗ под ожирением подразумевали индекс Кетле свыше 30 кг/м<sup>2</sup>. Артериальная гипертензия классифицировалась по критериям ВОЗ/МОГ (1999) при систолическом артериальном давлении не менее 140 мм рт. ст. и диастолическом артериальном давлении не менее 90 мм рт. ст., а также при нормальных значениях на фоне приема гипотензивных препаратов.

Гиперхолестеринемия классифицировалась при концентрации общего холестерина более 5,0 ммоль/л, гипертриглицеридемия – триглицеридов более 1,7 ммоль/л, высокий уровень липопротеидов низкой плотности (ЛПНП) – при значениях более 3,0 ммоль/л, низкий уровень липопротеидов высокой плотности (ЛПВП) – при значениях менее 1,0 ммоль/л, гипергликемия натощак – глюкозы более 5,6 ммоль/л.

Для оценки уровня депрессии использовали валидизированную в России госпитальную шкалу тревоги и депрессии (Hospital Anxiety and Depression Scale – HADS), для определения восприимчивости к стрессу – шкалу Perceived Stress Scale. По полученным порядковым рядам рассчитывали 75-й перцентиль, значения выше рассматривались как фактор риска: по шкале депрессии – 8 баллов и выше, по уровню стресса – 5 баллов и выше.

Оценка наличия ИБС проводилась по сумме трех эпидемиологических критериев: на основе кодирования ЭКГ-изменений по Миннесотскому коду, опросника Rose (стенокардия напряжения) и инфаркта миокарда в анамнезе.

Расчет дополнительного популяционного риска ИБС, обусловленного факторами риска, проводился по единым алгоритмам, но с помощью разных методов статистического анализа.

На первом этапе рассчитан вклад изучаемых факторов риска в значения распространенности ИБС по всей выборке. Для этой цели в первом случае применялся логистический регрессионный анализ, во втором – деревья классификации. При использовании логистического регрессионного анализа для устранения возможного модифицирующего влияния пола и возраста их также вводили в анализ. В качестве показателя вклада фактора риска в распространенность ИБС оценивались значения  $B$ -коэффициента.

При использовании деревьев классификации применялся метод дискриминантного одномерного ветвления для категориальных и порядковых предикторов. В качестве критериев точности прогноза взяты равные цены неправильной классификации объектов и априорные вероятности, пропорциональные размерам классов зависимой переменной. Остановка ветвления производилась по правилу отсекающего по ошибке классификации, при этом минимальное число неправильно классифицируемых объектов принималось равным 12, величина стандартной ошибки – 1,0. В качестве показателя вклада фактора риска в распространенность ИБС использовались значения рангов значимости предикторов.

Далее расчет проводился по единой схеме. Рассчитывалась распространенность факторов риска в половозрастных группах и в целом по выборке. Нагрузка факторами риска ИБС рассчитывалась как сумма произведений распространенности факторов риска с их вкладом в риск развития ИБС по формуле

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

где  $P$  – нагрузка факторами риска ИБС;  $R$  – распространенность фактора риска, %;  $C$  – вклад фактора риска в значения распространенности ИБС.

Далее рассчитывалась разница нагрузки факторами риска в половозрастных группах по сравнению с общей выборкой по формуле

$$\Delta P = P_{гр.} - P_{в.}, \quad (2)$$

где  $\Delta P$  – разница нагрузки факторами риска в половозрастных группах по сравнению с общей выборкой по формуле;  $P_{гр.}$  – нагрузка факторами риска в половозрастных группах;  $P_{в.}$  – нагрузка факторами риска в общей выборке.

На следующем этапе с помощью линейного регрессионного анализа определена ассоциация частоты ИБС со значениями нагрузок факторами риска в половозрастных группах. По полученному в ходе линейного регрессионного анализа  $B$ -коэффициенту разница нагрузок факторами риска в половозрастных группах переводилась в значения популяционного риска по формуле

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

где  $P\%$  – дополнительный к популяционному риск ИБС, обусловленный факторами риска, %;  $B$  –  $B$ -коэффициент связи частоты ИБС с нагрузкой факторами риска в линейном регрессионном анализе.

Критическим уровнем статистической значимости и  $p$ -уровнем для выбора переменной ветвления (для деревьев классификации) принимались значения 0,05.

**Результаты и их обсуждение.** По результатам логистического регрессионного анализа статистически значимые (либо приближающиеся к таковым,  $0,1 > p > 0,05$ ) ассоциации с ИБС, с учетом возраста и пола, наблюдаются по факторам риска: гипертензия, гипертриглицеридемия, гипергликемия, ожирение, сахарный диабет, курение, стресс, депрессия, отсутствие высшего образования и работы. Для дальнейшего анализа использовались только данные факторы риска.

В табл. 1 представлены  $B$ -коэффициенты по регрессионному анализу (от 0,046 – гипергликемия до 0,491 – сахарный диабет) и ранги по деревьям классификации (от 20 – у курения до 100 – у ожирения) выбранных факторов риска, которые далее использовались в расчете нагрузки факторами риска по формуле (1) по двум методам. Необходимо отметить, что про-

Таблица 1

Влияние факторов риска на вероятность ИБС по данным регрессионного анализа (с учетом пола и возраста) и значения рангов по данным деревьев классификации

Фактор риска	ОШ	95%-ный ДИ	<i>p</i> -уровень	В-коэфф.	Ранг
Гипертензия	1,28	0,97–1,70	0,082	0,118	68
Гиперхолестеринемия	0,89	0,67–1,18	0,41	–	–
Гипертриглицеридемия	1,66	1,22–2,25	0,0013	0,355	80
Высокие уровни ЛПНП	0,92	0,68–1,24	0,58	–	–
Низкие уровни ЛПВП	0,72	0,21–2,49	0,61	–	–
Гипергликемия	1,37	0,99–1,90	0,060	0,046	85
Ожирение	1,49	1,13–1,97	0,0048	0,260	100
Сахарный диабет	1,96	1,13–3,41	0,016	0,491	81
Курение	1,59	1,16–2,16	0,0036	0,359	20
Низкая физическая активность	0,97	0,70–1,34	0,83	–	–
Стресс	1,56	1,15–2,11	0,0042	0,311	74
Депрессия	1,88	1,39–2,55	0,000049	0,402	47
Доход	0,98	0,65–1,47	0,91	–	–
Образование	1,57	1,17–2,11	0,0026	0,292	71
Работа	1,40	1,04–1,88	0,028	0,170	83
Семья	1,02	0,76–1,37	0,89	–	–

гностическая значимость математической модели вероятности ИБС по сумме факторов риска, построенной с помощью логистического регрессионного анализа, составляет 83,8 %, с помощью деревьев классификации – 71,9 %.

Между значениями В-коэффициентов и рангами значимости факторов риска отсутствует статистически значимая связь, коэффициент корреляции составляет –0,32 при  $p = 0,37$ . Следовательно, разные методы статистического анализа дают разные значения вклада изучаемых факторов риска в показатели распространенности ИБС.

В табл. 2 и 3 представлена распространенность факторов риска и ИБС в половозрастных группах и в целом по выборке. В общей выборке распространенность факторов риска составляет от 3,9 % (сахарный диабет) до 60,7 % (отсутствие высшего образования). Распространенность ИБС в целом по выборке достигает 16,8 %, у мужчин – 13,5 %, у женщин – 19,2 %.

Рассчитанная нагрузка факторами риска по формуле (1) и разница нагрузки в половозрастных группах по сравнению с общей выборкой по формуле (2) по двум методам приведены в табл. 2 и 3. Нагрузка факторами риска, рассчитанная с помощью логистического регрессионного анализа, достигает в общей выборке 71,8, у мужчин – 72,0, у женщин – 71,7. С возрастом данная нагрузка закономерно увеличивается с 49,6 до 93,8 у мужчин и с 45,6 до 97,0 у женщин.

Аналогичные закономерности наблюдаются по нагрузке факторами риска, рассчитанной с помощью деревьев классификации. В общей выборке нагрузка составляет 19 459,6, у мужчин – 18 831,7, у женщин – 19 947,0. С возрастом нагрузка увеличивается с 10 570,0 до 28 512,9 у мужчин и с 10 250,0 до 30 457,8 у женщин.

Построенные с помощью линейного регрессионного анализа математические модели показали, что нагрузки факторами риска объясняют на 62,6 % (логистическая регрессия) и на 71,7 % (деревья классификации) частоту ИБС в половозрастных группах. Полученные по данным моделям В-коэффициенты регрессии использовались для перевода разницы нагрузок факторами риска в значения популяционного риска по формуле (3).

При использовании логистического регрессионного анализа В-коэффициент равен 0,4345, при использовании деревьев классификации – 0,0012.

Дополнительный к популяционному риск ИБС, обусловленный факторами риска, рассчитанный по формуле (3), представлен на рис. 1 ( $a$  – регрессионный анализ;  $b$  – деревья классификации). По обоим методам дополнительный риск в младших возрастных группах (до 46 лет) ниже популяционного от –2 до –11 %, после 45 лет достигает значений выше популяционных от 0,5 до 13,0 %. Увеличение дополнительного риска ИБС в возрастных группах практически линейное, с небольшими откло-

Таблица 2

Распространенность факторов риска, ИБС и нагрузка факторами риска в возрастных группах мужчин и в общей выборке

Фактор риска	Возрастная группа, лет (n)								Все мужчины, (700)	Вся выборка
	до 30 (86)	31–35 (85)	36–40 (88)	41–45 (65)	46–50 (98)	51–55 (94)	56–60 (107)	61–65 (77)		
Гипертензия, %	20,9	37,6	44,3	56,9	51,0	69,1	61,7	71,4	51,7	43,3
Гипертриглицеридемия, %	14,1	17,6	21,8	24,6	25,5	25,5	27,4	23,7	22,7	20,3
Гипергликемия, %	8,2	7,1	11,5	7,7	17,3	28,7	22,6	43,4	18,5	17,2
Ожирение, %	12,9	22,3	21,6	35,4	36,7	35,5	34,6	37,7	29,7	35,2
Сахарный диабет, %	1,2	0,0	0,0	0,0	4,1	5,3	4,7	14,7	3,7	3,9
Курение, %	47,7	47,1	46,6	47,7	50,0	48,9	43,0	33,8	45,7	30,5
Стресс, %	10,5	15,3	12,5	9,2	13,3	11,7	14,0	16,9	13,0	22,6
Депрессия, %	8,1	11,8	10,2	9,2	13,3	13,8	16,8	27,3	13,9	19,0
Образование, %	43,0	51,8	61,4	55,4	70,4	71,3	63,6	63,6	60,6	60,7
Работа, %	9,3	9,4	9,1	7,7	16,3	26,6	32,1	64,9	22,0	25,4
ИБС, %	3,5	3,5	5,9	9,2	10,3	22,3	21,5	30,3	13,5	16,8
Нагрузка 1	49,6	59,9	63,3	66,2	78,2	82,4	79,9	93,8	72,0	71,8
Разница 1	-22,2	-11,9	-8,5	-5,6	6,4	10,6	8,1	22,0	0,2	0
Нагрузка 2	10570	13885,1	15345	16671,4	19941,2	22920,1	22014,8	28512,9	18831,7	19459,6
Разница 2	-8889,6	-5574,5	-4114,6	-2788,6	481,6	3460,5	2555,2	9053,3	-627,9	0,0

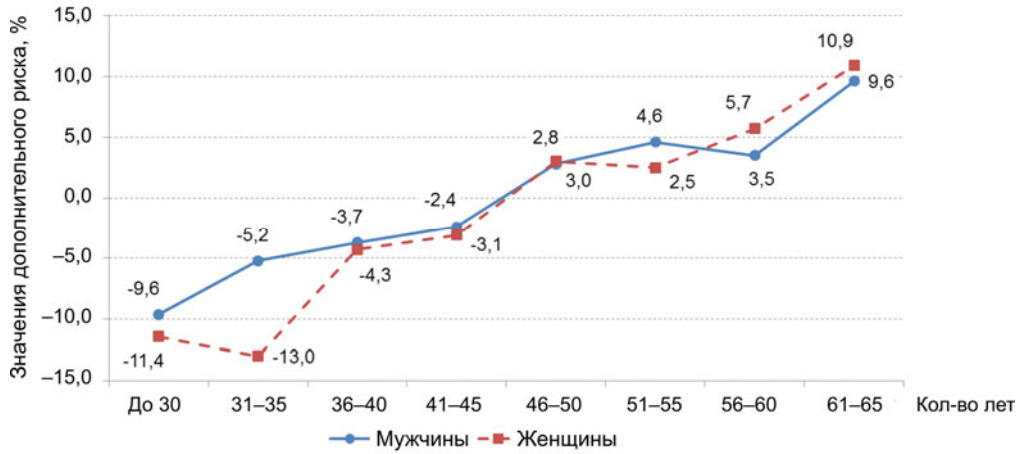
Таблица 3

Распространенность факторов риска, ИБС и нагрузка факторами риска в возрастных группах женщин

Факторы риска	Возрастная группа, лет (n)								Все женщины (928)
	до 30 (97)	31–35 (97)	36–40 (86)	41–45 (91)	46–50 (112)	51–55 (159)	56–60 (170)	61–65 (116)	
Гипертензия, %	10,3	8,2	18,6	30,8	42,0	50,3	52,3	57,8	37,2
Гипертриглицеридемия, %	5,2	9,3	9,3	17,6	26,4	23,4	20,4	28,1	18,5
Гипергликемия, %	0,0	2,1	12,8	13,2	15,5	15,8	26,9	32,5	16,2
Ожирение, %	9,3	14,4	27,9	41,8	46,4	44,9	50,6	61,2	39,4
Сахарный диабет, %	0,0	0,0	1,2	2,2	2,7	5,1	6,5	10,6	4,0
Курение, %	26,8	29,9	29,1	26,4	17,9	17,0	10,0	6,9	19,0
Стресс, %	26,8	26,8	29,4	27,8	36,6	28,9	31,4	29,3	29,8
Депрессия, %	11,3	7,2	17,4	17,6	23,2	28,9	31,8	31,9	22,8
Образование, %	48,5	33,0	59,3	51,6	71,4	63,5	69,4	76,7	60,9
Работа, %	20,6	12,4	24,4	11,0	14,3	18,9	44,7	63,7	27,9
ИБС, %	4,1	14,6	13,2	15,4	18,0	25,2	26,5	25,2	19,2
Нагрузка 1	45,6	41,8	62,0	64,6	78,8	77,6	85,0	97,0	71,7
Разница 1	-26,2	-30,0	-9,8	-7,2	7	5,8	13,2	25,2	-0,1
Нагрузка 2	10250	9211,9	15794,9	16971,6	21557,3	21452,6	25717,1	30457,8	19947
Разница 2	-9209,6	-10247,7	-3664,7	-2488	2097,7	1993	6257,5	10998,2	487,4

нениями: у женщин снижение риска в 31–35 и в 51–55 лет, у мужчин – в 56–60 лет по сравнению с предыдущей возрастной группой. При этом отклонения от линейности увеличения дополнительного риска ИБС с возрастом наблюдается по обоим методам – и при использовании логистического регрессионного анализа, и деревьев классификации.

Между значениями дополнительного популяционного риска, рассчитанного с помощью регрессионного анализа и с помощью деревьев классификации, определяется статистически значимая ( $p < 0,05$ ) сильная положительная связь, коэффициент корреляции равен 0,99. Разница дополнительного популяционного риска незначительная, колеблется от 0,1 до 2,3 %; в большин-

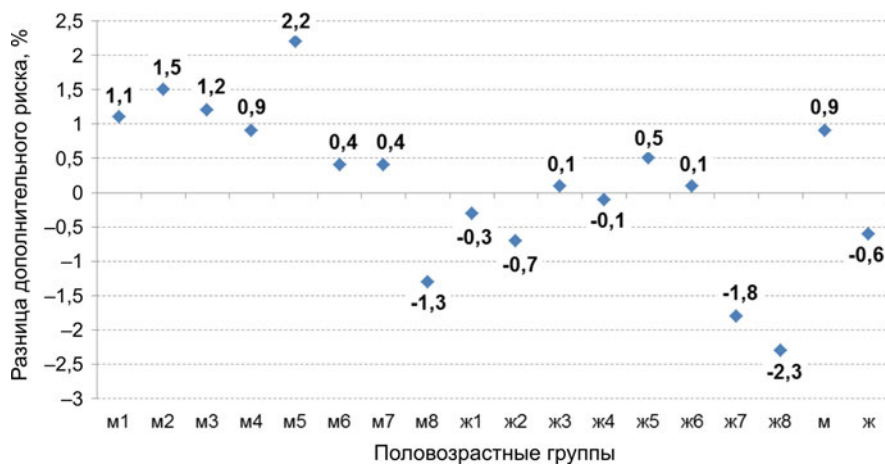


а



б

Рис. 1. Дополнительный к популяционному риск ИБС, обусловленный факторами риска: а – регрессионный анализ; б – деревья классификации



Примечание: м – мужчины, ж – женщины, 1 – до 30 лет, 2 – 31–35 лет, 3 – 36–40 лет, 4 – 41–45 лет, 5 – 46–50 лет, 6 – 51–55 лет, 7 – 56–60 лет, 8 – 61–65 лет.

Рис. 2. Разница дополнительного популяционного риска, рассчитанного по двум методам статистического анализа

стве половозрастных групп разница не превышает 1,5 % (рис. 2). Максимальные различия отмечаются среди женщин в возрасте 61–65 лет, мужчин 46–50 лет и женщин 56–60 лет.

**Выводы.** Таким образом, результаты исследования свидетельствуют о том, что оба используемых метода статистической обработки данных позволяют проводить комплексную оценку факторов риска ИБС на достаточно высоком прогностическом уровне. Прогностическая значимость математической модели при использовании логистического регрессионного анализа составляет 83,8 %, с помощью деревьев классификации – 71,9 %. В то же время два используемых метода статистического анализа дают разные значения вклада изучаемых факторов риска в показатели распространенности ИБС, что следует из отсутствия статистически значимой связи между значениями В-коэффициентов и рангами значимости факторов риска.

Полученные данные позволяют рассчитывать нагрузку факторами рисками. При этом по обоим методам (регрессионный анализ и деревья классификации) разница нагрузки в половозрастных группах по сравнению с общей выборкой изменяется с отрицательных значений в возрастных группах до 45 лет включительно

до положительных значений у лиц более старшего возраста. Аналогичным образом изменяется дополнительный к популяционному риск ИБС, обусловленный нагрузкой факторами риска. Увеличение дополнительного риска ИБС в возрастных группах по обоим методам практически линейное, с небольшими отклонениями. Если у лиц обоего пола до 30 лет дополнительный риск ИБС ниже популяционного на 9,6–11,4 %, то к 50 годам он начинает превышать популяционный (на 0,6–2,8 %), а к 65 годам превышение достигает максимальных значений (на 9,6–13,2 %).

Несмотря на то что разные методы статистического анализа дают разные значения вклада изучаемых факторов риска в показатели распространенности ИБС, между значениями дополнительного популяционного риска, рассчитанного с помощью регрессионного анализа и с помощью деревьев классификации, связь статистически значимая сильная. Разница дополнительного популяционного риска, рассчитанного по разным методам статистического анализа, незначительная, как правило, не превышает 1,5 %. Следовательно, оба метода показывают схожие результаты и могут в равной степени использоваться при расчете популяционного риска ИБС.

### Список литературы

1. Батюшин М.М. Модернизация шкалы SCORE оценки десятилетнего риска сердечно-сосудистой смертности // Российский кардиологический журнал. – 2005. – № 6. – С. 40–44.
2. Значимость мультифокального атеросклероза для модификации шкалы отдаленного риска смертности GRACE у больных острым коронарным синдромом с подъемом сегмента ST / М.В. Зыков, Д.С. Зыкова, В.В. Кашталап, Т.Б. Печерина, О.Л. Барбараш // Атеросклероз. – 2012. – Т. 8, № 1. – С. 14–20.
3. Концевая А.В., Суворова Е.И., Худяков М.Б. Экономическая эффективность ренальной денервации у пациентов с резистентной артериальной гипертензией: результаты марковского моделирования // Кардиология. – 2014. – Т. 54, № 1. – С. 41–47.
4. Концевая А.В., Шальнова С.А. Популяционные модели прогнозирования сердечно-сосудистого риска: целесообразность моделирования и аналитический обзор существующих моделей // Кардиоваскулярная терапия и профилактика. – 2015. – Т. 14, № 6. – С. 54–58.
5. Максимов С.А., Индукаева Е.В., Артамонова Г.В. Интегральная оценка риска ишемической болезни сердца в эпидемиологических исследованиях (ЭССЕ-РФ в Кемеровской области). Сообщение I. Возрастно-половые детерминанты // Профилактическая медицина. – 2015. – № 6. – С. 34–39.
6. Максимов С.А., Скрипченко А.Е., Артамонова Г.В. Интегральная оценка факторов риска профессиональной обусловленности артериальной гипертензии // Российский кардиологический журнал. – 2015. – Т. 120, № 4. – С. 38–42.
7. Максимов С.А., Табакаев М.В., Артамонова Г.В. Интегральная оценка риска ишемической болезни сердца в эпидемиологических исследованиях (ЭССЕ-РФ в Кемеровской области). Сообщение II. Социально-экономические детерминанты // Профилактическая медицина. – 2016. – Т. 19, № 1. – С. 24–29.
8. Мамедов М.Н., Чепурина Н.А. Суммарный сердечно-сосудистый риск: от теории к практике: пособие для врачей / под ред. акад. РАМН Р.Г. Оганова. – М., 2007. – 23 с.
9. Мнение врачей о роли отдельных факторов смертности от болезней системы кровообращения в регионах Российской Федерации / С.А. Бойцов, М.А. Ватолина, И.В. Самородская, О.Л. Барбараш, О.А. Овчаренко, Н.В. Кондрикова // Комплексные проблемы сердечно-сосудистых заболеваний. – 2015. – № 4. – С. 53–60.

10. Новый способ оценки индивидуального сердечно-сосудистого суммарного риска для населения России / Р.Г. Оганов, С.А. Шальнова, А.М. Калинина, А.Д. Деев, О.С. Глазачев, Е.И. Гусев, И.А. Беляева, А.М. Сударев // Кардиология. – 2008. – № 5. – С. 85–89.
11. Assmann G. Assessment of cardiovascular risk – PROCAM and new algorithms // Biomed. Tech. (Berl). – 2005. – Vol. 50, № 7–8. – P. 227–232.
12. Cardiovascular risk and cardiometabolic risk: an epidemiological evaluation / D. Vanuzzo, L. Pilotto, R. Miolo, S. Pirelli // G. Ital. Cardiol. (Rome). – 2008. – Vol. 9, № 4. – P. 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, № 2. – P. 91–99.
14. Screening for cardiovascular risk in asymptomatic patients / J.S. Berger, C.O. Jordan, D. Lloyd-Jones, R.S. Blumenthal // J. Am. Coll. Cardiol. – 2010. – Vol. 55, № 12. – P. 1169–1177.
15. The impact of prevention on reducing the burden of cardiovascular disease / R. Kahn, R.M. Robertson, R., Smith D. Eddy // Circulation. – 2008. – Vol. 118. – P. 576–585.

*Максимов С.А., Цыганкова Д.П., Артамонова Г.В. Применение регрессионного анализа и деревьев классификации для расчета дополнительного популяционного риска ишемической болезни сердца // Анализ риска здоровью. – 2017. – № 3. – С. 31–39. DOI: 10.21668/health.risk/2017.3.04*

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.

---

© 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).



## 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. Zikov M.V., Zykova 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 analiticheskiy 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 kardiologicheskii 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 smertnosti 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]. *Kompleksnye problemy serdechno-sosudistykh 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. *Biomed. 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

Получена: 05.06.2017

Принята: 20.09.2017

Опубликована: 30.09.2017