

УДК 57.087.1

## ВОЗРАСТЗАВИСИМЫЙ ПОДХОД К ПОИСКУ ГЕНЕТИЧЕСКИХ ВАРИАНТОВ, СВЯЗАННЫХ С РИСКОМ ИНФАРКТА МИОКАРДА

© 2020 г. Г. Ж. Осмак<sup>a, b, \*</sup>, А. Р. Сидько<sup>a</sup>, И. С. Киселев<sup>a, b</sup>, О. О. Фаворова<sup>a, b</sup>

<sup>a</sup>Национальный медицинский исследовательский центр кардиологии Министерства здравоохранения Российской Федерации, Москва, 121552 Россия

<sup>b</sup>Российский национальный исследовательский медицинский университет им. Н.И. Пирогова Министерства здравоохранения Российской Федерации, Москва, 117997 Россия

\*e-mail: german.osmak@gmail.com

Поступила в редакцию 12.03.2020 г.

После доработки 12.03.2020 г.

Принята к публикации 20.03.2020 г.

Инфаркт миокарда (ИМ) – одно из наиболее распространенных проявлений старения сердечно-сосудистой системы, нередко оканчивающееся летальным исходом. Подавляющее число исследований генетической предрасположенности к возрастзависимым заболеваниям проводится с использованием дизайна “случай–контроль”. Однако его применение связано с рядом сложностей, большинство из которых возникает на этапе формирования контрольной группы из условно здоровых индивидов. В работе проведен анализ функций дожития у носителей альтернативных полиморфных вариантов 18 генов, ассоциация которых с ИМ исследована нами ранее методом “случай–контроль”, и дана оценка величины смещения возраста дебюта заболевания в зависимости от индивидуальных вариаций генома. Выявлены ассоциированные с возрастом дебюта ИМ варианты риска rs2430561\*А гена *IFNG* (HR = 1.3,  $p = 0.043$ ), rs1799889\*5 гена *PAI-1* (HR = 1.3,  $p = 0.039$ ), rs1800896\*GG гена *IL10* (HR = 1.5,  $p = 0.0048$ ), rs1800471\*С гена *TGFBI* (HR = 1.5,  $p = 0.043$ ) и rs11614913\*ТТ гена *MIR196A2* (HR = 1.5,  $p = 0.035$ ), носительство которых способствует развитию заболевания на 3–6 лет раньше в сравнении с носителями альтернативных вариантов. Результаты исследования сопоставлены с ранее полученными на той же выборке оценками ассоциаций методом “случай–контроль”. Оказалось, что по большей части оценки, получаемые двумя методами, не совпадают, однако возрастзависимый подход опирается на меньшее число допущений, которые могут быть проверены, что, на наш взгляд, делает этот подход более перспективным по сравнению с подходом “случай–контроль”.

**Ключевые слова:** возрастзависимые заболевания, инфаркт миокарда, анализ дожития, машинное обучение, однонуклеотидные полиморфизмы

**DOI:** 10.31857/S0026898420040138

### ВВЕДЕНИЕ

Инфаркт миокарда (ИМ) – одна из наиболее тяжелых форм ишемической болезни сердца (ИБС) и главная причина смертности от неинфекционных заболеваний в мире [1]. ИМ в большинстве случаев возникает как осложнение прогрессирующего атеросклероза, который по своей этиологии принадлежит к многофакторным полигенным заболеваниям [2]. Риск как атеросклероза, так и ИМ повышается с возрастом, вследствие чего их относят к возрастзависимым заболеваниям [3]. Анализ генетической архитектуры атеросклероза и его частых осложнений показал, что от 40 до 60% предрасположенности к разви-

тию этих состояний определяется отягощенной наследственностью [4], однако полногеномные ассоциативные исследования полиморфных вариантов объясняют менее 20% наследуемости [5, 6].

Подавляющее число исследований генетической предрасположенности к ИМ проводится с использованием дизайна “случай–контроль”. Однако его применение связано с рядом сложностей, большинство из которых возникает на этапе формирования контрольной группы из условно здоровых индивидов [7].

В то же время исследование предрасположенности к возрастзависимым заболеваниям можно проводить путем оценки времени до наступления

Сокращения: ИБС – ишемическая болезнь сердца; ИМ – инфаркт миокарда; HR (hazard ratio) – отношение рисков; OR (odds ratio) – отношение шансов; 95% CI (95% confidence interval) – 95%-ный доверительный интервал.

события (time-to-event (survival) data analysis) [8]. Применительно к генетическим исследованиям исходят из допущения, что индивидуальные генетические вариации задают начальный уровень риска возрастзависимых заболеваний (в частности ИМ), определяющий их последующую динамику. В таком случае не требуется сравнения с контрольной группой, а в качестве зависимой (целевой) переменной принимается время от рождения индивида до дебюта заболевания. Очевидно, что при рассмотрении таких заболеваний возрастание числа факторов риска будет смещать возраст их дебюта в сторону уменьшения, что активно используется при генетических исследованиях старения и долгожительства [9], а также при оценках риска сердечно-сосудистых событий с помощью прогностических шкал.

Ранее методом “случай–контроль” мы изучали связь с риском развития ИМ полиморфных вариантов 18 генов и выявили значимую ассоциацию с ИМ для шести из них: *TGFB1*, *FGB*, *CRP*, *IFNG*, *PTGS1* и *MIR196A2* [10, 11]. В представленной работе мы оценили смещение возраста дебюта ИМ в зависимости от носительства вариантов исследуемых генов в выборке больных ИМ, сформированной в цитированных исследованиях. Полученные результаты сопоставлены с предшествующими данными [10, 11], оценена их согласованность, проанализированы достоинства и недостатки двух подходов при использовании в качестве инструментов оценки риска развития возрастзависимых заболеваний.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**В работе использовали** экспериментальные данные по геномному типированию полиморфных участков генов *CRP* (rs1130864), *IL10* (rs1800896), *CTLA4* (rs231775), *CCR5* (rs333), *FGA* (rs6050), *FGB* (rs1800788, rs1800790), *PDE4D* (rs152312), *IL4* (rs2243250), *TNF* (rs1800629), *LTA* (rs909253), *IL6* (rs1800795), *PAI-1* (rs1799889), *PTGS1* (rs3842787), *IFNG* (rs2430561), *TGFB1* (rs1800469, rs1982073, rs1800471), *MIR146A* (rs2910164), *MIR196A2* (rs11614913) и *MIR499A* (rs3746444), полученные нами ранее на выборке из 289 пациентов (193 мужчин и 96 женщин) с диагностированным первым “острым инфарктом миокарда” [10, 11].

**Поиск генетических вариантов**, от которых может зависеть возраст дебюта ИМ, осуществляли посредством сравнения функций выживаемости Каплана–Майера, рассчитывая отношение рисков (hazard ratio, HR) дебюта ИМ и его 95%-ный доверительный интервал (confidence interval, 95% CI) у групп носителей альтернативных аллелей или генотипов в модели пропорциональных рисков Кокса. Для целей настоящего исследования отсчет времени при построении кривых выживаемости ве-

ли со дня рождения пациента, а в качестве конечной точки принимали первый случай ИМ.

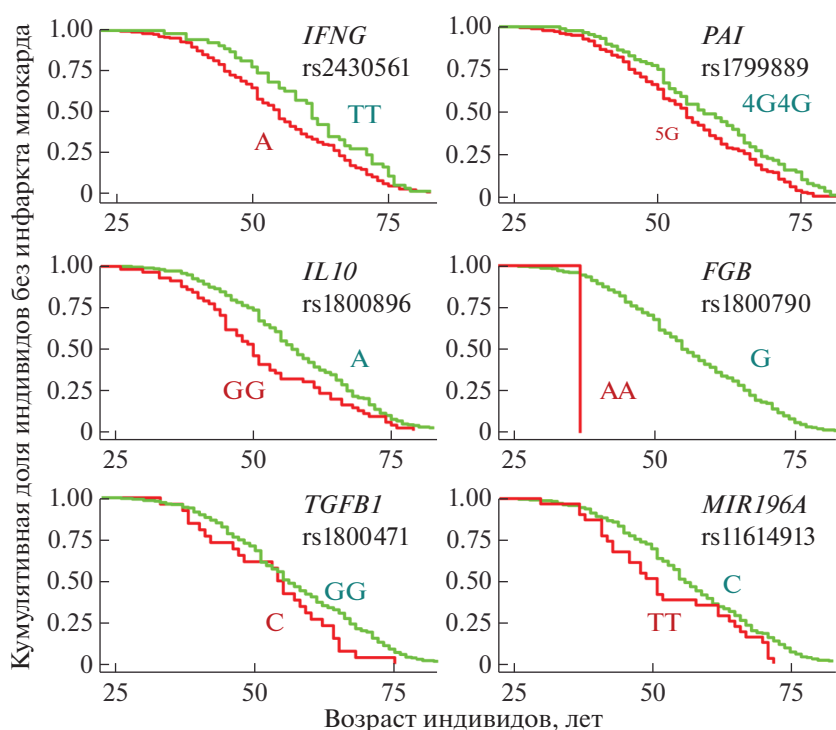
**Статистическую обработку данных проводили** при помощи средств языка программирования R. Исследование функций выживаемости, включая проверку пропорциональности рисков при помощи теста остатков Шенфильда, проводили с использованием пакета “survival” в соответствии с документацией (<https://www.rdocumentation.org/packages/survival/versions/3.1-11>). Построение регрессионной модели для предсказания возраста ИМ осуществляли при помощи библиотеки scikit-learn для языка Python 3. Для отбора наиболее важных признаков использовали Lasso-регрессию с регуляризацией по L1-норме [12]. Подбор величины штрафа осуществляли прямым перебором, реализованным в классе `sklearn.linear_model.LassoCV` по сетке параметров с 10-кратной кросс-валидацией для повышения устойчивости модели и снижения вероятности переобучения. Разбиение выборки на обучающую и тестовую проводили в пропорции 4 : 1 соответственно, используя функцию `train_test_split` той же библиотеки.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Анализируя кривые выживаемости для больных ИМ, носителей альтернативных вариантов 18 исследуемых генов, наблюдали значимую ассоциацию с возрастом дебюта ИМ носительства вариантов генов *IFNG*, *PAI-1*, *IL10*, *FGB*, *TGFB1* и *MIR196A* ( $p$ -value < 0.05, 95% CI не пересекает единицу). На рис. 1 представлено смещение кривых выживаемости у носителей различных вариантов этих генов. Проверка условия пропорциональности рисков при помощи теста остатков Шенфельда показала, что оценки HR для вариантов этих генов постоянны и значимо не зависят от возраста индивидов ( $p$  < 0.05). Обращает на себя внимание резкое смещение функции выживаемости в случае носителей генотипа AA rs1800790 гена *FGB*, у всех из них первый ИМ произошел в возрасте до 40 лет.

В табл. 1 приведены значения  $p$ -value и HR (95% CI) для тех аллелей или генотипов приведенных на рис. 1 генов, носительство которых значимо снижает возраст дебюта ИМ (т.е. вариантов риска ИМ). Значения HR для всех из них, за исключением генотипа AA гена *FGB*, лежат в довольно узком диапазоне: от 1.3 до 1.5. Для *FGB*\*AA значение HR составляет 11, при том, что лежит в достаточно узком диапазоне 95% CI: 9.15–13.3.

Для оценки числа лет, на которые смещается возраст дебюта ИМ в зависимости от генотипа, мы сформулировали задачу предсказания возраста дебюта ИМ при носительстве тех или иных полиморфных вариантов. Такая постановка вопро-



**Рис. 1.** Кривые Каплана–Майера возраста дебюта ИМ для носителей альтернативных полиморфных вариантов указанных генов (представлены данные только для значимых ассоциаций). В гене *PAI-1* исследован инсерционно-делеционный полиморфизм 4G/5G; остальные полиморфизмы – однонуклеотидные замены. Кривые для аллелей/генотипов, повышающих риск раннего ИМ, выделены красным цветом, а для протективных вариантов – зеленым.

са позволяет перейти к моделям на основе линейной регрессии. В этом случае коэффициенты уравнения регрессии при полиморфных вариантах будут соответствовать величине смещения возраста дебюта ИМ у носителей аллелей или генотипов риска относительно среднего возраста дебюта в исследуемой выборке.

Модель обучали, минимизируя квадратичную функцию потерь с введением штрафа по L1-норме для прореживания признаков полиморфных вариантов. В результате получены оценки смещения возраста дебюта ИМ в сторону

более раннего его развития в диапазоне от 3.08 лет для носителей rs2430561\*A гена *IFNG* до 5.77 лет для носителей rs1800896\*GG гена *IL10* (табл. 1). Исключение составляет ген *FGB*, для которого оценка смещения возраста составила 0 лет.

Совокупность представленных результатов свидетельствует об ассоциации вариантов генов *IFNG*, *PAI-1*, *IL10*, *TGFBI* и *MIR196A* с риском ИМ. Использование регуляризации по L1-норме позволило исключить из числа значимых для оценки возраста дебюта ИМ ген *FGB*, что согласуется с отмеченными выше отклонениями кривых

**Таблица 1.** Полиморфные варианты генов, значимо ассоциированные с более ранним возрастом дебюта ИМ

Ген, идентификационный номер полиморфизма	Аллель/генотип риска	<i>p</i> -Value	HR (95% CI)	Величина смещения возраста ИМ (лет)
<i>IFNG</i> , rs2430561	A	0.043	1.3 (1.06–1.66)	3.08
<i>PAI-1</i> , rs1799889	5G*	0.039	1.3 (1.05–1.56)	3.02
<i>IL10</i> , rs1800896	GG	0.0048	1.5 (1.23–1.81)	5.77
<i>TGFBI</i> , rs1800471	C	0.043	1.5 (1.11–1.93)	3.42
<i>MIR196A</i> , rs11614913	TT	0.035	1.5 (1.12–1.88)	3.34
<i>FGB</i> , rs1800790	AA	0.022	11.0 (9.15–13.3)	0

\* Аллель инсерционно-делеционного полиморфизма.

выживаемости и значениями HR для этого гена относительно других генов.

### ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Нами исследована ассоциация полиморфных вариантов генов *CRP* (rs1130864), *IL10* (rs1800896), *CTLA4* (rs231775), *CCR5* (rs333), *FGA* (rs6050), *FGB* (rs1800788, rs1800790), *PDE4D* (rs152312), *IL4* (rs2243250), *TNF* (rs1800629), *LTA* (rs909253), *IL6* (rs1800795), *PAI-1* (rs1799889), *PTGS1* (rs3842787), *IFNG* (rs2430561), *TGFBI* (rs1800469, rs1982073, rs1800471), *MIR146A* (rs2910164), *MIR196A2* (rs11614913) и *MIR499A* (rs3746444) с возрастом дебюта ИМ. Выявлены ассоциированные с возрастом дебюта ИМ рисковые варианты генов *IFNG*, *PAI-1*, *IL10*, *TGFBI* и *MIR196A*, носительство которых способствует его развитию на 3–6 лет раньше в сравнении с носителями альтернативных вариантов.

Неоднозначные результаты, полученные нами для носителей генотипа rs1800790\*AA гена *FGB*, привели к исключению этого гена из рассмотрения. Скорее всего, эти результаты связаны с малым числом ( $n = 3$ ) таких носителей и плохо отражают свойства генеральной совокупности больных ИМ.

Сравнивая результаты настоящего исследования с ранее полученными для той же выборки больных ИМ методом “случай–контроль” [10, 11] (табл. 2), можно видеть, что качественно совпадают результаты только для варианта гена *IFNG*. В то же время при анализе методом “случай–контроль” [10] наблюдали ассоциацию с ИМ других полиморфных участков генов *TGFBI* (rs1982073) и *FGB* (rs1800788), а также варианта гена *CRP* rs1130864.

Таким образом, результаты, полученные при исследовании ассоциации генетических вариантов с возрастом дебюта ИМ, отличаются от результатов, полученных методом “случай–контроль”. В рамках проведенного исследования мы не можем проанализировать все возможные причины наблюдаемых расхождений, но некоторые из них следует отметить.

При дизайне “случай–контроль” для исследования многофакторных возрастзависимых заболеваний невозможно гарантировать, что они не разовьются у индивидов контрольной группы спустя некоторое время после ее формирования. Искажение результатов может происходить вследствие вмешательства негенетических (посторонних) факторов, которые не были учтены при формировании групп сравнения, – так называемых конфаунд-эффектов [13]. Смещение оценок может происходить также из-за несоответствия между поставленными задачами, способом формирования контрольной группы (эксклюзивный, конкурентный, инклюзивный), распространенностью заболевания и особенностями исследуемых факторов риска. Все это неизбежно приводит к невыполнению статистических условий и некорректной трактовке получаемых результатов [14]. Выше перечисленные факторы могут приводить к получению результатов, не в полной мере отражающих реальное положение вещей, и при их интерпретации следует соблюдать осторожность.

В свою очередь, возрастзависимый подход к исследованию генетической предрасположенности к ИМ основан на оценке возраста дожития до первого ИМ и предполагает, что эффекты тех или иных генетических вариантов на риск возникновения ИМ постоянны в разных возрастах и накапливаются кумулятивно по мере старения, что и определяет возрастзависимый характер заболевания. В случае же нелинейных эффектов будет нарушаться пропорциональность рисков – важное условие для применения регрессионной модели Кокса, что приведет к ложноположительному или ложноотрицательному результату анализа. В проведенном исследовании с помощью теста остатков Шенфельда проверена пропорциональность рисков и показано, что это условие соблюдалось, о чем свидетельствует также отсутствие пересечения кривых дожития, построенных для носителей альтернативных генотипов (см. рис. 1).

Таким образом, при исследовании генетической архитектуры ИМ с использованием возрастзависимого

**Таблица 2.** Сравнение результатов представленного исследования с полученными ранее методом “случай–контроль” для той же выборки больных ИМ

Носительство аллеля или генотипа	Отношение рисков (HR) по результатам проведенного исследования	Отношение шансов (OR) по результатам исследования методом “случай–контроль” [10, 11]
<i>IFNG</i> rs2430561*A <sup>a</sup>	1.3	2.9
<i>MIR196A2</i> rs11614913*TT	1.5	0.5
<i>IL10</i> rs1800896*GG	1.5	Н.З. <sup>b</sup>
<i>PAI-1</i> rs1799889*5G	1.3	Н.З.
<i>TGFBI</i> rs1800471*C	1.5	Н.З.

<sup>a</sup>В составе эпистатического сочетания. <sup>b</sup>Н.З. – незначимо ( $p$ -value > 0.05).

мого подхода вводится меньше допущений, чем при подходе “случай–контроль”, причем все они могут быть проверены. Это и определяет преимущество возрастзависимого подхода перед дизайном типа “случай–контроль”. Неслучайно методы диагностики, основанные именно на возрастзависимых оценках, получили широкое распространение в практической медицине в составе шкал оценок риска неблагоприятных сердечно-сосудистых событий (шкала SCORE [15], Фрамингемская шкала [16], шкала ASCVD [17] и др.).

Исследование поддержано грантом РФФИ № 19-315-80019.

Настоящая статья не содержит каких-либо исследований с участием людей или животных в качестве объектов исследований.

Авторы сообщают об отсутствии конфликта интересов.

### СПИСОК ЛИТЕРАТУРЫ

1. Roberts R. (2014) Genetics of coronary artery disease. *Circ. Res.* **114**(12), 1890–1903.
2. Wu M.-Y., Li C.-J., Hou M.-F., Chu P.-Y. (2017) New insights into the role of inflammation in the pathogenesis of atherosclerosis. *Int. J. Mol. Sci.* **18**(10), pii: E2034.
3. Head T., Daunert S., Goldschmidt-Clermont P.J. (2017) The aging risk and atherosclerosis: a fresh look at arterial homeostasis. *Front. Genet.* **8**, 216.
4. (2002) The Genetic Basis of Common Diseases. Eds King R.A., Rotter J.I., Motulsky A.G. Oxford: Oxford University Press, vol. 44.
5. Hartiala J., Schwartzman W.S., Gabbay J., Ghazalpour A., Bennett B.J., Allayee H. (2017) The genetic architecture of coronary artery disease: current knowledge and future opportunities. *Curr. Atheroscler. Rep.* **19**(2), 6.
6. Marian A.J. (2012) The enigma of genetics etiology of atherosclerosis in the post-GWAS era. *Curr. Atheroscler. Rep.* **14**(4), 295–299.
7. Tenny S., Hoffman M.R. (2020) Case Control Studies. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing.
8. Kleinbaum D.G., Klein M. (2010) Survival Analysis, vol. 3. Springer.
9. Slagboom P.E., van den Berg N., Deelen J. (2018) Phenome and genome based studies into human ageing and longevity: an overview. *Biochim. Biophys. Acta Mol. Basis Dis.* **1864**(9), 2742–2751.
10. Barsova R.M., Lvovs D., Titov B.V., Matveeva N.A., Shakhnovich R.M., Sukhinina T.S., Kukava N.G., Ruda M.Ya., Karamova I.M., Nasibullin T.R., Mustafina O.E., Osmak G.J., Tsareva E.Yu., Kulakova O.G., Favorov A.V., Favorova O.O. (2015) Variants of the coagulation and inflammation genes are replicably associated with myocardial infarction and epistatically interact in Russians. *PLoS One.* **10**(12), e0144190. <https://doi.org/10.1371/journal.pone.0144190>
11. Осьмак Г.Ж., Матвеева Н.А., Титов Б.В., Фаворова О.О. (2018) Связь полиморфизма гена *MIR196A2* с инфарктом миокарда и возможное вовлечение микроРНК miR-196a2 в сигнальные пути, участвующие в формировании патологического фенотипа. *Молекуляр. биология.* **52**(6), 1006–1013.
12. Tibshirani R. (1996) Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**(1), 267–288.
13. Beaglehole R., Bonita R., Kjellström T. (1993) Basic Epidemiology. World Health Organization, Geneva.
14. Rodrigues L., Kirkwood B.R. (1990) Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Int. J. Epidemiol.* **19**(1), 205–213.
15. Conroy R.M., Pyörälä K., Fitzgerald A.P., Sans S., Menotti A., Backer G.D., Bacquer D.D., Ducimetière P., Jousilahti P., Keil U., Njølstad I., Oganov R.G., Thomsen T., Tunstall-Pedoe H., Tverdal A., Wedel H., Whincup P., Wilhelmsen L., Graham I.M. (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur. Heart J.* **24**(11), 987–1003.
16. Wilson P.W., D’Agostino R.B., Levy D., Belanger A.M., Silbershatz H., Kannel W.B. (1998) Prediction of coronary heart disease using risk factor categories. *Circulation.* **97**(18), 1837–1847.
17. Goff D.C., Lloyd-Jones D.M., Bennett G., Coady S., D’agostino R.B., Gibbons R., Greenland P., Lackland D.T., Levy D., O’donnell C.J. (2014) 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Part B. *J. Am. Coll. Cardiol.* **63**(25), 2935–2959.

## AGE-DEPENDENT APPROACH TO SEARCHING FOR GENETIC VARIANTS ASSOCIATED WITH MYOCARDIAL INFARCTION

G. J. Osmak<sup>1, 2, \*</sup>, A. R. Sidko<sup>1</sup>, I. S. Kiselev<sup>1, 2</sup>, and O. O. Favorova<sup>1, 2</sup>

<sup>1</sup>National Medical Research Center for Cardiology, Moscow, 121552 Russia

<sup>2</sup>Pirogov Russian National Research Medical University, Moscow, 117997 Russia

\*e-mail: german.osmak@gmail.com

Myocardial infarction (MI) is one of the most common manifestations of the cardiovascular system aging, which is often fatal. The vast majority of studies of the genetic susceptibility to age-dependent diseases are

carried out using the “case–control” study design. However, its use is followed by a number of difficulties, most of which arise at the stage of establishing the control group of relatively healthy individuals. In this work, we analyzed the survival functions for carriers of alternative polymorphic variants of eighteen genes, which were previously tested by us for the association with MI using the “case–control” approach, and we further estimated the magnitude of the shift in the age of the disease onset depending on individual variations of the genome. We identified risk variants associated with the age of MI: rs2430561\*A of *IFNG* gene (HR = 1.3,  $p = 0.043$ ), rs1799889\*5 of *PAI-1* gene (HR = 1.3,  $p = 0.039$ ), rs1800896\*GG of *IL10* gene (HR = 1.5,  $p = 0.0048$ ), rs1800471\*C of *TGFB1* gene (HR = 1.5,  $p = 0.043$ ), and rs11614913\*TT of *MIR196A2* gene (HR = 1.5,  $p = 0.035$ ). Carriage of these variants leads to the development of the disease 3–6 years earlier in the comparison with carriers of alternative variants. Results of this study were compared with data on the association with MI previously obtained on the same sample using the “case–control” approach. It turned out that estimates based on two methods mostly do not coincide. However, the age-dependent approach relies on fewer assumptions that can be additionally verified. In our opinion, it makes this approach more promising relative to the “case–control” approach.

**Keywords:** age-dependent diseases, myocardial infarction, survival analysis, machine learning, single nucleotide polymorphisms