DOI: https://doi.org/10.63769/3034-7459-2026-2-2-99-112

CC BY 4.0

Кросс-секционное исследование пациентов в целях разработки верифицированных балльных шкал риска остеоартрита

И.Ю. Торшин[1], А.Н. Громов[1], И.А. Рейер[1], Б.И. Беликов[2], В.А. Семёнов[3], Т.Р. Гришина[2], О.А. Лиманова[2], Л.Э. Федотова[2], О.А. Громова[1, 2]

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия;
[2]Ивановский государственный медицинский университет, Иваново, Россия;
[3]Кемеровский государственный медицинский университет, Кемерово, Россия

Аннотация

Обоснование. Выявление пациентов с высоким риском развития заболеваний и своевременное назначение им фармаконутрицевтиков – насущная задача современной медицины. Для этой цели часто используются верифицированные балльные опросники, позволяющие врачу оценивать риски конкретного пациента и своевременно осуществлять профилактические и терапевтические (в том числе фармакотерапевтические) мероприятия. Современные подходы к разработке таких опросников основаны на комплексе методов, известных как интеллектуальный анализ данных.

Цель исследования – разработка и верификация балльных шкал риска остеоартрита (ОА) на основе данных крупномасштабных кросс-секционных исследований пациентов.

Методы. Проведен анализ 2 независимых выборок пациентов в возрасте 35–90 лет из Базы данных Института микроэлементов. Данные первого кросс-секционного исследования (n = 3440) включили информацию об анамнезе, симптоматике и фармакотерапии пациентов, в том числе данные пациентов с ОА (n = 107). Данные второго исследования (n = 1290) включили информацию преимущественно об анамнезе и симптоматике, включая пациентов с ОА (n = 450). Разработка и верификация балльных шкал осуществлены в контексте алгебраического подхода к распознаванию с использованием метода опорных функций.

Результаты. На основе алгебраического интеллектуального анализа данных разработаны две 100-балльные шкалы для оценки риска ОА. В результате анализа первого кросс-секционного исследования выборки разработана и верифицирована шкала из 27 пунктов, позволяющая выявлять пациентов с высоким риском ОА (чувствительность 88 %, специфичность 100 %). На основании данных второго кросс-секционного исследования разработана более простая шкала из 16 пунктов, позволяющая выявлять пациентов с высоким риском развития ОА (чувствительность 83 %, специфичность 74 %).

Заключение. Пациентам с неблагоприятными показателями разработанных бальных шкал рекомендуется заблаговременное назначение хондропротекторных препаратов на основе стандартизированных субстанций хондроитина сульфата и глюкозамина сульфата, что позволит существенно снизить прогрессирование и тяжесть течения заболевания.

Ключевые слова: прогнозирование; профилактика; топологический подход к распознаванию; искусственный интеллект; балльная шкала риска заболевания.

Как цитировать: Торшин И.Ю., Громов А.Н., Рейер И.А., Беликов Б.И., Семёнов В.А., Гришина Т.Р., Лиманова О.А., Федотова Л.Э., Громова О.А. Кросс-секционное исследование пациентов в целях разработки верифицированных балльных шкал риска остеоартрита // Фармакоинформатика и фармаконутрициология. Современный анализ данных 2026;2(2):99–112. DOI: https://doi.org/10.63769/3034-7459-2026-2-2-99-112

A cross-sectional study of patients to develop verified osteoarthritis risk scores

Ivan Yu. Torshin[1], Andrey N. Gromov[1], Ivan A. Reyer[1], Bogdan I. Belikov[2], Vladimir A. Semenov[3], Tatiana R. Grishina[2], Olga A. Limanova[2], Lyubov E. Fedotova[2], Olga A. Gromova[1, 2]

[1]Computer Science and Control, Russian Academy of Sciences, Moscow, Russia;
[2]Ivanovo State Medical University, Ivanovo, Russia;
[3]Kemerovo State Medical University, Kemerovo, Russia

Abstract

***BACKGROUND*:** Identifying patients at high risk for developing diseases and promptly prescribing pharmaconutraceuticals is a pressing issue in modern medicine. Verified patient scoring questionnaires are often used for this purpose, allowing physicians to assess a specific patient's risk and implement timely preventive and therapeutic (including pharmacotherapeutic) measures. Modern approaches to developing such questionnaires are based on a set of methods known as data mining (DMA).

***AIM*:** To develop and validate osteoarthritis (OA) risk scoring scales based on data from large-scale cross-sectional patient studies.

***METHODS*:** An DMA-based analysis was conducted of two independent samples of patients aged 35–90 years from the Institute of Microelements Database (IMBD). Data from the first cross-sectional study (n = 3,440) included information on the patient's medical history, symptoms, and pharmacotherapy, including data from patients with OA (n = 107). Data from the second study (n = 1,290) included information primarily on medical history and symptoms, including patients with OA (n = 450). The scoring scales were developed and verified within the context of an algebraic approach to recognition using the support function method.

***RESULTS*:** Based on the algebraic DMA, two 100-point scales for OA risk assessment were developed. Based on the analysis of the first cross-sectional study sample, a 27-item scale was developed and verified, enabling the identification of patients at high risk of OA (sensitivity 88 %, specificity 100 %). Based on data from the second cross-sectional study, a simpler 16-item scale was developed, enabling the identification of patients at high risk of developing OA (sensitivity 83 %, specificity 74 %).

***CONCLUSION*:** For patients with unfavorable scores on the developed scoring scales, early administration of chondroprotective medications based on standardized chondroitin sulfate and glucosamine sulfate is recommended. This can significantly reduce the progression and severity of the disease.

Keywords: prognosis; prevention; topological approach to recognition; artificial intelligence; disease risk scoring system.

To cite this article: Torshin I.Yu., Gromov A.N., Reyer I.A., Belikov B.I., Semenov V.A., Grishina T.R., Limanova O.A., Fedotova L.E., Gromova O.A. A cross-sectional study of patients to develop verified osteoarthritis risk scores. Pharmacoinformatics and pharmaconutritiology. Current trends in data analysis 2026;2(2):99–112. DOI: https://doi.org/10.63769/3034-7459-2026-2-2-99-112

Обоснование

Важным направлением дальнейшего развития подходов к интеллектуальному анализу биомедицинских данных является разработка теоретического обоснования, алгоритмов и программных систем для построения и верификации балльных опросников состояния пациента. Такие опросники широчайшим образом используются практически во всех разделах клинической медицины и терапии (неврология, кардиология и др.) и позволяют врачу проводить количественную оценку состояния пациента с той или иной патологией.

Существенными недостатками подавляющего большинства имеющихся опросников являются:

- произвол в выборе рассматриваемых показателей состояния пациента (например, часто используются возраст и пол, которые зачастую малоинформативны для большинства хронических патологий);
- произвол в назначении весов выбранных показателей состояния пациента (как правило, вес назначается в соответствии с личным мнением соответствующего эксперта);
- произвол в процедурах верификации результатов применения опросника (зачастую опросники не верифицируются, т. е. не тестируются на независимой контрольной выборке пациентов и не проводится сравнение баллов по опроснику с какими-то объективно определяемыми показателями состояния пациента (например, с уровнями того или иного метаболита в крови и др.)).

В настоящей работе применяли комплекс проблемно-ориентированных теорий, разрабатываемых в на-

учной школе Ю.И. Журавлёва и К.В. Рудакова, для построения и верификации балльных опросников состояния пациента. В рамках предлагаемого подхода опросники строятся как аддитивные модели с корректорами (опорными функциями), позволяющие прогнозировать значения интересующей врача переменной (наличие того или иного диагноза, уровни определенных метаболитов, белков, ионов в крови и др.). Разработанные шкалы верифицируются в рамках кросс-валидационного вычислительного эксперимента.

Данный метод был приложен к разработке балльных шкал, которые позволили бы выявлять пациентов с высоким риском формирования и прогрессирования остеоартрита (ОА). Важной особенностью настоящего метода является то, что в качестве входной информации используются не только какие-то отдельные симптомы ОА, а весь комплекс данных о пациенте. Такой подход обоснован комплексной патофизиологией ОА, включающей определенные сочетания коморбидных диагнозов (желчнокаменная болезнь, хроническая ишемия мозга, атеросклероз, тромбофлебит), нарушения уровней микроэлементов, нуклеотидные варианты генетических полиморфизмов [1], влияние на формирование ОА назначения тех или иных лекарственных средств (например, при ишемической болезни сердца назначаются антигиперлипидемические препараты, диуретики, бета-блокаторы, антиаритмические препараты, которые потенциально могут влиять на риск ОА) [2].

Отметим, что такая комплексная информация практически никогда не собирается в ходе, так сказать, «стандартных» исследований ОА и других хронических патологий. Подобного рода исследования, как правило, отличаются выраженной однобокостью вследствие волюнтаризма «специалистов» и недостаточного понимания постгеномной парадигмы современных биомедицинских исследований, приверженностью к редукционистским моделям формирования заболеваний, равно как и вследствие прочих субъективных факторов.

Информация в Базе данных Института микроэлементов (ИМБД), собранная в ходе обследования нескольких тысяч пациентов в рамках исследовательских программ Московского сотрудничающего центра Института микроэлементов при ЮНЕСКО (http://trace-elements.ru), позволяет проводить кросс-секционные исследования, результатом которых является построение верифицированных балльных шкал. Такие клинико-эпидемиологические исследования позволят разрабатывать алгоритмы, которые могут с высокой аккуратностью выявлять пациентов с предрасположенностью к ОА в рамках доносологического скрининга. Для анализа комплексных взаимодействий между описаниями таких пациентов были использованы современные методы топологического анализа данных, разрабатываемые в рамках алгебраического подхода к распознаванию [3].

Методы

База данных

ИМБД содержит медицинскую информацию для нескольких тысяч пациентов, обследованных в рамках исследовательских программ [4]. Пополнение ИМБД осуществляется не только за счет проведения собственных исследований, но и за счет данных, содержащихся в таких базах, как NCBI, EUROCAT, GPRD и др. Для каждого пациента в ИМБД вводятся демографические параметры, род занятий, антропометрия, состояние сердечно-сосудистой системы, оценка физической активности, данные об употреблении алкоголя и курении табака, стандартный и биохимический анализы крови (в том числе уровни глюкозы, инсулина, гомоцистеина, С-пептида, гликированного гемоглобина, витаминов, магния, меди, цинка, железа, ферритина, С-реактивного белка, различных гормонов, омега-3-индекс и др.), подробный медицинский анамнез (в том числе кардиологический, эндокринологический, дерматологический, акушерский/гинекологический, урологический и др.), оценки психофизиологического состояния по различным шкалам (САН, SF-36 и др.), оценки потребления различных витаминов, макро- и микроэлементов по опросникам и дневникам диеты.

Для целей настоящего исследования важно, что в ИМДБ содержится не только подробная анамнестическая информация, но и сведения о принимаемых пациентом лекарственных препаратах (антибиотики, противовирусные, противогрибковые, противомалярийные, противотуберкулезные средства, противодиабетические, антигиперлипидемические, антиаритмические препараты, пероральные контрацептивы и др.).

В результате анализа ИМДБ были сформированы 2 выборки пациентов в возрасте 35–90 лет, на основе анализа которых были получены верифицированные шкалы оценки риска ОА. Первая выборка (выборка 1) включила данные о 2120 признаках для 3440 пациентов (в том числе 107 пациентов с ОА), у которых исследовано наличие ОА и была собрана информация об анамнезе, симптоматике и фармакотерапии. Вторая выборка (выборка 2) включила данные о 1839 признаках для 1290 пациентов (в том числе 450 пациентов с ОА), для которых была собрана информация преимущественно об анамнезе и симптоматике, эпизодически встречающейся у этих пациентов (не реже чем 1 раз в 6 мес).

Методы анализа данных

Для статистической обработки результатов исследования использовали критерии χ^2, Т-критерий Вилкоксона–Манна–Уитни и тест Стьюдента (программный пакет Statistica 10.0 и электронные таблицы Microsoft Excel). Для сравнения 2 групп с нормально распределенными числовыми данными использовали

t-критерий Стьюдента. Для сравнения числовых показателей, распределение которых отличалось от нормального, применяли критерий Вилкоксона. Анализ всей совокупности корреляций между показателями состояния пациентов выполняли с помощью современных методов интеллектуального анализа данных: метод анализа метрических сгущений в пространстве параметров, метод метрических карт [5], методы прогнозирования числовых таргетных переменных [6]. Математические детали использованных методов (в том числе сравнения с другими подходами и алгоритмами) приведены в цикле работ по метрическому анализу данных [3–6].

Результаты

В результате анализа 2 выборок ИМДБ были разработаны две 100-балльных шкалы, позволяющие выявлять пациентов с высоким риском ОА с приемлемыми чувствительностью (83–88 %) и специфичностью (74–100 %). Рассмотрим более подробно анализ выборки 1 и соответствующую балльную шкалу.

Выборка 1 (n = 3440) включила данные о 2120 показателях состояния здоровья пациентов, для которых была собрана весьма подробная информация об анамнезе (сердечно-сосудистом, респираторном, желудочно-кишечном, онкологическом, ревматологическом, эндокринологическом и др.), а также об эпизодической симптоматике и фармакотерапии. Методами топологического анализа данных был выделен 21 признак, которые, являясь наиболее информативными предикторами ОА, могут использоваться для построения балльных шкал. Эти признаки представлены на так называемой метрической карте на рис. 1.

Метрическая карта на рис. 1 наглядно иллюстрирует весь массив достоверных корреляций между ОА и предикторами ОА. Метрическая карта представляет каждый показатель точкой на плоскости так, что расстояние между каждой парой точек пропорционально степени взаимодействия (корреляции) между соответствующими параметрами. Из метрической карты очевидно, что предикторы ОА можно подразделить на 3 группы: 1) *анамнестические* (пиелонефрит, язва желудка, острая реакция на стресс, псориаз, заболевания щитовидной железы и др.); 2) *симптоматические* (общая слабость, сыпь на коже, боли в колене, в области шеи или поясницы); 3) *фармакотерапевтические* (прием антипсориатических, противомалярийных средств, эстрогенсодержащих препаратов, антиаритмических препаратов, кардиоселективных бета-адреноблокаторов, ингибиторов протонной помпы и др.).

Метрическая карта на рис. 1 позволила провести анализ сочетаний отдельных предикторов ОА, которые

Рис. 1. *Метрическая карта, отражающая ассоциации между остеоартритом и различными показателями состояния пациентов в выборке 1 (см. текст). МКБ-10 – Международная классификация болезней 10-го пересмотра; ПЦР – полимеразная цепная реакция*

Fig. 1. *Metric map reflecting associations between osteoarthritis and various characteristics of patients' conditions in sample 1 (see text). ICD-10 – International Classification of Diseases 10th revision; PCR – polymerase chain reaction*

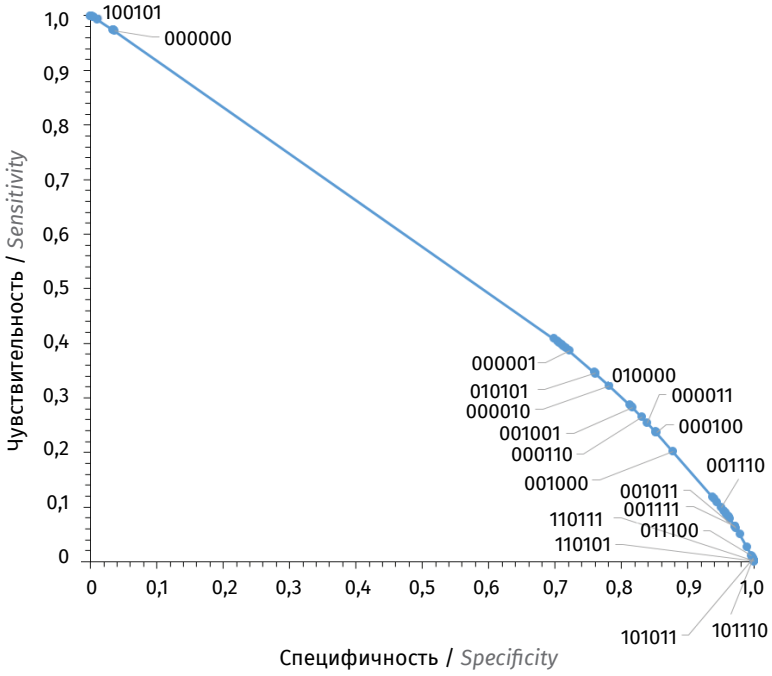

Рис. 2. *Кривая «специфичность–чувствительность» для сочетаний предикторов остеоартрита, которые отличались наибольшей частотой встречаемости в выборке 1. Каждое сочетание обозначено комбинацией единиц и нулей так, что крайне левая (1-я) позиция сочетания соответствует наличию гипертонии (есть – 1, нет – 0), 2-я позиция – приему статинов, 3-я – ингибиторов протонной помпы, 4-я – приему антиаритмических средств, 5-я – приему диуретиков, 6-я – приему кардиоселективных бета-блокаторов. На рисунке обозначены прежде всего наиболее часто встречающиеся сочетания*

Fig. 2. *Specificity–sensitivity curve for combinations of osteoarthritis predictors with the highest frequencies in sample 1. Each combination is represented as a combination of zeros and ones: leftmost (1st) position describes presence of hypertension (1 – present, 0 – absent), 2nd position – administration of statins; 3rd – inhibitors of proton pump; 4th – antiarrhythmic medications; 5th – diuretics; 6th – cardioselective beta blockers. The figure shows the most common combinations*

отличались наибольшей частотой встречаемости в выборке 1. К этим предикторам относились наличие у пациента гипертонии (15 % обследованных), прием статинов (18 %), ингибиторов протонной помпы (43 %), антиаритмических препаратов (22 %), диуретиков (32 %), кардиоселективных бета-блокаторов (35 %). Результаты представлены на рис. 2 и в табл. 1.

Анализ сочетаний наиболее часто встречающихся признаков пациентов с ОА (см. рис. 1) позволил не только описать наиболее характерные группы риска пациентов с ОА (см. первые 5 строк табл. 1), но и выявить сочетания признаков, приводящих к ложноотрицательному классифицированию пациентов с ОА. В частности, интересно отметить, что сочетание признаков, обозначенное как «001000» (пациенты без гипертонии, принимающие ингибиторы протонной помпы, без приема диуретиков, антиаритмиков, бета-блокаторов, статинов), встречалось у 18 % обследованных и соответствовало небольшому повышению риска ОА (всего на 39 %, отношение шансов (ОШ) для риска ОА 1,39). В то же время данное сочетание соответствовало 5-кратному повышению риска ошибочной классификации пациента с ОА (ОШ 5,52). Аналогичная картина наблюдалась для сочетания «000100» (пациенты без гипертонии, принимающие антиаритмики, без приема диуретиков, ингибиторов протонной помпы, бета-блокаторов, статинов), найденного у 8 % обследованных (см. табл. 1). Анализ сочетаний, приведенных в табл. 1, показал, что более низкая информативность сочетания признаков/предикторов приводит к бóльшей ошибке при прогнозировании ОА (рис. 3).

Результаты анализа сочетаний признаков/предикторов на рис. 1 подтверждаются результатами корреляционного анализа. Например, с повышенным риском ОА коррелирует прием ингибиторов протонной помпы (коэффициент корреляции 0,53). Результаты корреляционного анализа и анализа сочетаний признаков были использованы для разработки балльной шкалы № 1 (табл. 2).

Таблица 1. *Сочетания наиболее часто встречающихся предикторов ОА. Каждое сочетание обозначает (слева направо) наличие у пациента АГ (есть – 1, нет – 0), приема статинов, ингибиторов протонной помпы, антиаритмических средств, диуретиков, кардиоселективных бета-блокаторов*

Table 1. *Combinations of the most common OA predictors. Each combination describes (from left to right) presence of AH (present – 1, absent – 0), administration of statins, proton pump inhibitors, antiarrhythmic medications, diuretics, cardioselective beta blockers*

Сочетание Combination	*n*	ОШ ОА OA OR	ОШ ошибки Error OR	Группа риска пациентов Risk group
011101	9	5,08	1,55	Пациенты без АГ, принимающие статины, ингибиторы протонной помпы, антиаритмики и бета-блокаторы, без приема диуретиков Patients without AH taking statins, proton pump inhibitors, antiarrhythmic medications, cardioselective beta blockers but not taking diuretics
110001	13	3,48	2,33	Пациенты с АГ, принимающие бета-блокаторы, без приема ингибиторов протонной помпы, статинов, диуретиков, антиаритмиков Patients with AH taking beta blockers, but not proton pump inhibitors, statins, diuretics, antiarrhythmic medications
101011	20	3,32	2,56	Пациенты с АГ, принимающие ингибиторы протонной помпы и бета-блокаторы, без приема статинов, диуретиков, антиаритмиков Patients with AH taking proton pump inhibitors and beta blockers, but not statins, diuretics, antiarrhythmic medications
111000	19	2,37	3,50	Пациенты с АГ, принимающие статины, ингибиторы протонной помпы, без приема диуретиков, антиаритмиков и бета-блокаторов Patients with AH taking statins, proton pump inhibitors but not diuretics, antiarrhythmic medications and beta blockers
011100	22	2,32	3,63	Пациенты без АГ, принимающие статины, ингибиторы протонной помпы, антиаритмики, без приема диуретиков и бета-блокаторов Patients without AH taking statins, proton pump inhibitors, antiarrhythmic medications but not diuretics and beta blockers
001100	194	2,19	4,26	Пациенты без АГ, принимающие ингибиторы протонной помпы и антиаритмики, без приема диуретиков, статинов и бета-блокаторов Patients without AH taking proton pump inhibitors and antiarrhythmic medications but not diuretics, statins and beta blockers
001010	276	2,13	4,33	Пациенты без АГ, принимающие ингибиторы протонной помпы и диуретики, без приема антиаритмиков, статинов и бета-блокаторов Patients without AH taking proton pump inhibitors and diuretics, but not antiarrhythmic medications, statins and beta blockers
001011	141	2,03	4,57	Пациенты без АГ, принимающие ингибиторы протонной помпы, диуретики и бета-блокаторы, без приема антиаритмиков или статинов Patients without AH taking proton pump inhibitors, diuretics and beta blockers but not antiarrhythmic medications or statins
001111	45	2,03	4,37	Пациенты без АГ, принимающие ингибиторы протонной помпы, диуретики, антиаритмики и бета-блокаторы, без приема статинов Patients without AH taking proton pump inhibitors, diuretics, antiarrhythmic medications and beta blockers but not statins
001000	1393	1,39	5,52	Пациенты без АГ, принимающие ингибиторы протонной помпы, без приема диуретиков, антиаритмиков, бета-блокаторов, статинов Patients without AH taking proton pump inhibitors but not diuretics, antiarrhythmic medications, beta blockers and statins
000100	576	1,39	4,68	Пациенты без АГ, принимающие антиаритмики, без приема диуретиков, ингибиторов протонной помпы, бета-блокаторов, статинов Patients without AH taking antiarrhythmic medications but not diuretics, proton pump inhibitors, beta blockers, statins
000011	295	1,31	6,19	Пациенты без АГ, принимающие бета-блокаторы и диуретики, без приема ингибиторов протонной помпы, статинов, антиаритмиков Patients without AH taking beta blockers and diuretics but not proton pump inhibitors, statins, antiarrhythmic medications
000110	171	1,28	6,99	Пациенты без АГ, принимающие бета-блокаторы и диуретики, без приема ингибиторов протонной помпы, статинов, антиаритмиков Patients without AH taking beta blockers and diuretics, but not proton pump inhibitors, statins, antiarrhythmic medications

Окончание табл. 1
End of table 1

Сочетание Combination	*n*	ОШ ОА OA OR	ОШ ошибки Error OR	Группа риска пациентов Risk group
100100	41	0,98	4,66	Пациенты с АГ, принимающие антиаритмики, без приема ингибиторов протонной помпы, диуретиков, бета-блокаторов, статинов Patients with AH taking antiarrhythmic medications but not diuretics, proton pump inhibitors, beta blockers, statins
010010	103	0,75	10,43	Пациенты без АГ, принимающие статины и диуретики, без приема антиаритмиков, ингибиторов протонной помпы, бета-блокаторов Patients without AH taking statins and diuretics but not antiarrhythmic medications proton pump inhibitors, beta blockers
010100	34	0,54	13,99	Пациенты без АГ, принимающие статины и антиаритмики, без приема диуретиков, ингибиторов протонной помпы, бета-блокаторов Patients without AH taking statins and antiarrhythmic medications but not diuretics, proton pump inhibitors, beta blockers

Примечание. *ОА — остеоартрит; АГ — артериальная гипертония; ОШ ОА — отношение шансов для оценки риска ОА при наличии у пациента данного сочетания признаков; ОШ ошибки — отношение шансов того, что наличие данного сочетания приведет к ошибочной (ложнонегативной) интерпретации пациента с ОА. Сочетания предикторов упорядочены по убыванию значений ОШ ОА.*
Note. *OA — osteoarthritis; AH — arterial hypertension; OA OR — odds ratio for risk of OA in a patients with this combination of characteristics; error OR — odds ratio for probability that this combination will lead to erroneous (false negative) interpretation of patients with OA. Predictor combinations are listed by decrease in OA OR.*

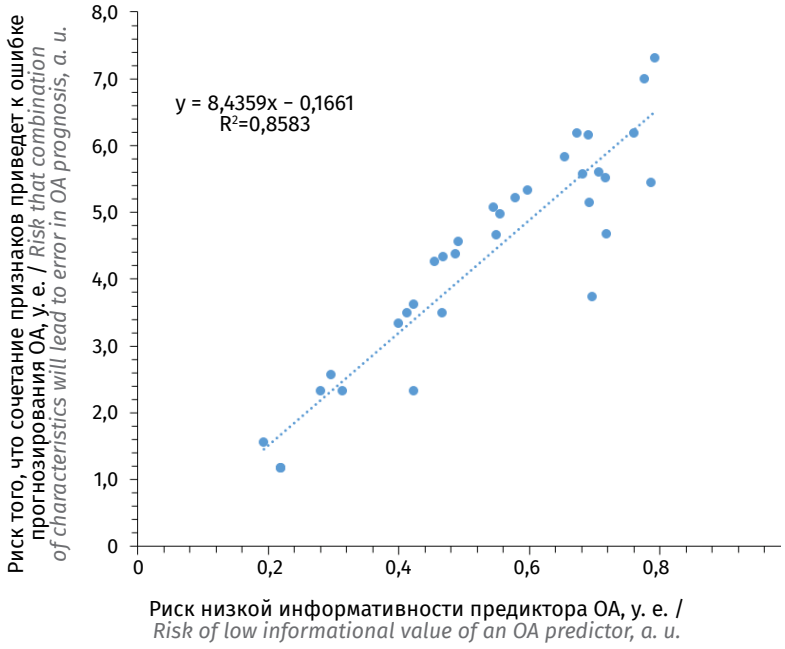

Рис. 3. *Соотношение между информативностью типичных сочетаний предикторов остеоартрита (ОА) и ошибками классификации пациентов с ОА. Риск низкой информативности соответствует делению единицы на показатель ОШ ОА (отношение шансов для оценки риска ОА при наличии у пациента данного сочетания признаков (см. табл. 1)). Риск того, что сочетание признаков приведет к ошибке соответствует показателю ОШ ошибки (отношение шансов того, что наличие данного сочетания приведет к ложнонегативной классификации пациентов с ОА (см. табл. 1))*
Fig. 3. *Relationship between informational value of typical combinations of osteoarthritis (OA) predictors and errors in classification of patients with OA. Risk of low informational value corresponds to division of one by OA OR (odds ratio for estimation of OA risk in a patient with this combination of characteristics (see Table 1)). Risk that a combination of characteristics with lead to error corresponds to error OR (odds ratio for false negative classification of patients with OA having this combination (see Table 1))*

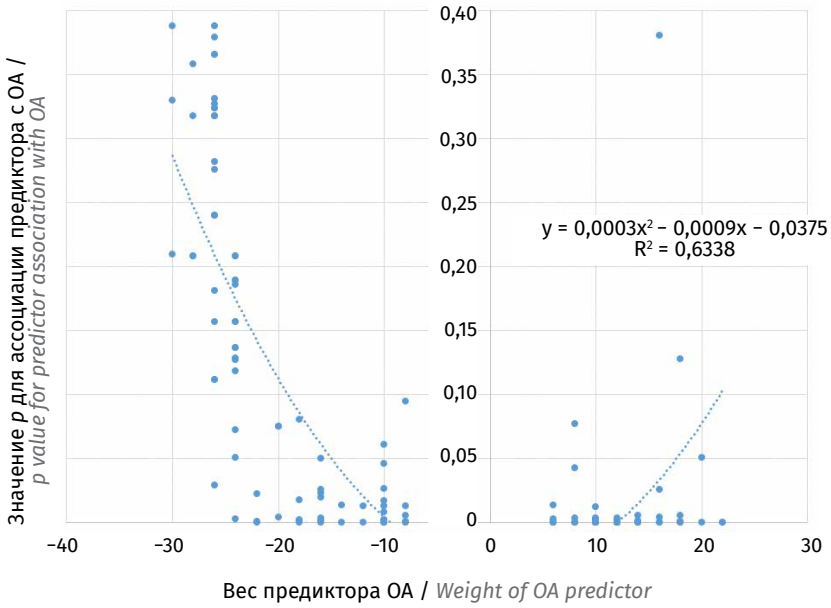

Рис. 4. *Соотношение между статистической значимостью отдельных предикторов (значение p по t-тесту) и весами предикторов в линейной схеме прогнозирования остеоартрита (ОА) ($y = \Sigma\, \omega_i p_i > C$)*
Fig. 4. *Ratio between statistical significance of individual predictors (p value per t-test) and predictor weights in linear scheme for osteoarthritis prediction (OA) ($y = \Sigma\, \omega_i p_i > C$)*

Балльная шкала строилась на основании поиска оптимального набора весов признаков в рамках решения задачи линейного прогнозирования. Иначе говоря, отклик y (т. е. наличие ОА) прогнозировался на основании суммирования признаков/предикторов p_i с весами ω_i по формуле $y = \sum \omega_i p_i$. Условием наличия ОА у пациента являлось превышение откликом $y = \sum \omega_i p_i$ некоторого порогового значения C, $y > C$. В результате решения данной задачи методами топологического анализа данных были найдены оптимальные значения весов ω_i. Интересно отметить, что бóльшие абсолютные значения весов признаков могут в ряде случаев быть обратно пропорциональны их статистической значимости (рис. 4).

В целом в результате анализа выборки 1, включившей информацию об анамнезе, симптоматике и фармакотерапии 3440 пациентов из ИМБД, была сформирована балльная шкала № 1 (табл. 2).

Таблица 2. *Балльная шкала № 1 для оценки риска остеоартрита у пациентов на основе данных об анамнезе, эпизодической симптоматике и фармакотерапии (выборка 1, n = 3440). В каждом из 3 разделов вопросы упорядочены по убыванию весов*
Table 2. *Scoring system No. 1 for estimation of osteoarthritis risk in patients based on anamnesis, episodic symptoms and pharmacotherapy (sample 1, n = 3440). In each of the 3 sections, questions are ordered by decrease in weight*

Вопрос Question	Вес Weight	Балл Score
Анамнез Anamnesis		
Пиелонефрит неуточненный Pyelonephritis unspecified	23,5	
Язва желудка Gastric ulcer	13,7	
Острая реакция на стресс (код F43.0 по Международной классификации болезней 10-го пересмотра) Acute stress response (F43.0 code in the International Classification of Diseases 10th revision)	6,3	

Продолжение табл. 2
Continuation of table 2

Вопрос **Question**	**Вес** **Weight**	**Балл** **Score**
Псориаз Psoriasis	4,8	
Заболевания щитовидной железы Thyroid disorders	3,8	
Проведена кожная проба для исключения туберкулеза Skin test was performed to rule out tuberculosis	3,6	
Инфекции мочевыводящих путей Urinary tract infections	3,5	
Гипертония Hypertension	3,2	
Проведен тест с помощью полимеразной цепной реакции для исключения гепатита Polymerase chain reaction test was performed to rule out hepatitis	2,8	
Бронхит Bronchitis	2,3	
Симптоматика (встречается эпизодически, не реже чем 1 раз в 6 мес) **Symptoms (episodic, occur at least once every 6 months)**		
Общая слабость General malaise	1,5	
Боли в колене Knee pain	1,4	
Сыпь на коже Skin rash	1,3	
Боли в пояснице Lower back pain	1,1	
Боли в шее Neck pain	0,8	
Фармакотерапия (в анамнезе) **Pharmacotherapy (anamnesis)**		
Антипсориатические средства Antipsoriatic drugs	11,8	
Противомалярийные хинолоны Antimalaria quinolones	7,8	
Эстрогенсодержащие препараты *per os* Estrogen-containing drugs *per os*	2,6	
Антиаритмики* Antiarrhythmics*	1,3	
Кардиоселективные бета-адреноблокаторы Cardioselectivity beta blockers	1,0	
Ингибиторы протонной помпы** Proton pump inhibitors**	0,7	
Диуретики Diuretics	0,6	

Окончание табл. 2
End of table 2

Вопрос **Question**	**Вес** **Weight**	**Балл** **Score**
Статины Statins	0,5	
Цитрат магния (регулярный прием) Magnesium citrate (regular administration)	–1,0	
Антиандрогенные препараты Antiandrogen drugs	–1,0	
Препараты лития (регулярно) Lithium drugs (regularly)	–1,1	
Цитрат кальция (регулярно) Calcium citrate (regularly)	–1,1	
Препараты цинка (регулярно) Zinc drugs (regularly)	–1,1	
Итого ***Total***		

**Если антиаритмические агенты употребляются при отсутствии гипертонии, приема статинов, ингибиторов протонной помпы, диуретиков и кардиоселективных бета-блокаторов, то вес данного признака умножается на 4.*
***Если ингибиторы протонной помпы употребляются при отсутствии гипертонии, приема статинов, антиаритмических агентов и диуретиков, то вес этого признака умножается на 5.*
Инструкция к использованию балльной шкалы. *При наличии признака поставьте «1» в крайней правой колонке. Затем сложите веса признаков, отмеченных «1» (с учетом * и **), и запишите сумму в ячейку «Итого». Величина полученного результата более 45 баллов соответствует повышенному риску остеоартрита (чувствительность – 88 %, специфичность – 100 %).*
**If antiarrhythmic agents are taken in the absence of hypertension, administration of statins, proton pump inhibitors, diuretics and cardioselective beta blockers, weight of this characteristics is multiplied by 4.*
***If proton pump inhibitors are taken in the absence of hypertension, administration of statins, antiarrhythmic agents and diuretics, weight of this characteristics is multiplied by 5.*
Instructions for using the scoring system. *If a characteristics is present, put "1" in the far-right column. Then calculate the sum of all the weights with "1" (taking into account * and **) and put it in the "Total" cell. Value above 45 corresponds to elevated risk of osteoarthritis (sensitivity 88 %, specificity 100 %).*

Для верификации шкалы № 1 (см. табл. 2) были проведены кросс-валидационные эксперименты, в ходе которых выборка 1 (n = 3440) случайным образом делилась на обучающую (n = 1720) и контрольную (n = 1720) выборки пациентов. Кросс-валидационные эксперименты (10 случайных разделений на обучающую и контрольную выборки) показали, что наиболее приемлемые результаты получаются при использовании порогового значения $C = 45$ в формуле прогнозирования $y = \sum \omega_i p_i > C$. Величина $\sum \omega_i p_i$ более 45 баллов соответствует повышенному риску ОА (чувствительность – 88 %, специфичность – 100 % (рис. 5)).

Аналогичным образом при анализе выборки 2 (n = 1290) была получена шкала № 2 для оценки риска ОА (табл. 3). Анамнестические предикторы в шкале № 2 включили наличие в анамнезе пациента анемии, хронических запоров, атеросклероза, нарушений зрения, герпесвирусной инфекции, опоясывающего лишая (вирусное заболевание с односторонними герпетиформными высыпаниями на коже и выраженной болевой симптоматикой), контагиозного моллюска (вирусная инфекция, вызываемая особой разновидностью вирусов оспы, который поражает кожу и слизистые оболочки) и др. Эпизодическая симптоматика (встречается не реже чем 1 раз в 6 мес) включила болезненные ощущения при пальпации брюшной полости, повышенное артериальное давление, боли в районе тазобедренного сустава, запястье, в ногах при ходьбе, при мочеиспускании, а также частые эпизоды покашливания, не связанные с острой респираторной вирусной инфекцией (см. табл. 3).

Настройка весов предикторов в шкале № 2 проводилась по той же линейной схеме прогнозирования ОА, т. е. $y = \sum \omega_i p_i > C$. Кросс-валидационный анализ показал, что оптимальным является пороговое значение C, равное 38 баллам: величина $\sum \omega_i p_i$ более 38 баллов соответствует повышенному риску ОА (чувствительность – 83 %, специфичность – 74 % (рис. 6)).

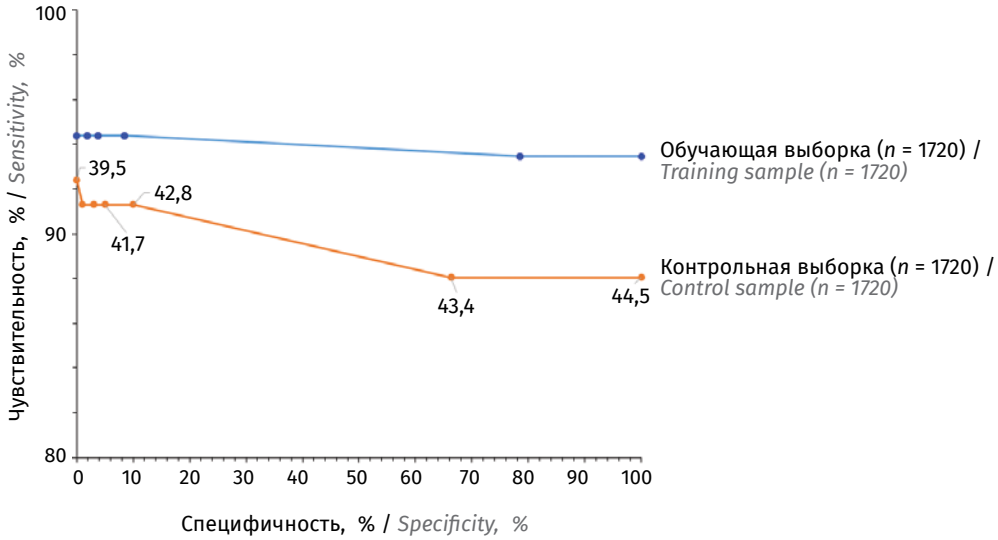

Рис. 5. *Верификация балльной шкалы № 1. Приведены примеры кривых «специфичность–чувствительность» для случайного разделения выборки 1 на обучающую и контрольную. Рядом с точками показаны различные пороговые значения баллов (параметр C в линейной схеме прогнозирования остеоартрита $y = \Sigma\, \omega_i p_i > C$)*

Fig. 5. *Verification of scoring system No. 1. Examples of «specificity–sensitivity» curves for random division of sample 1 into training and control samples are presented. Various threshold values (parameter C in the linear scheme of osteoarthritis prognosis $y = \Sigma\, \omega_i p_i > C$) are shown near the dots*

Таблица 3. *Балльная шкала № 2 для оценки риска остеоартрита у пациентов на основе данных об анамнезе и симптоматике (выборка 2, n = 1290). Вопросы расположены по убыванию весов*

Table 3. *Scoring system No. 2 for estimation of osteoarthritis risk in patients based on anamnesis and symptoms (sample 2, n = 1290). Questions are ordered by decrease in weight*

Вопрос Question	Вес Weight	Балл Score
Анамнестические данные **Anamnestic data**		
Герпесвирусная инфекция, опоясывающий лишай, ветряная оспа в анамнезе History of herpes virus infection, zoster, chicken pox	11,8	
Хронические запоры Chronic constipation	5,2	
Анемия в анамнезе History of anemia	3,4	
Пациент категорически не употребляет в пищу что-то из перечисленного: свежие овощи, фрукты, каши, кисломолочные продукты, свежую рыбу Patients never eats something of the following: fresh vegetables, fruits, porridge, dairy, fresh fish	3,3	
Нарушения зрения (близорукость, дальнозоркость, катаракта) Eye conditions (myopia, far-sightedness, cataract)	3,2	
Регулярный прием диуретиков Regular administration of diuretic	3,2	
Регулярный прием пероральных контрацептивов Regular administration of oral contraceptives	3,1	
Атеросклероз Atherosclerosis	2,2	

Окончание табл. 3
End of table 3

Вопрос **Question**	**Вес** **Weight**	**Балл** **Score**
Симптоматика (встречается эпизодически, не реже чем 1 раз в 6 мес) **Symptoms (episodic, occur at least once every 6 months)**		
Боли в районе тазобедренного сустава Pain around the hip joint	20,3	
Повышенное артериальное давление Elevated blood pressure	12,4	
Скованность в суставах и мышцах и/или фибромиалгия Joint and muscle stiffness and/or fibromyalgia	12,4	
Частые эпизоды покашливания, не связанные с острой респираторной вирусной инфекцией Frequent episodes of coughing not associated with acute respiratory viral infection	5,6	
Боли в запястье Wrist pain	4,2	
Боль или жжение при мочеиспускании Pain or burning during micturition	3,3	
Боли в ногах при ходьбе Leg pain during walking	3,2	
Болезненные ощущения при пальпации брюшной полости Painful sensations during abdominal palpation	3,2	
Итого ***Total***		

Инструкция к использованию балльной шкалы. *При наличии признака поставьте «1» в крайней правой колонке. Затем сложите веса признаков, отмеченных «1» и запишите сумму в ячейку «Итого». Величина полученного результата более 38 баллов соответствует повышенному риску остеоартрита (чувствительность – 83 %, специфичность – 74 %).*
Instructions for using the scoring system. *If a characteristic is present, put "1" in the far right column. Then calculate the sum of all the weights with "1" and put it in the "Total" cell. Value above 38 corresponds to elevated risk of osteoarthritis (sensitivity 83 %, specificity 74 %).*

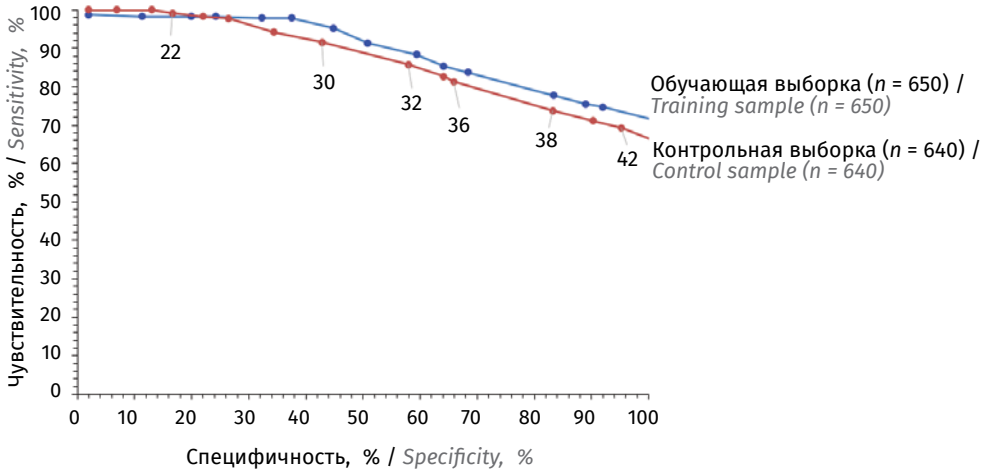

Рис. 6. *Верификация балльной шкалы № 2. Рядом с точками показаны различные пороговые значения баллов (параметр C в линейной схеме прогнозирования остеоартрита $y = \Sigma\, \omega_i p_i > C$)*
Fig. 6. *Verification of scoring system No. 2. Various threshold values (parameter C in the linear scheme of osteoarthritis prognosis ($y = \Sigma\, \omega_i p_i > C$) are shown near the dots*

Заключение

Общепринятыми признаками ОА являются болезненность и ограничение объема движений в суставе при ходьбе, ночью, в положении сидя, ощущения «нестабильности» сустава, ограничение дистанции ходьбы и, конечно, рентгенологические признаки (сужение суставной щели, плотные утолщения по краю суставной щели, признаки воспаления и др.). Однако эти признаки указывают на то, что заболевание уже сформировалось. Поэтому весьма важно выявление предикторов ОА, основанных не только на упоминаемых выше симптомах ОА, но и на данных о перенесенных ранее заболеваниях, фармакоанамнезе пациента и симптоматике, напрямую не связанной с ОА. Для формирования балльных шкал для оценки риска ОА на основании такого рода «неочевидных» предикторов перспективно использовать методы анализа больших данных (тысячи пациентов, тысячи признаков), разрабатываемые в рамках алгебраического подхода к распознаванию (см. ресурс http://www.bigdata-mining.ru).

В настоящей работе предложены 2 шкалы оценки риска ОА, основанные на «неочевидных» предикторах ОА. Шкалы № 1 и № 2 были получены в результате анализа 2 выборок пациентов в возрасте 35–90 лет и характеризуются приемлемо высокими чувствительностью (83–88 %) и специфичностью (74–100 %). Для пациентов с высокими баллами по любой из этих шкал перспективно своего рода «упреждающее» назначение хондропротекторов на основе высокоочищенных субстанций хондроитина сульфата и глюкозамина сульфата, стандартизированных по содержанию действующих веществ.

Дополнительная информация / Additional information

Источник финансирования. Систематический анализ данных проведен в рамках текущего финансирования организации Федерального исследовательского центра «Информатика и управление» РАН.
Funding source. Analysis of the a systematic data was conducted within the framework of the organization's current funding Computer Science and Control, Russian Academy of Sciences.

Раскрытие интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.
Disclosure of interests. The authors declare that they have no apparent or potential conflicts of interest related to the publication of this article.

Вклад авторов. Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, подготовку статьи, прочли и одобрили финальную версию перед публикацией). И.Ю. Торшин – разработка дизайна исследования, анализ данных, обзор публикаций, написание текста, редактирование; А.Н. Громов – получение данных для анализа, анализ данных, редактирование; И.А. Рейер, Б.И. Беликов, О.А. Лиманова – получение данных для анализа, анализ данных; В.В. Семёнов, Т.Р. Гришина – получение данных для анализа, обзор публикаций по теме статьи; Л.Э. Федотова – получение данных для анализа, редактирование; О.А. Громова – разработка дизайна исследования, получение данных для анализа, анализ данных, обзор публикаций по теме статьи, написание текста, редактирование.
Authors' contribution. All authors confirm that they meet the international ICMJE criteria for authorship (all authors have made a significant contribution to the development of the concept, preparation of the article, and have read and approved the final version before publication). I.Yu. Torshin – design, data analysis, publication review, article writing, editing; A.N. Gromov – data acquisition for analysis, data analysis, editing; I.A. Reyer, B.I. Belikov, O.A. Limanova – data acquisition for analysis, data analysis; V.A. Semenov, T.R. Grishina – data acquisition for analysis, publication review on the topic of the article; L.E. Fedotova – data acquisition for analysis, editing; O.A. Gromova – design, data acquisition for analysis, data analysis, review of publications on the topic of the article, article writing, editing.

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовали.
Generative artificial intelligence statement. No generative artificial intelligence technologies were used in the creation of this article.

Список литературы / References

1. Торшин И.Ю., Лила А.М., Наумов А.В. и др. Перспективы персонификации профилактики и терапии остеоартрита на основании анализа коморбидного фона, генетических полиморфизмов и микроэлементного статуса. Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология 2021;14(1):28–39. DOI: 10.17749/2070-4909/FARMAKOEKONOMIKA.2021.077
Torshin I.Yu., Lila A.M., Naumov A.V. et al. Prospects for personalizing the prevention and treatment of osteoarthritis based on the analysis of comorbid background, genetic polymorphisms, and microelement status. Farmakoekonomika. Sovremennaya farmakoekonomika i farmakoepidemiologiya = Pharmacoeconomics. Modern Pharmacoeconomics and Pharmacoepidemiology 2021;14(1):28–39. (In Russ.). DOI: 10.17749/2070-4909/FARMAKOEKONOMIKA.2021.077
2. Nukala S., Puvvada S.R., Luvsannyam E. et al. Hyperlipidemia and statin use on the progression of osteoarthritis: a systematic review. Cureus 2021;13(6):e15999. DOI: 10.7759/cureus.15999
3. Torshin I.Yu., Rudakov K.V. On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification. Patt Recognit Imag Anal (Adv Math Theory Appl) 2015;25(4):577–87. DOI: 10.1134/S1054661815040252
4. Лиманова О.А., Торшин И.Ю., Сардарян И.С. и др. Обеспеченность микронутриентами и женское здоровье: интеллектуальный анализ клинико-эпидемиологических данных. Вопросы гинекологии, акушерства и перинатологии 2014;13(2):5–15.
Limanova O.A., Torshin I.Yu., Sardaryan I.S. et al. Micronutrient status and women's health: an intelligent analysis of clinical and epidemiological data. Voprosy ginekologii. akusherstva i perinatologii = Issues of Gynecology, Obstetrics and Perinatology 2014;13(2):5–15. (In Russ.).

5. Torshin I.Yu., Rudakov K.V. On the procedures of generation of numerical features over partitions of sets of objects in the problem of predicting numerical target variables. Patt Recognit Imag Anal 2019;29(4):654–67. DOI: 10.1134/S1054661819040175

6. Torshin I.Yu., Rudakov K.V. On metric spaces arising during formalization of recognition and classification problems. part 1: properties of compactness. Patt Recognit Imag Anal (Adv Math Theory Appl) 2016;26(2):274.

Об авторах / Authors' info

Торшин Иван Юрьевич, канд. физ.-мат. наук, канд. хим. наук;
Ivan Yu. Torshin, Cand. Sci. (Physics and Mathematics), Cand. Sci. (Chemical);
ORCID: 0000-0002-2659-7998; e-mail: tiy135@yahoo.com

Громов Андрей Николаевич;
Andrey N. Gromov;
ORCID: 0000-0001-7507-191X; e-mail: gromlogin@gmail.com

Рейер Иван Александрович, канд. тех. наук;
Ivan A. Reyer, Cand. Sci. (Technology);
ORCID: 0000-0002-7928-053X; e-mail: reyer@forecsys.ru

Беликов Богдан Ильич;
Bogdan I. Belikov;
ORCID: 0009-0002-3538-7324

Семёнов Владимир Александрович, д-р мед. наук, профессор;
Vladimir A. Semenov, MD, Dr. Sci. (Medicine), Professor;
ORCID: 0000-0002-2902-6816; e-mail: semenov_v_a.717@mail.ru

Гришина Татьяна Романовна, д-р мед. наук, профессор;
Tatiana R. Grishina, MD, Dr. Sci. (Medicine), Professor;
ORCID: 0000-0002-1665-1188

Лиманова Ольга Адольфовна, канд. мед. наук;
Olga A. Limanova, MD, Cand. Sci. (Medicine);
ORCID: 0000-0002-2342-3036

Федотова Любовь Эдуардовна, канд. мед. наук;
Lyubov E. Fedotova, MD, Cand. Sci. (Medicine);
ORCID: 0000-0002-0778-1562

*__Громова Ольга Алексеевна,__ д-р мед. наук, профессор;
*__Olga A. Gromova,__ MD, Dr. Sci. (Medicine), Professor;
ORCID: 0000-0002-7663-710X; e-mail: unesco.gromova@gmail.com

Рукопись получена: 15.05.2026. **Рукопись одобрена:** 01.07.2026. **Опубликована онлайн:** 10.10.2026.
Submitted: 15.05.2026. **Accepted:** 01.07.2026. **Published online:** 10.10.2026.

*Автор, ответственный за переписку
*Corresponding author