

Компьютерный анализ эмоциональной модальности 20 млн публикаций в базе данных PUBMED указывает на пути повышения эффективности фармакотерапии посредством идентификации псевдонаучных публикаций, направленных на негативную эмоциональную «накачку» врачей

**Максимов В. А.¹, Торшин И. Ю.^{2,3}, Громова О. А.^{2,3}, Галустян А. Н.⁴,
Гоголева И. В.⁵, Гришина Т. Р.⁵, Громов А. Н.^{2,3}, Калачева А. Г.⁵, Кобалава Ж. Д.⁶,
Коденцова В. М.⁷, Лиманова О. А.⁵, Малявская С. И.⁸, Рудаков К. В.^{2,9},
Сардарян И.С.⁴, Сорокин А. И.^{2,3}, Стаховская Л. В.¹⁰, Тапильская Н. И.¹¹,
Тетраушвили Н. К.¹², Федотова Л. Э.⁵, Хаджидис А. К.⁴**

¹ – ФГБОУ ДПО «Российская медицинская академия непрерывного профессионального образования»
Министерства здравоохранения Российской Федерации, Москва, Россия

² – ФГУ «Федеральный исследовательский центр «Информатика и управление» Российской академии наук»,
Москва, Россия

³ – Центр хранения и анализа больших данных ФГБОУ ВО «Московский государственный университет
им. М.В. Ломоносова», Москва, Россия

⁴ – ФГБОУ ВО «Санкт-Петербургский государственный педиатрический медицинский университет»
Министерства здравоохранения Российской Федерации, Санкт-Петербург, Россия

⁵ – ФГБОУ ВО «Ивановская государственная медицинская академия» Министерства здравоохранения
Российской Федерации, Иваново, Россия

⁶ – ФГАОУ ВО «Российский университет дружбы народов» (Медицинский институт), Москва, Россия

⁷ – ФГБУН «Федеральный исследовательский центр питания, биотехнологии и безопасности пищи»,
Москва, Россия

⁸ – ФГБОУ ВО «Северный государственный медицинский университет» Министерства здравоохранения
Российской Федерации, Архангельск, Россия

⁹ – ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)»,
Москва, Россия

¹⁰ – ФГБУ «Федеральный центр цереброваскулярной патологии и инсульта» Министерства здравоохранения
Российской Федерации, Москва, Россия

¹¹ – ФГБНУ «Научно-исследовательский институт акушерства, гинекологии и репродуктологии имени Д.О. Отта»,
Санкт-Петербург, Россия

¹² – ФГБУ «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии
имени академика В.И. Кулакова» Министерства Здравоохранения Российской Федерации, Москва, Россия

Аннотация. Поиск оригинальных публикаций по фундаментальной и клинической медицине наивысшего научного качества – насущная необходимость для каждого врача-исследователя. Такие публикации принципиально необходимы, в частности, для разработки надёжных стандартов лечения. Существенной подмогой в решении этой задачи являются англоязычные ресурсы PUBMED и EMBASE. Однако существует очевидная проблема оценки качества находимых исследований. В работе сформулирована методика анализа текстов биомедицинских публикаций, в основе которой лежит алгоритмическая оценка эмоциональной модальности медицинских текстов (т. н. сентимент-анализа). Применение топологической теории анализа данных позволило разработать комплекс высокоточных алгоритмов для выявления 16 типов сентиментов (манипулятивные обороты речи, исследования без положительных результатов, пропаганда, подделка результатов, негативное личное отношение, агрессивность текста, негативный эмоциональный фон и др.). На основе разработанных алгоритмов получена балльная шкала оценки сентимент-качества исследований, которую мы назвали «β-баллом»: чем выше β-балл, тем в меньшей степени оцениваемый текст содержит манипулятивные языковые конструкции. В результате разработана и апробирована система ANTIFAKE (<http://antifake-news.ru>), предназначенная для анализа сентимент-качества англоязычных научных текстов. Анализ ~20 млн абстрактов из PUBMED показал, что публикации с низким сентимент-качеством (β-балл < 0, т. е. преобладание манипулятивных конструкций над содержательными) составляют всего 19 %. В подавляющем большинстве тематических рубрик (27 090 из 27 840 рубрик MESH-системы PUBMED) показана положительная динамика сентимент-качества текстов публикаций по годам. В то же время, в результате исследования были выделены 249 рубрик с резко отрицательной динамикой сентимент-качества и с выраженным нарастанием манипулятивных сентиментов, характерных для «жёлтой» англоязычной прессы. Данные рубрики включают десятки тысяч публикаций в реферируемых журналах,

которые направлены на (1) легализацию этически порочных практик (эвтаназия, т. н. «контроль популяций» и т. п.), (2) дискредитацию психиатрии как науки, (3) медийную войну против микронутриентов и (4) дискредитацию доказательной медицины под видом разработки т. н. «международных стандартов доказательной медицины». В целом, разработанная система искусственного интеллекта позволяет отфильтровывать псевдонаучные публикации, текст которых перегружен эмоциональной манипуляцией и которые публикуются под маской «доказательных стандартов».

Ключевые слова: фармакоинформатика; доказательная медицина; интеллектуальный анализ данных; алгоритмы оценки качества публикаций

Для цитирования:

Максимов В.А., Торшин И.Ю., Громова О.А., Галустян А.Н., Гоголева И.В., Гришина Т.Р., Громов А. Н., Калачева А.Г., Кобалава Ж.Д., Коденцова В.М., Лиманова О.А., Маливская С.И., Рудаков К.В., Сардарян И.С., Сорокин А.И., Стаховская Л.В., Тапильская Н.И., Тетруашвили Н.К., Федотова Л.Э., Хаджидис А.К. Компьютерный анализ эмоциональной модальности 20 млн публикаций в базе данных PUBMED указывает на пути повышения эффективности фармакотерапии посредством идентификации псевдонаучных публикаций, направленных на негативную эмоциональную «накачку» врачей // *Фармакокинетика и фармакодинамика*. – 2020. – № 4. – С. 19–40. DOI: 10.37489/2587-7836-2020-4-19-40

Computer analysis of the emotional modality of 20 million publications in PUBMED database indicates ways to increase the effectiveness of pharmacotherapy by identifying pseudoscientific publications aimed at negative emotional "pumping" of doctors

Maximov VA¹, Torshin IYu^{2,3}, Gromova OA^{2,3}, Galustyan AN⁴, Gogoleva IV⁵, Grishina TR⁵, Gromov AN^{2,3}, Kalacheva AG⁵, Kobalava ZD⁶, Kodentsova VM⁷, Limanova OA⁵, Malyvskaya SI⁸, Rudakov KV^{2,9}, Sardaryan IS⁴, Sorokin AI^{2,3}, Stakhovskaya LV¹⁰, Tapilskaya NI¹¹, Tetrushvili NK¹², Fedotova LE⁵, Hadzhidis AK⁴

¹ – FSBEI FPE "Russian Medical Academy of Continuous Professional Education" of the Ministry of Healthcare of the Russian Federation, Moscow, Russia

² – Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia

³ – Big Data Storage and Analysis Center FSEI HPE Lomonosov Moscow State University, Moscow, Russia

⁴ – Federal State Budgetary Institution of Higher Professional Education "St. Petersburg State Pediatric Medical University" of the Ministry of Healthcare of the Russian Federation, Saint-Petersburg, Russia

⁵ – FSBEI HE "Ivanovo State Medical Academy" of the Ministry of Healthcare of the Russian Federation, Ivanovo, Russia

⁶ – Peoples' Friendship University of Russia, Moscow, Russia

⁷ – FSBSI "Federal Research Center for Nutrition, Biotechnology and Food Safety", Moscow, Russia

⁸ – FSBEI HE "Northern State Medical University" of the Ministry of Health of the Russian Federation, Arkhangelsk, Russia

⁹ – Moscow Institute of Physics and Technology, Moscow, Russia

¹⁰ – FSBS "Federal Center for Cerebrovascular Pathology and Stroke" of the Ministry of Health of the Russian Federation, Moscow, Russia

¹¹ – FSBSI "The Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott", Saint-Petersburg, Russia

¹² – FSBI "National Medical Research Center for Obstetrics, Gynecology and Perinatology named after Academician V.I. Kulakov"

Ministry of Healthcare of the Russian Federation, Moscow, Russia

Abstract. The search for original publications on fundamental and clinical medicine that would produce results of the highest scientific quality represents an urgent need for every medical researcher. Such publications are essential, in particular, for the development of reliable treatment standards. The English-language resources PUBMED and EMBASE are essential to help in solving this problem. However, there is an obvious problem in assessing the quality of the studies found. The paper formulates a method for analyzing the texts of biomedical publications, which is based on an algorithmic assessment of the emotional modality of medical texts (so-called sentiment analysis). The use of the topological theory of data analysis made it possible to develop a set of high-precision algorithms for identifying 16 types of sentiments (manipulative turns of speech, research without positive results, propaganda, falsification of results, negative personal attitude, aggressiveness of the text, negative emotional background, etc.). On the basis of the developed algorithms, a point scale for assessing the sentiment quality of research was obtained, which we called the "β-score": the higher the β-score, the less the evaluated text contains manipulative language constructions. As a result, the ANTIFAKE system (<http://antifake-news.ru>) was developed to analyze the sentiment-quality of English-language scientific texts. An analysis of ~ 20 million abstracts from PUBMED showed that publications with low sentiment quality (β-score <0, that is, that the prevalence of manipulative constructions over meaningful ones) is only 19 %. In the overwhelming majority of thematic headings (27,090 out of 27,840 headings of the MESH system PUBMED), a positive dynamics of sentiment quality of the texts of publications is shown by years). At the same time, as a result of the study, 249 headings were identified with sharply negative dynamics of sentiment quality and with a pronounced increase in manipulative sentiments characteristic of the "yellow" English-language press. These headings include tens of thousands of publications in peer-reviewed journals, which are aimed at (1) legalizing ethically unacceptable practices (euthanasia, perversions, so-called "population control", etc.), (2) discrediting psychiatry as a science, (3) media the war against micronutrients and (4) discrediting evidence-based medicine under the guise of developing the so-called "international standards of evidence-based medicine". In general, the developed system of artificial intelligence allows researchers to filter out pseudoscientific publications, the text of which is overloaded with emotional manipulation and which are published under the guise of "evidence-based standards".

Keywords: pharmacoinformatics; evidence-based medicine; data mining; algorithms for evaluating the quality of publications

For citations:

Maximov VA, Torshin IYu, Gromova OA, Galustyan AN, Grishina TR, Gromov AN, Kalacheva AG, Kobalava ZD, Kodentsova VM, Limanova OA, Malyvskaya SI, Rudakov KV, Sardaryan IS, Sorokin AI, Stakhovskaya LV, Tapilskaya NI, Tetrushvili NK, Fedotova LE, Hadzhidis AK. Computer analysis of the emotional modality of 20 million publications in PUBMED database indicates ways to increase the effectiveness of pharmacotherapy by identifying pseudoscientific publications aimed at negative emotional "pumping" of doctors. *Farmakokinetika i farmakodinamika*. 2020;(4):19–40. DOI: 10.37489/2587-7836-2020-4-19-40

Введение

Для адекватного планирования научных исследований и разработки соответствующих национальных программ по здоровью необходимы нахождение и анализ максимально возможного массива публикаций по той или иной теме из области биомедицины. Существенный объём публикаций в имеющихся базах данных (30 млн публикаций в PUBMED, 40 млн – в EMBASE и др.) подчёркивают насущную необходимость использования адекватных алгоритмов поиска.

Несмотря на то что тематический поиск в уже опубликованной литературе является наиважнейшим шагом в разработке адекватных рекомендаций и в планировании исследований, к этому шагу большинство исследователей, незнакомых с методами интеллектуального анализа данных, относятся достаточно поверхностно. Например, ряд важных моментов, относящихся к стратегиям поиска литературы, обсуждался в [1]. Среди них авторы упомянули несколько очевидных психологических «табу», которые мешают эффективному поиску информации. Одним из таких запретов является замечательное «правило 90 секунд», которое даёт много возможностей поразмышлять о печальных результатах выработки условных рефлексов по И.П. Павлову у людей. То есть, многие клиницисты даже не пытаются найти информацию, если не думают, что смогут сделать это за полторы минуты [1]. Другими словами, реорганизация повседневной рутины для поиска лучших исследований кажется невыполнимым или крайне противным большинству этих исследователей. Если информация не может быть найдена за полторы минуты, то делаем выводы наподобие «такой информации не имеется» или «нет данных» или «эффект не доказан» [2]. В настоящее время массовой распространённости «клипового сознания» более приемлемо было бы говорить о «правиле 30 секунд» или о ещё более коротких интервалах времени [3].

Кроме того, взирая на огромный объём информации (миллионы статей), многие исследователи поддаются под влияние самого настоящего искушения как-то сократить объём статей за счёт использования некоторых якобы «важных» и «общепринятых» ключевых слов вроде «... AND review», «... AND Cochrane», «... AND evidence-based» и т. п. Однако наш весьма об-

ширный опыт (см. ниже) показывает, что при использовании таких «оптимизаций» поиска может теряться до 95 % оригинальных исследований и метаанализов.

Тем не менее, при использовании адекватных наборов ключевых слов при целенаправленном и тщательном поиске литературы [3] уже имеющиеся алгоритмы поиска по базам данных PUBMED, EMBASE и др. позволяют достаточно точно выделять подмножества публикаций, релевантных той или иной теме исследований. Однако ни одна из имеющихся поисковых систем не позволяет оценивать качество находимых таким образом публикаций. А ведь вопрос об оценке качества отбираемых публикаций принципиально важен для нахождения объективной научной информации.

Наш опыт анализа доказательных данных из литературы был получен, в частности, в ходе работы над восемью монографиями, в каждой из которых цитируется от 1 000 до 2 500 оригинальных исследований и/или метаанализов (рис. 1, см. ресурс <http://pharmacoinformatics.ru/монографии-3/>). При подготовке каждой из монографий был проведён анализ от 12 000 до 27 000 источников из баз данных PUBMED/MEDLINE и ELIBRARY. В результате, была систематизирована обширная доказательная база по отдельным микронутриентам и микронутриентным комплексам и были сделаны важные выводы о насущной необходимости алгоритмических оценок качества публикаций.

В частности, при работе над монографией по витамину D [4], в которой представлены ссылки приблизительно на 2 500 оригинальных статей, мы провели детальный анализ массива из 12 700 публикаций по фундаментальным и клиническим исследованиям витамина D. Экспертный анализ всего этого массива публикаций показал, что не менее трети клинических публикаций носили ярко выраженный негативный характер: в тексте таких публикаций не было описано (и даже не было упомянуто) *ни одного* положительного клинического эффекта использования витамина D по отношению к той или иной патологии.

Поясним вышесказанное. Очевидно, что в полноценном научном исследовании всегда получается набор положительных и отрицательных результатов: какие-то факторы влияют на эффективность терапии, какие-то – не влияют. В то же время, если текст статьи

Систематизирована обширная доказательная база по отдельным микронутриентам и микронутриентным комплексам



Рис. 1. Примеры систематизации доказательной базы

представляет только отрицательные результаты (или, что бывает гораздо реже, только положительные результаты), то такая манера изложения указывает на научную односторонность исследования, вызывает серьёзные сомнения в мотивации такого «исследователя» и, вероятно, на определённого рода конфликт интересов.

Для целей настоящей статьи важен тот факт, что «негативные» или «безрезультатные» исследования отличаются особым стилем изложения, преобладанием определённой лексики и грамматических конструкций. Так, тексты безрезультатных исследований характеризуется избытком конструкций отрицания («did not», «no significant effect» и др.), излишней юридизацией и вычурным формализмом языковых конструкций: пациенты называются не «пациентами», а какими-то бесполовыми и аморфными «участниками» (англ. «participants»), вместо указания чётких клинических диагнозов по МКБ-10 фигурируют некие «основные точки» (англ. «primary outcome»), причём, зачастую, без чёткого определения того, что же «исследователи» в действительности называют этой «основной точкой» и т. п. Иначе говоря, экспертный анализ позволяет однозначно утверждать, что «безрезультатные» исследования написаны по некоторому негативному «шаблону» или «трафарету».

В ходе нашего анализа различных выборок статей особого внимания заслужил и тот факт, что тексты определённых публикаций характеризуются преобладанием штампов манипулятивного характера, которые в англоязычной литературе широко известны как «spin words» («слова для накручивания сознания») или «weasel words» (дословно, «хорьковые слова», т. е. слова, используемые для подмены обсуждения по сути на психологическое давление и манипуляцию). Сам термин восходит к шекспировской фразе «I can suck melancholy out of a song, as a weasel sucks eggs» («Я могу высосать меланхолию из песни как хорёк высасывает яйца», «Как вам это понравится», 2-ой акт) и подразумевает слова, которые «высасывают смысл из слов рядом с ними и оставляют только треск пустой скорлупы» [5]. Напомним, что здесь речь идёт о публикациях, зарегистрированных на столь серьёзном ресурсе, как PUBMED, а вовсе не о «жёлтой прессе», не об интернет-ресурсах и не о социальных сетях.

Например, словосочетание «adverse outcomes» (что можно перевести как «тяжёлые побочные эффекты») является стандартным и общеизвестным «хорьковым словом» [6], типичным не только для безрезультатных исследований, но и для бульварной прессы [7] («Healthy Daily», «Daily Mail», «BBC News» и т. п.). В бульварных листках данное словосочетание применяется исключительно для запугивания читателя, а не для представления какой-либо важной информации. К сожалению, как показывают результаты настоящего исследования, лексика англоязычных бульварных листков (англ. tabloid) широко используется и в некоторых англоязычных медицинских журналах.

Существование определённых «шаблонов» и «языковых штампов» в текстах биомедицинских публикаций позволяет предположить, что процесс отделения манипулятивных, псевдонаучных публикаций от нормальных публикаций может быть автоматизирован посредством использования современных методов машинного обучения и интеллектуального анализа данных. Поэтому, в настоящей работе анализ массива из 19 918 308 публикаций, найденных в базе данных PUBMED по запросу «(1970 [Date – Publication] : 2019 [Date – Publication])» и включившего аннотированные публикации в англоязычных рецензируемых научных журналах с полным текстом абстракта. Анализ всего этого массива текстов был проведён с использованием современных методов машинного обучения, основанных на топологической теории сентимент-анализа (т. е. анализа эмоциональной модальности текста), развиваемой в рамках алгебраического подхода к распознаванию и классификации.

Материалы и методы

Центральной задачей топологической теории сентимент-анализа является разработка алгоритмов классификации текстов, позволяющих отличать (с максимально возможной специфичностью и селективностью) тексты заданного класса («безрезультатные», «манипулятивные», «негативные коннотации» и др., см ниже) от всех остальных текстов. Системы машинного обучения оперируют с множествами прецедентов, представляющих собой материал обучения таких алгоритмов. Отдельный прецедент или объект (отдельное предложение, абстракт или полная научная публикация) состоит из признакового описания (т. е. слов и словосочетаний текста) и информации о принадлежности этого объекта к тому или иному классу объектов (например, класс K_1 – «манипулятивные тексты», класс K_2 – «нормальные тексты» и др.). Такие классы принадлежности объектов (текстов), естественно, задаются экспертом.

После того как методами топологической теории анализа данных [8–10] получены признаковые описания объектов и определены соответствующие классы K_1/K_2 , исследуется разрешимость поставленной задачи классификации текстов. Под разрешимостью задач понимается непротиворечивость множеств прецедентов (т. е. существования решения у задачи). В случае задачи поиска результативных исследований, условие разрешимости этой задачи записывается следующим образом:

$$\bigvee_{Pr} (A^1, K^1), (A^2, K^2) \times \left(\bigvee_{\alpha=1}^{|P|} p_{\alpha} : p_{\alpha}(A^1, P) = p_{\alpha}(A^2, P) \right) \Rightarrow K^1 = K^2, \quad (1)$$

где: Pr – множество всех текстовых прецедентов; $N = |Pr|$ – число текстов во множестве прецедентов; $(A^1, K^1), (A^2, K^2)$ – произвольные прецеденты из множества Pr . В прецеденте (A, K) множество A обозначает признаковое описание (например, набор слов или словосочетаний из всех текстов во множестве Pr), K – класс, к которому принадлежит прецедент (K_1 или K_2). P – множество всех значений признаков, найденных в прецедентах из Pr , $p_\alpha(A, P)$ – значение α -го бинарного признака из множества P в прецеденте (A, K) такое, что $p_\alpha(A, P) = 1$ когда ключевое слово содержится в прецеденте A и $p_\alpha(A, P) = 0$ в противном случае).

Для эффективного установления выполнимости условия (1) вводится некоторый функционал $D_{инф}(\alpha)$ [11–14], позволяющий оценить информативность α -го признака относительно классов K_1/K_2 . Тогда множество всех значений признаков P можно линейно упорядочить в соответствии со значением оценки информативности (т. е. по убыванию значений $D_{инф}$). После упорядочения множества P большим значениям $D_{инф}(\alpha)$ соответствуют меньшие значения ранга информативности α , так что условие (1) записывается как критерий разрешимости на упорядоченном множестве признаков:

$$\forall_{Pr} (i, j, i \neq j) : A^i \in K_1, A^j \in K_0 \Rightarrow \exists_P p_\alpha, \alpha \rightarrow \min : p_\alpha(A^i) > p_\alpha(A^j), \quad (2)$$

где: $K_0 \subset P$ – минимальное по размеру множество признаков, на котором выполнено условие (2).

В качестве $D_{инф}$ был применен один из D -функционалов [14–16], отвечающий следующим качественным представлениям эксперта об информативности терминов: (а) наиболее интересны термины, встречающиеся «часто» в выборке K_1 (например, десятки раз) и «достаточно редко» (единицы) – в выборке K_2 ; (б) термины, встречающиеся «очень часто» (сотни и тысячи раз) должны учитываться только по мере необходимости для выполнения условия разрешимости (2); (в) термины, встречающиеся «редко» (единицы) в выборке K_1 характеризуются наименьшей информативностью; (г) наибольший интерес представляют термины, наиболее релевантные по смыслу (например, «адекватная терапия»).

Значения $D_{инф}$ для различных ключевых слов вычислялись следующим образом. Пусть α -й термин встречается в n_1^α абстрактах выборки K_1 , и в n_2^α абстрактах выборки K_2 , размеры выборок K_1 и K_2 – N_1 и N_2 абстрактов, соответственно, а частоты встречаемости термина – $v_1^\alpha = n_1^\alpha / (N_1 + N_2)$ и $v_2^\alpha = n_2^\alpha / (N_1 + N_2)$. Требованиям (а, б, в, г) вполне отвечает D -функционал вида (3):

$$D_{инф} = \gamma_\alpha \cdot (n_1^\alpha - \varepsilon_2 \cdot n_2^\alpha / D_1^\alpha(v_1^\alpha, \varepsilon_2)), \quad (3)$$

где: γ_α – смысловая релевантность α -го термина в соответствии с заданным экспертами словарем терминов; $\varepsilon_0 = N_1 / (N_1 + N_2)$ – отношение размеров выборок K_1 и K_2 ; D_1^α – кусочно-линейная функция вида (3'), указывающая, насколько чаще α -й термин встречается в K_1 , чем в K_2 . Например, $D_1^\alpha = 1,0$ соответствует тому, что α -й термин встречается только среди абстрактов выборки K_1 и ни разу – в K_2 .

$$D_1^\alpha(v_1^\alpha, v_2) = \begin{cases} v_1^\alpha \leq v_2 : 0 \\ v_1^\alpha > v_2 : \frac{v_1^\alpha - v_2}{1 - v_2} \end{cases} \quad (3')$$

Некоторые признаки могут быть избыточны, так как встречаются в устойчивых сочетаниях слов (например, «лекарственная терапия»), семантически связанных терминов или синонимов (например, «сердечно-сосудистая система») и др. Поэтому, при анализе условия (2) важно принимать во внимание *парные взаимодействия терминов*. Для произвольного класса K «расстояние» между i -м и j -м терминами оценивалось как $r(i, j) = |T_i \Delta T_j| / |T_i \cup T_j|$, где: $T_i = \{A \in K : p_i(A) = 1\}$, $T_j = \{A \in K : p_j(A) = 1\}$. С использованием метрики $r(i, j)$ проводится кластеризация терминов, так что в условии (2) фигурировали уже не отдельные слова/термины, а кластера терминов, установленные методом анализа метрических сгущений [8, 12].

Вычисление условия (2) на множестве прецедентов и лежит в основе использованного метода отбора результативных публикаций для мегаанализа. Условие (2) позволяет вычислить множество $K_0 \subset P$, т. е. отобрать *наиболее информативные признаки*, гарантирующие разрешимость задачи классификации текстов. Далее, с использованием методов порождения синтетических признаков и прогнозирования числовых переменных, включающих использование алгоритмов логических правил, ранжирования, нейронных сетей и регрессионных подходов [16], над множеством признаков K_0 строились алгоритмы, позволяющие различать тексты классов K_1/K_2 .

Разработанная система распознавания классов текстов K_1/K_2 позволяет естественным образом ввести *балльные оценки объектов* в условии (2), на основе которых оценивается «сентимент-качество» соответствующих текстов. После отнесения объекта A к классу K_1 балльная оценка этого объекта относительно класса K_1 вычисляется как $\beta(A) = |\{\alpha | p_\alpha(A, K_0) = 1, \alpha = 1, \dots, |K_0|\}|$. Иначе говоря, балл $\beta(A)$ для объекта A суть число наиболее информативных признаков, по которым объект A был отнесен к классу K_1 .

В случае настоящего исследования, признаками являются соответствующие конструкции языка или «сентименты» (т. к. подразумевается исследование эмоционально нагруженных языковых конструкций).

Было изучено 16 классов текстов и построены соответствующие алгоритмы для распознавания этих классов (табл. 1). В результате обучения алгоритмов на коллекции из 2 529 638 текстов были выделены 7 116 языковых конструкций, позволяющих с достаточной точностью распознавать классы текстов (среднее значение площади под кривой работы алгоритма – 0,88).

Таблица 1

Классы текстов с сентиментами, изученные в настоящем исследовании

№	Классы текстов	N	AUC	m	β-балл
1	Манипулятивные обороты	12 201	0,90	74	β ₁ , β _{манип.}
2	Адекватное изложение клинических результатов	15 827	0,87	25	β ₂ , β _{клин.}
3	Положительные результаты	14 394	0,84	13	β ₃ , β _{полж.}
4	Безрезультатные исследования	5 672	0,96	16	β ₄ , β _{отриц.}
5	Пропаганда	14 229	0,88	202	β ₅ , β _{проп.}
6	НКО-иноагенты	4 525	0,89	38	β ₆ , β _{НКО}
7	Подделка результатов	6 733	0,83	101	β ₇ , β _{отозв.}
8	Негативное личное отношение	18 920	0,91	132	β ₈ , β _{нег.}
9	Позитивное личное отношение	20 992	0,92	143	β ₉ , β _{поз.}
10	Сеяние раздора, агрессии, массовых беспорядков	8 830	0,86	199	β ₁₀ , β _{раздор}
11	Язык тоталитарных сект	5 221	0,84	783	β ₁₁ , β _{секта}
12	Позитивный эмоциональный фон	459 993	0,83	1790	β ₁₂ , β _{поз.эм.}
13	Негативный эмоциональный фон	475 952	0,93	1117	β ₁₃ , β _{нег.эм.}
14	Конструктивный диалог	459 993	0,82	2009	β ₁₄ , β _{констр.}
15	Неконструктивный диалог	475 952	0,89	113	β ₁₅ , β _{неконстр.}
16	Дипломатический «слэнг»	530 204	0,88	361	β ₁₆ , β _{дипл.}
—	ИТОГО	2 529 638	0,88	7 116	—

Примечания: N – число текстов в обучающей выборке; AUC – площадь под кривой как оценка точности алгоритма в кросс-валидации; m – общее число выявленных наиболее информативных признаков (языковых конструкций, сентиментов); β-балл – обозначения соответствующего β-балла (см. текст).

В качестве источников обучающих выборок были использованы собственные выборки текстов, накопленные авторами за последние 20 лет исследований (адекватное изложение клинических результатов, положительные результаты исследований, безрезультатные исследования, пропаганда, публикации с установленной фальсификацией результатов, данные о публикациях в Pubmed, поддержанных некоммерческими организациями-иноагентами, официально зарегистрированными в госдепартаментах соответствующих стран, информация с сайтов www.state.gov, www.ngoadvisor.net др.), тексты из англоязычных ресурсов, приводимых в монографиях [5–7, 17–19] и на сайте www.earthnewspapers.com (манипулятивные обороты, тоталитарные секты, раздор, агрессия, пропаганда), а также публично доступные интернет-ресурсы, контент которых представлялось возможным использовать для проведения сентимент-анализа, в т. ч. базы данных сообщений из ресурсов Twitter, IMDB, Wikileaks и др. (эмоциональный фон, конструктивность дискурса, дипломатический «слэнг»). Полученные алгоритмы отличались достаточно высокой точностью (AUC = 0,88; 95 % ДИ: 0,82–0,96) и, в среднем, позволяют идентифицировать 90 % текстов с соответствующими сентиментами с ложнопозитивными определениями всего в 8–15 %.

Общая балльная оценка (β-балл) для текста каждого исследования вычислялась как $\beta(\text{текст}) = \beta_+(\text{текст}) - \beta_-(\text{текст})$, т. е. разность между значением балльной оценкой «положительных» сентиментов $\beta_+ = \beta_2 + \beta_3 + \beta_9 + \beta_{12} + \beta_{14}$ и балльной оценкой «отрицательных» сентиментов $\beta_- = \beta_1 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \beta_{10} + \beta_{11} + \beta_{13} + \beta_{15} + \beta_{16}$ (см. табл. 1).

О соответствиях между предлагаемой методологией сентимент-анализа и идентификацией псевдонаучных текстов

Перед рассмотрением результатов скрининга базы данных PUBMED заметим, что предлагаемая математическая методология сентимент-анализа биомедицинских текстов имеет непосредственное отношение к проблеме различения «научных» и «псевдонаучных» текстов. В традиционной философии науки имеются особые критерии разграничения научного от ненаучного (псевдонаучного) и, прежде всего, критерий фальсифицируемости К.Р. Поппера [20]. Согласно последнему, утверждение относится к области науки тогда и только тогда, когда существует возможность его экспериментального или иного опровержения с использованием доступного в настоящее время научного инструментария. Если эксперт использует критерий фальсифицируемости и помещает в класс K₁ тексты, удовлетворяющие данному критерию (соответственно, в класс K₂ помещаются тексты, не удовлетворяющие этому критерию), то формулы (1–3) позволяют «об-

учать» алгоритмы для решения задачи классификации «научных» и «псевдонаучных» текстов.

Однако возможен и косвенный подход к проблеме идентификации псевдонаучных текстов — посредством анализа эмоциональной «нагрузки» текста. Хотя сам К.Р. Поппер достаточно мягко относился к проблеме псевдонаучных публикаций («...наука часто ошибается, и псевдонаука может случайно натолкнуться на истину» [20]), для здравомыслящего учёного очевидна недопустимость использования излишне эмоциональных, манипулятивных и пропагандистских sentiments в тексте научной статьи. В настоящем исследовании мы считаем, что преобладание такого рода sentiments косвенно указывает на псевдонаучность соответствующего текста — ведь излишняя эмоциональность — (1) ассоциирована с нарушениями логики, (2) может отражать скрываемый конфликт интересов, (3) указывает на необъективность индивидуума, пишущего такой текст. Эти факторы не могут не провоцировать и формирование заведомо ложных утверждений в рамках соответствующего текста и, следовательно, нарушения принципа фальсифицируемости. Поэтому, сформулированные выше количественные оценки sentiment-качества текстов ($\beta_1 \dots \beta_{16}$ в табл. 1, β_+ и β_-) могут также интерпретироваться и как количественные оценки научности/псевдонаучности текстов в соответствии со следующими критериями:

- Изложение научного результата должно использовать конструктивную лексику (β_{14}), содержать хотя бы какие-нибудь положительные результаты (β_3) и проводиться на основе специальной терминологии соответствующей научной области (β_2).

- Излишне депрессивный эмоциональный фон текста (β_- , β_4 , β_8 , β_{13}) несовместим с конструктивным изложением результатов (β_{14}).

- Эмоциональный фон научного текста должен быть эмоционально нейтральным (β_{12} , $\beta_9 \sim 0$) или слегка положительным ($\beta_+ > 0$, что интуитивно ясно, т. к. настоящий исследователь всегда радуется новому открытию).

- Научный текст в принципе не должен содержать языковых конструкций пропагандистского характера (β_5) или конструкций, указывающих выраженное "личное отношение", в особенности эмоциональное приятие/неприятие обсуждаемого результата (β_8).

- Изложение научного результата не может содержать манипулятивных приёмов изложения (β_1), в т. ч. используемых при вербовке прозелитов тоталитарных сект (β_{11}), в работе некоммерческих организаций являющихся иноагентами (β_6), при организации массовых митингов (β_3) или массовых беспорядков (β_{10}).

- Изложение научного результата не должно быть излишне эмоциональным. Направленная манипуляция эмоциями (характерная, в частности, для ретрагированных исследований, β_7), абсолютно недопустима.

- Манипулятивные языковые конструкции, повсеместно используемые политиками или средствами массовой информации (β_{15}), совершенно неприемлемы для научных текстов. Двухлинейный языковой стандарт дипломатического слэнга также недопустим (β_{16}).

Очевидно, что значения суммарного балла $\beta > 0$ для текста некоторой публикации соответствуют преимущественному выполнению этих критериев, а значения $\beta < 0$ — преобладанию соответствующих негативных и неконструктивных sentiments в тексте соответствующей публикации по биомедицине. Заметим, что разработанная система алгоритмов была обширно апробирована ранее при проведении многочисленных систематических анализов литературы по различным аспектам фундаментальной и клинической медицины и показала свою практическую эффективность (см. примеры далее).

Результаты

В результате применения комплекса алгоритмов для вычисления β -балла был проведён анализ текстов абстрактов 19,92 млн публикаций, зарегистрированных в PUBMED за последние 50 лет (1970–2019). Была изучена встречаемость 16 классов sentiments (манипулятивные обороты речи, пропаганда, негативное личное отношение, и др., см. табл. 1). В течение всего этого периода, количество публикаций с низким sentiment-качеством (β -балл текста меньше 0, что соответствует преобладанию манипулятивных и негативных sentiments) росло гораздо медленнее, чем общее число публикаций в PUBMED (рис. 2).

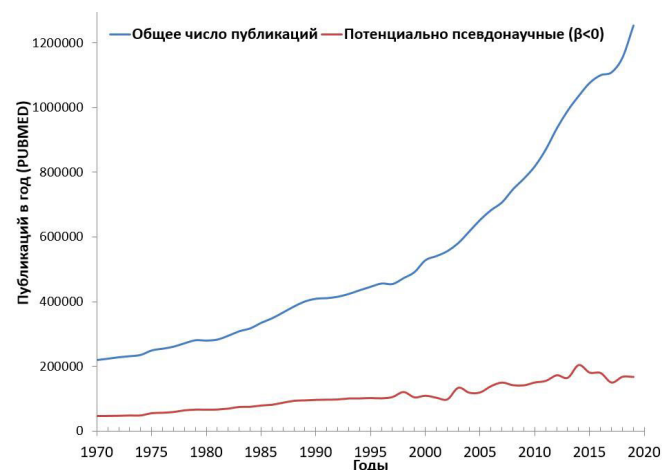


Рис. 2. Динамика публикаций в PUBMED, в т. ч. публикаций с отрицательным β -баллом

Визуальный анализ распределений значений β -балла (рис. 3) показывает, что полученный результат весьма позитивен — ведь большинство текстов публикаций, зарегистрированных в PUBMED (более 80 %), характеризуются положительными значениями β -балла.

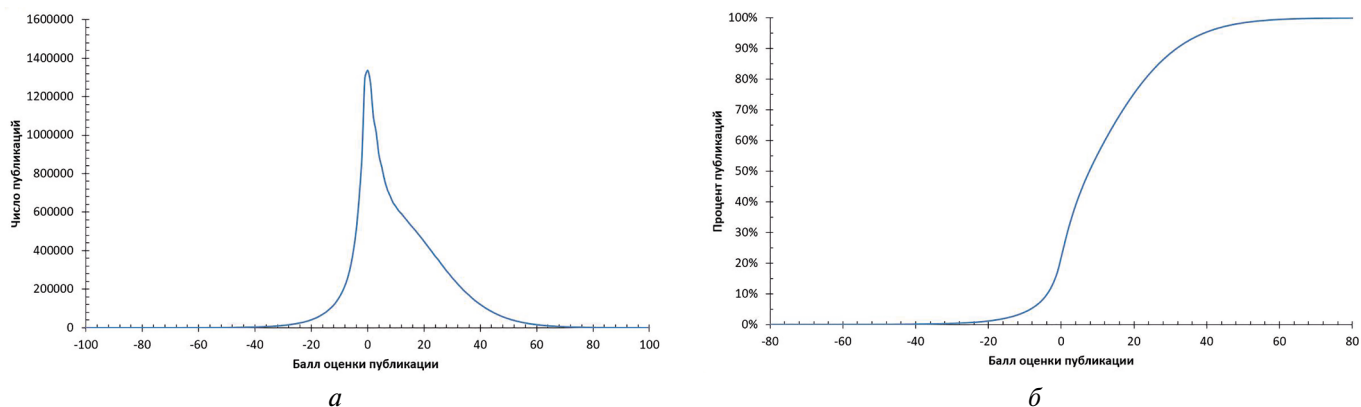


Рис. 3. Эмпирическая функция распределения (э.ф.р.) значений β -балла по 19,9 млн публикаций в PUBMED: *a* — плотность вероятности э.ф.р.; *б* — интегральная форма э.ф.р.

Анализ распределения значений β_+ и β_- (рис. 4) показывает, что большинство исследований (более 90 %, включая некоторые из исследований со значениями $\beta < 0$), характеризуются низкими значениями и балла β_+ , и балла β_- . Иначе говоря, большинство исследований в PUBMED практически не содержит манипулятивных или эмоционально перегруженных конструкций, будь то положительные или отрицательные эмоциональные коннотации. Этот факт соответствует приемлемому sentiment-качеству подавляющего числа

публикаций в PUBMED и ещё раз показывает, что *PUBMED* — ценнейший ресурс для поиска информации по фундаментальной медицине, клинической медицине и по смежным областям наук о живых системах.

Тем не менее, 18,5 % публикаций в PUBMED (3,68 млн из 19,92 млн) всё же характеризовались отрицательными значениями β -балла (рис. 5). Интересно отметить, что наибольшим вкладом в отрицательное значение β -балла этих публикаций вносили «сектантская» лексика ($\beta_{11} = 6,18$), негативные сентименты ($\beta_{13} = 4,92$) и манипулятивная лексика ($\beta_1 = 4,24$). «Сектантская» лексика включала, в частности, конструкции, основанные на терминах, подразумевающих подавление воли адептов секты, в т. ч. наведение жёсткой дисциплины тоталитарного характера и нетерпимость к «уклонистам» («personnel must be», «necessary to use», «aberrations», «mandatory», «staff management», «staff clearance», «training program», «mental training» и др.). Конечно, по отдельности такие термины могут использоваться и во вполне нормальных статьях по тренировке медицинского персонала или по психологии. Однако определённые сочетания этих терминов, наряду с избыточно частым их употреблением, распознаются разработанным алгоритмом анализа текстов именно как языковые шаблоны, характерные для менеджмента адептов тоталитарной секты, а вовсе не для тренировки медицинского персонала.

Удивительно, что лингвистические конструкции на основе терминов явно публицистического характера (которые, к тому же, широко используются организаторами массовых беспорядков, балл β_{10} в табл. 1) также нередко встречаются в предположительно «научной» литературе ($\beta_{10} = 2,89$): «resistance forces» (силы сопротивления), «political» в сочетании с «regressions» (политические репрессии), «struggle against» (борьба против [режима]), «attack» в связке с «military» (атаковать/атакуйте военнослужащих) и т. п. При этом в статье могут обсуждаться вовсе не какие-то военные операции или явно противоправные действия, а казалось бы совершенно другие темы (см. левую колонку табл. 2).

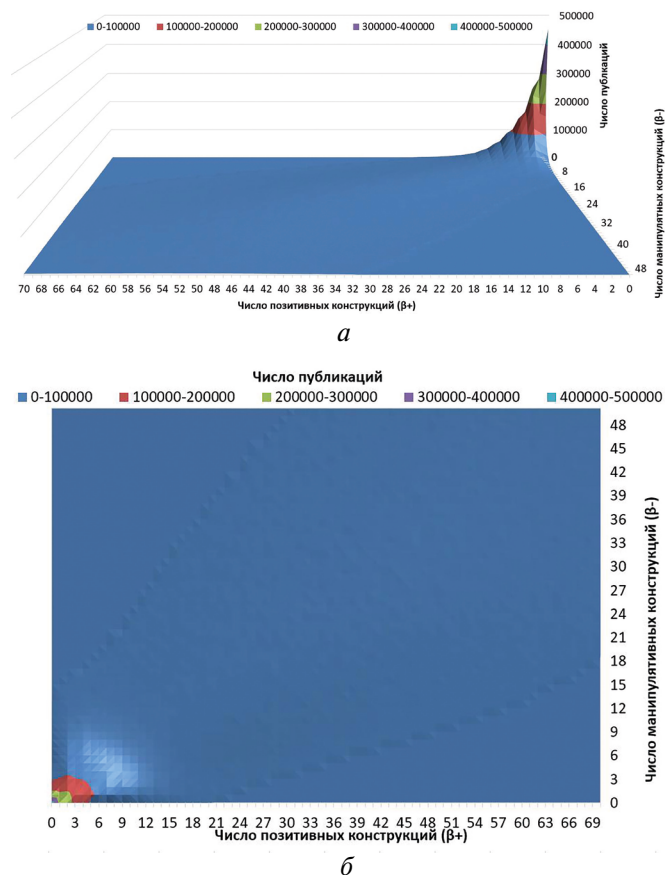


Рис. 4. Распределение значений β_+ и β_- по 19,9 млн публикаций в PUBMED: *a* — вид сбоку; *б* — вид сверху

Негативные сентименты ($\beta_{13} = 4,92$) также были весьма типичны для текстов с $\beta < 0$. Действительно, чрезмерно частое употребление конструкций на основе таких слов и словосочетаний как «waste» (понапрасну), «is poorly» (есть плохо), «the worst» (наихудшее), «poor quality» (очень низкого качества), «a worse» (ещё хуже), «zero potential» (нулевой потенциал), «has to avoid» (должно избегать) и других может и в самом деле целенаправленно «портить настроение» читающего по отношению к затрагиваемой в тексте публикации научной теме.

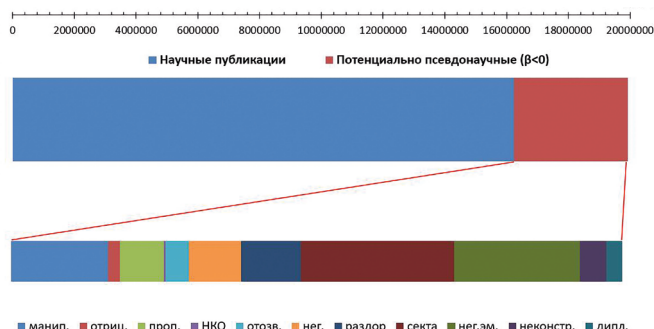


Рис. 5. Вклад различных негативных сентиментов в формирование отрицательных значений β -балла для 3,68 млн из 19,92 млн публикаций в PUBMED
Примечание: Длины прямоугольников на нижней полосе пропорциональны значениям соответствующих β -баллов (см. табл. 1)

Анализ стран, из которых исходят англоязычные публикации по биомедицине (рис. 6) показал, что наибольшими значениями β -балла характеризовались публикации из ОАЭ ($\beta = 15,46$), Ирана ($\beta = 13,67$) и Тайваня ($\beta = 12,14$), затем шли южная Корея, Сингапур, Кувейт и Иордания ($\beta = 10-11$) и отдельные европейские страны (Греция, Словения, Нидерланды, Швейцария, Ирландия, $\beta = 9-11$). Таким образом, исследователи из этих стран наименее склонны использовать манипулятивную и эмоционально-перегруженную лексику при написании научных статей.

Рубрикация всего массива публикаций по 27 840 рубрикам MESH-системы PUBMED позволила выявить рубрики, наиболее затронутые и наименее затро-

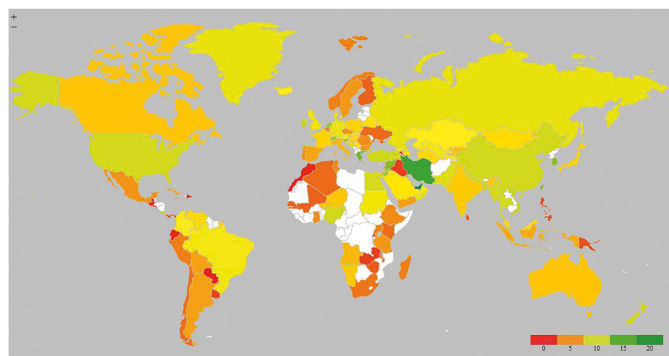


Рис. 6. Распределение значений β -балла по странам (среднее за последние 50 лет, 1970–2019 гг.)

нутые манипулятивными сентиментами. Практически все 3,68 млн текстов с отрицательными значениями β -балла попали в 3 520 из 27 840 рубрик PUBMED. Наиболее характерные (если не сказать, одиозные) примеры таких рубрик приведены в табл. 2.

Манипулятивная англоязычная лексика наиболее распространена в публикациях, рубрицированных под разделами «экономика» ($\beta = -17,15$) и «юриспруденция» ($\beta = -12,60$). Весьма «чувствительны» в этом отношении числовые данные по медицинской статистике западных стран ($\beta = -6,41$), в т. ч. рассматривающие вопросы стоимости лекарств ($\beta = -3,86$), перспективы т. н. «планирования семьи» ($\beta = -3,61$), легализации эвтаназии ($\beta = -1,53$) и других этически порочных практик.

Чрезвычайно интересно отметить, что даже в случае явных изданий типа «жёлтой прессы», каким-то образом попавших в PUBMED (например, «Нью Йорк Таймс», рубрика «SO – N Y Times Web») среднее значение β -балла было намного выше ($\beta = -0,64$), чем в случае отдельных «научных» журналов (табл. 3).

В то же время наибольшими положительными значениями β -балла отличались публикации по спортивной медицине («Resistance Training», «Muscle Strength/*physiology», «Athletic Performance», «Energy Intake/*physiology», «Physical Endurance/*physiology»), системной биологии («Systems Biology/*methods», «Metabolomics») и нутрициологии («Dietary Supplements/analysis»), биофизике («Molecular Dynamics Simulation», «Nonlinear Dynamics», «Spectroscopy, Fourier Transform Infrared/*methods»), математическим методам в биологии и медицине («Finite Element Analysis», «Computational Biology/*methods», «Principal Component Analysis»), в т. ч. математическим методам «искусственного интеллекта» и «машинного обучения» («Artif Intell Med», «Pattern Recognition, Automated/*methods», «Neural Networks (Computer)», «Problem-Based Learning», «Artificial Intelligence»).

Представляет интерес рассмотреть, как изменялась динамика β -балла публикаций по годам по каждой из исследованных 27 840 рубрик PUBMED. Для этого временные ряды в координатах «год– β -балл» для каждой из рубрик были представлены в виде символьных последовательностей в рамках подхода символьной динамики [21]. Парные расстояния между полученными символьными последовательностями вычислялись с использованием метрики Левинштейна [22] и была проведена кластеризация на основе метода анализа метрических сгущений [8, 12–14]. В результате проведения кластеризации положительная динамика сентимент-качества исследований (т. е. систематическое повышение β -балла по годам) было установлено для 27 090 из 27 840 исследованных рубрик (97,3 %). Наиболее интенсивная положительная динамика найдена для исследований по генетике, физиологии, фармакологии и геронтологии (рис. 7, а). Выделены 249 рубрик с резко отрицательной динамикой каче-

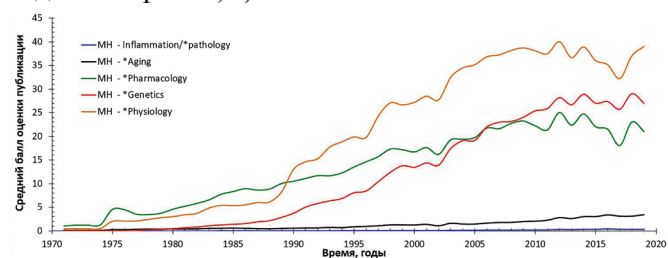
Таблица 2

Примеры рубрик БД PUBMED с наиболее отрицательными и наиболее положительными значениями β -балла по всем публикациям в данной рубрике за последние 50 лет (1970-2019 гг.)

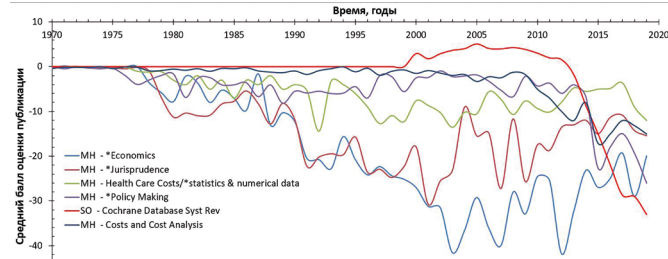
Рубрика	Количество статей	β -балл	Рубрика	Количество статей	β -балл
MH – *Economics	7 855	-17,15	MH – *Resistance Training	2 282	28,94
MH – *Jurisprudence	13 221	-12,60	MH – Muscle Strength/*physiology	4 561	28,81
MH – Health Care Costs/*statistics & numerical data	4 225	-6,41	MH – Systems Biology/*methods	2 249	26,03
MH – Mental Health Services/*economics	871	-4,30	MH – *Molecular Dynamics Simulation	8 896	25,77
MH – *Drug Costs	2 671	-3,86	MH – Learning/*physiology	10 886	24,70
MH – *Policy Making	4 876	-3,79	MH – *Metabolomics	2 106	23,06
SO – Fam Plann Perspect	1 389	-3,61	SO – Artif Intell Med	1 309	23,05
MH – *Insanity Defense	918	-2,73	MH – Athletic Performance/*physiology	3 640	22,82
MH – Health Policy/*legislation & jurisprudence	3 943	-2,59	MH – Maternal Behavior/*physiology	937	22,79
MH – Commitment of Mentally Ill/*legislation & jurisprudence	2 055	-2,29	MH – Dietary Supplements/analysis	1 095	22,69
MH – Homosexuality, Male/*statistics & numerical data	1 163	-2,22	MH – Maze Learning/*physiology	2 007	22,17
MH – *Medication Errors	2 358	-2,07	MH – Drug Resistance, Multiple, Bacterial/*genetics	1 909	22,12
MH – Condoms/*statistics & numerical data	2 582	-2,04	OT – NAFLD	1 148	21,83
SO – Cochrane Database Syst Rev	12 405	-1,86	OT – BDNF	1 905	21,79
MH – Mental Competency/*legislation & jurisprudence	1 250	-1,70	MH – Proteomics/*methods	14 267	21,41
MH – Developing Countries/*economics	916	-1,58	MH – Muscle, Skeletal/*physiology	16 346	20,98
MH – Suicide, Assisted/*legislation & jurisprudence	1 206	-1,53	MH – *Finite Element Analysis	3 408	20,86
MH – *Population Control	1 326	-1,48	MH – Pattern Recognition, Automated/*methods	14 042	20,84
MH – *Abortion, Legal	2 394	-1,33	MH – Gait/*physiology	8 099	20,78
MH – HIV Infections/*transmission	2 754	-1,23	MH – *Gene Expression Profiling	21 965	20,71
MH – Violence/*legislation & jurisprudence	108	-1,28	MH – Tandem Mass Spectrometry/*methods	11 254	20,70
MH – *Conflict of Interest	3 680	-1,03	MH – Physical Endurance/*physiology	5 745	20,70
MH – Costs and Cost Analysis	34 779	-0,94	MH – Computational Biology/*methods	17 691	20,51
MH – Child Abuse/*legislation & jurisprudence	1 214	-0,88	MH – *Principal Component Analysis	986	20,32
MH – *Lobbying	2 123	-0,83	MH – *Droughts	2 520	20,10
MH – *Fraud	1 495	-0,69	MH – Artificial Intelligence	5 325	19,59
MH – *Scientific Misconduct	2 196	-0,67	MH – Nonlinear Dynamics	5 409	19,41
SO – N Y Times Web	2 170	-0,64	MH – Caregivers/*psychology	11 592	19,24
MH – Marketing of Health Services/*methods	1 882	-0,62	MH – Spectroscopy, Fourier Transform Infrared/*methods	2 802	19,24
MH – Civil Rights/*legislation & jurisprudence	1 724	-0,61	MH – *Neural Networks (Computer)	14 038	19,24
MH – *Sterilization, Reproductive	1 674	-0,59	MH – Energy Intake/*physiology	2 905	19,11
MH – *Peer Review, Research	1 597	-0,58	MH – *Problem-Based Learning	2 226	19,09

Рубрика	Количество статей	β-балл	Рубрика	Количество статей	β-балл
MH – *Abortion, Induced	7 870	-0,54	SO – Proteomics	5 841	19,06
MH – Fees, Dental	1030	-0,53	MH – Thermodynamics	22 139	19,05
MH – Evidence-Based Medicine/*standards	1 392	-0,52	MH – Soccer/*physiology	1 543	19,04
OT – *Family Planning Programs	2 348	-0,44	MH – Knee Joint/physiology	2 528	18,96
MH – *Euthanasia	3 079	-0,40	MH – Genome-Wide Association Study/*methods	1 825	18,85
MH – *Embryo Research	878	-0,31	SO – FEBS J	5 728	18,83
MH – State Medicine/standards	1 088	-0,30	MH – Gastrointestinal Tract/*microbiology	2 647	18,82
MH – *Psychotropic Drugs	1 690	-0,27	MH – Insulin Resistance	7 251	18,79
ИТОГО	149 420	-2,23	ИТОГО	255 052	21,27

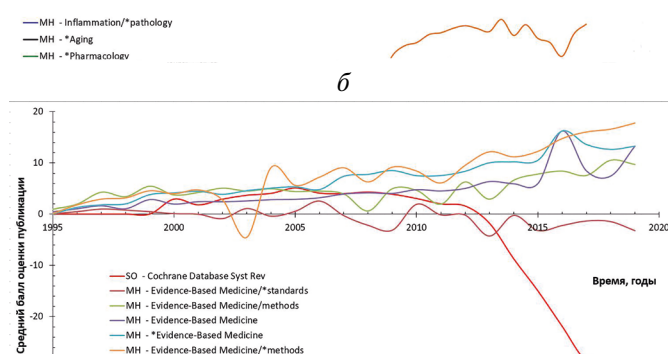
ства и с выраженным нарастанием манипулятивных сентиментов, подобных встречающимся в «жёлтой прессе» западных стран (отдельные примеры приведены на рис. 7, б).



а



б



в

Рис. 7. Примеры динамики β-балла по различным категориям PUBMED:

а — примеры положительной динамики; б — примеры отрицательной динамики; в — примеры из области доказательной медицины

Наиболее характерной и несколько неожиданной была динамика рубрики «SO – Cochrane Database Syst Rev», которая ассоциирована, очевидным образом, с проблематикой доказательной медицины. В самом деле, до 2008 года β-балл публикаций под этой рубрикой колебался вблизи значений +1,5...+2,2. После 2008 года началось постоянное снижение значений β-балла, так что в 2013 г. средний β-балл публикаций в этой рубрике приобрёл отрицательное значение. После 2013 г. началось крайне резкое падение значений β-балла (на 5,1 балла/год), так что к концу 2019 года среднее значение β-балла по этой рубрике составит -32,0. В то же время почти все остальные другие рубрики, имеющие непосредственное отношение к доказательной медицине («Evidence-Based Medicine/*methods», «Evidence-Based Medicine/*education», просто «*Evidence-Based Medicine» и др.) характеризовались отчётливой положительной динамикой (рис. 6В). Исключение составила рубрика «Evidence-Based Medicine/*standards», в которой также отмечена отрицательная динамика, хотя и не столь выраженная.

Столь резкое изменение сентимент-качества публикаций в двух упоминаемых выше рубриках позволяет говорить о дискредитации доказательной медицины под видом разработки т. н. «международных стандартов доказательной медицины» и полностью соответствует сформулированной ранее концепции захвата области доказательной медицины т. н. «пиратами» (термин работы [23]). В работе [23] проф. Ioannidis JPA с соавторами отмечают, что «доказательная медицина платит цену своего успеха: получив более широкое признание, ею манипулируют и злоупотребляют для поддержки извращенных «повесток дня», которые подрывают репутационную ценность доказательной медицины. Иногда конфликты, стоящие за этими повестками дня, настолько сильны, что можно беспокоиться о том, является ли этот рейдерский захват доказательной медицины обратимым» [23]. По результатам настоящего исследования можно предположить, что обсуждаемый

Таблица 3

Примеры отдельных англоязычных научных журналов с наиболее отрицательными и наиболее положительными значениями β -балла по всем публикациям за последние 50 лет (1970–2019 гг.)

Журнал	Количество статей	β -балл
SO – EBRI Issue Brief	214	-13,78
SO – Reprod Freedom News	143	-7,91
SO – Evid Rep Technol Assess (Full Rep)	73	-7,16
SO – Harv Bus Rev	1 357	-5,85
SO – Can HIV AIDS Policy Law Rev	395	-4,90
SO – Annu Rev Popul Law	589	-4,22
SO – J Am Health Policy	204	-3,74
SO – Fam Plann Perspect	1 389	-3,61
SO – Lancet HIV	311	-2,74
SO – Conscience	103	-2,21
SO – Bioethics	1 394	-2,07
SO – Cochrane Database Syst Rev	12 405	-1,86
SO – AIDS Treat News	808	-1,69
SO – AIDS Policy Law	4 225	-1,65
SO – N Engl J Hum Serv	48	-1,40
SO – Health Law Vigil	748	-1,11
SO – Evid Based Dent	842	-1,10
SO – Issues Reprod Genet Eng	54	-0,94
SO – Harv Health Lett	1 635	-0,84
SO – Autism Res	587	21,14
SO – Gigascience	237	21,27
SO – J Proteome Res	6 573	21,54
SO – Database (Oxford)	633	21,71
SO – Clin Epigenetics	390	21,85
SO – Wiley Interdiscip Rev RNA	429	21,87
SO – Ecology	3 693	22,05
SO – Fam Syst Health	351	22,09
SO – Birth Defects Res C Embryo Today	361	22,11
SO – Cell J	374	22,16
SO – Clin Proteomics	171	22,36
SO – Iran J Microbiol	381	22,61
SO – BMC Geriatr	1 225	22,82
SO – Eur J Sport Sci	686	22,84
SO – Nutr Diabetes	250	22,89
SO – Sports Med	2 396	23,84
SO – Front Plant Sci	5 955	23,85
SO – BMC Biotechnol	1 207	23,99
SO – J Comput Phys	59	24,49
SO – BMC Microbiol	3 104	24,92
SO – PLoS Comput Biol	4 966	25,19
SO – Metabolomics	310	27,39
SO – BMC Genomics	9 555	28,11
SO – Microbiome	321	30,80
SO – BMC Syst Biol	1 714	35,97

«рейдерский захват» произошёл в районе 2013–2014 годах (см. рис. 6, в).

Такого рода «медицинские пираты», зачастую, характеризуются полным отсутствием реальной клинической практики и, тем не менее, обладают существенным медийным ресурсом для навязывания своих произвольных мнений сотням тысяч реальных врачей во всем мире [24]. Поэтому ряд ведущих аналитиков вполне обоснованно утверждают, что в настоящее время метаанализы по типу кохрейновских не являются сколько-нибудь доказательными хотя бы вследствие повсеместного использования устаревших и, более того, ошибочных методов анализа данных [25]. Принимая во внимание, что в Австралии, Канаде, ЕС и США официально зарегистрирована торговая соответствующая марка [26], и что текстовые продукты под данной торговой маркой всё ещё пользуются популярностью среди врачей, «пираты», упоминаемые выше, действительно захватили весьма ценный властный ресурс. При этом резко упал уровень дискурса: в отличие от сдержанного и более или менее положительного дискурса до 2008 года ($\beta = 1,5-2,2$) дискурс большинства публикаций буквально «упал» до выражено манипулятивного нагнетания негативных эмоций, агрессии, сектантской нетерпимости и т. п. ($\beta = -32,0$ в 2019 году).

Типичным примером «текстового продукта», произведённого после этого «захвата», является опубликованная в 2018 г. статья под кодовым номером «CD003177». Данная статья, детально проанализированная нами ранее [27], является своего рода «лидером» среди публикаций с отрицательным β -баллом ($\beta = -42$, в PUBMED таких статей менее 0,5 %, см. распределение на рис. 2). При этом числа манипулятивных сентиментов ($\beta_1 = 55$), сентиментов, отражающих негативное личное отношение ($\beta_8 = 35$), негативный эмоциональный фон ($\beta_{13} = 37$), раздор/агрессию ($\beta_{10} = 23$) и «сектантскую» лексику ($\beta_{11} = 32$), даже взятые по отдельности, существенно превышают суммарный отрицательный балл (β_-) подавляющего большинства (99,98 %) из 19,9 млн исследованных публикаций (рис. 3, б).

Экспертный анализ текста «CD003177» указал на грубейшие нарушения, допущенные при сборе и анализе данных, включая отсутствие чёткой клинической логики при построении метаанализа, анализа клинической неоднородности когорт и незнание простейших основ фармакологии и фармации (например, различие между лекарственным препаратом, БАДом и пищевым продуктом) [27]. Сделанные в работе [28] выводы о «максимально возможном игнорировании основ фармакологии, биохимии и эпидемиологии на фоне догматизации шаблонных моделей мышления» весьма органично подтверждаются чрезвычайно высокими значениями баллов «сектантства», «раздора/агрессии» и негативного личного отношения для данной публикации (см. выше).

Описанные проблемы со «стандартами доказательной медицины» отнюдь не ограничены обсуждаемой выше проблематикой (рубрика «Evidence-Based Medicine/*standards» и т. п.). Используемая методология сентимент-анализа биомедицинских текстов указывает, в частности, и на такой аспект использования негативных языковых штампов под прикрытием «доказательности», как недобросовестная конкуренция.

Языковые манипуляции как средства недобросовестной конкуренции против микронутриентных препаратов

В ходе упоминаемых нами ранее анализов публикаций для написания ряда монографий по нутрициологии мы убедились в существовании огромного массива реальных доказательных исследований, указывающих на необходимость компенсации витаминно-минеральных дефицитов для эффективной профилактики многочисленных хронических патологий. Полученные в настоящей работе оценки результативности десятков тысяч исследований позволяют утверждать, что, начиная с 2012 года, в англоязычной медицинской литературе против микронутриентных препаратов ведётся настоящая «гибридная война» (см. примеры далее).

И фундаментальные, и клинические результаты научных исследований не дают никакого сомнения в том, что для сохранения здоровья человек нуждается в ежедневном потреблении, как минимум, 33 витаминов и микроэлементов. К сожалению, технология выращивания растений и животных в промышленных масштабах направлена на достижение максимальных объёмов продукции в кратчайшие сроки. В результате, витамины в современных продуктах питания не успевают накопиться. Отметим, что разговоры о том, что «раньше жили лучше без всяких витаминов» основаны на полном незнании статистики длительности жизни, материнской и младенческой смертности в прошлом, особенно в средневековье и даже в начале 20-го века.

Дефициты витаминов и микроэлементов существуют в самых разных странах: и в «бедных», и в «богатых», где выбор продуктов весьма широк. Например, крупномасштабное исследование женщин репродуктивного возраста показало, что суточное потребление витаминов часто не достигает нормы. Ситуация и в России, и в западной Европе приблизительно одинакова: одновременно всеми жизненно необходимыми микронутриентами обеспечено менее 10 % населения и в России, и в странах ЕС [29] (рис. 8).

Потребление необходимого числа калорий не означает потребления необходимого количества микронутриентов. В результате, развиваются полигиповитаминозы (т. е. дефициты сразу нескольких микронутриентов). Метрический анализ данных исследований [29, 30] показал, что если эти сочетанные дефициты витаминов и микроэлементов продолжают годами, то многократно возрастает риск коморбидных

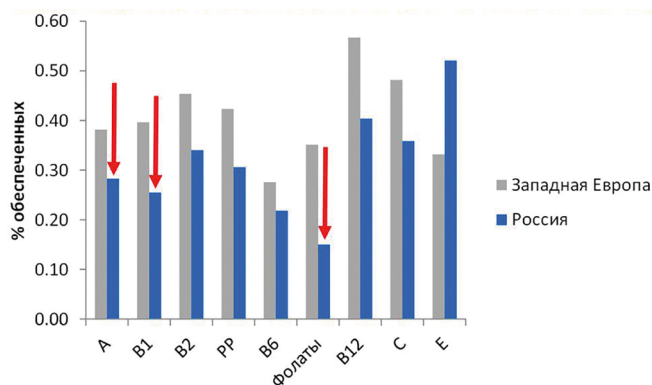


Рис. 8. Обеспеченность витаминами женщин репродуктивного возраста ($n = 2141$)

патологий: атеросклероза, тромбоза, артериальной гипертонии, инфаркта, инсульта, накопления токсических металлов, опухолей, ожирения и диабета (рис. 9). Всё это приводит к ускорённому старению и к сокращению длительности жизни.

Результаты цитируемого выше исследования, равно как и результаты десятков тысяч других исследований, указывают на существование дефицитов витаминов и связанные с этим риски для здоровья. Тысячи исследований показывают, что компенсация дефицитов витаминов эффективна для профилактики

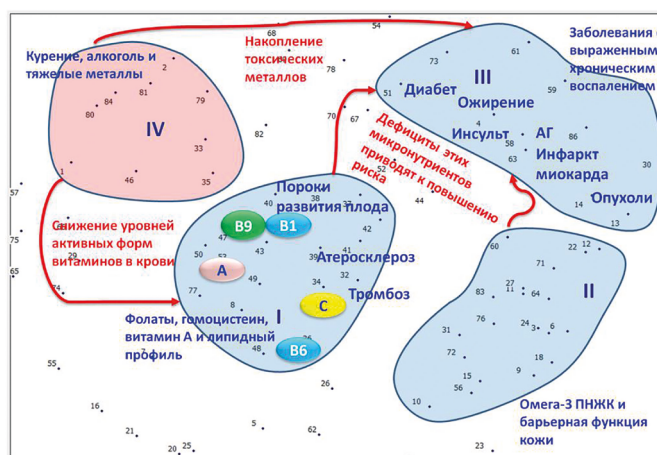


Рис. 9. Метрический анализ взаимосвязей хронических коморбидных патологий и микронутриентных дефицитов

практически каждой из т. н. «болезней цивилизации». Поэтому возникает вопрос: а почему столь важная информация (доступная, кстати, в огромном количестве в специализированной научной литературе) всячески вытесняется из медийного пространства? Тому есть несколько причин.

Во-первых, результаты настоящего сентимент-анализа текстов из PUBMED позволяют утверждать о резкой активизации скоординированных медийных атак на витамины и микронутриенты, которые проводятся особым «анти-витаминным» лобби. Очевидно, что чем больше больных, тем больше рынок для

сверх-дорогих лекарств. Поэтому, дотации витаминов и микроэлементов, являющиеся экономически выгодными и клинически эффективными методами борьбы с хроническими заболеваниями в масштабах популяций, совершенно не выгодны для продавцов сверхдорогих и весьма токсичных лекарств. При этом это именно скоординированная атака, т. к. включает далеко не только истерические публикации в жёлтой прессе, но и якобы «научные» публикации в престижных англоязычных журналах.

Приведём пример по фундаментальной и клинической медицине витамина D₃ [31]. Профилактика дефицита витамина D₃ действительно является чрезвычайно экономически выгодным способом снижения нагрузки населения хронической патологией. Помимо многочисленных экспертных оценок, приводимых в монографии [31], в крупномасштабном исследовании ($n = 7217$) была прямо продемонстрирована корреляция между более низким уровнем 25(OH)D₃ в крови (<20 нг/мл) и более высокими затратами на содержание пациента в стационаре ($p = 0,001$) [32].

Тем не менее, разработанная система сентимент-анализа продолжает регистрировать вал «безрезультатных» исследований витамина D₃ ($\beta < 0$, $\beta_4 > 10$), характеризующихся характерной шаблонной лексикой. Это и термины жёсткого отрицания клинических эффектов витамина D₃ («not», «no», «did not», «no significant difference», «any evidence», «no use of» и др.), употребляемых с аномально высокой частотой наподобие некой ритуальной мантры. Это и чересчур частое употребление словечка «все» («ALL»). Использование данного слова должно быть весьма осторожным: ведь всем известен парадокс критянина Эпименида («Все критяне — лжецы»). Поэтому, аномально частое употребление слова «все» указывает на необоснованное стремление к обобщению, причём не имея на то никаких оснований. С психологической точки зрения, неправомерно частое использование слова «все» может указывать на избыточную эмоциональность, дефицит логического мышления или же на преднамеренные попытки эмоционального давления на читателя, предпринимаемое под прикрытием авторитета «доказательной медицины».

В целом, анализ базы данных PUBMED показал, что с 2013 по 2018 гг. резко увеличилось число англоязычных «безрезультатных» исследований по витамину D, в которых приём витамина D, по заявлениям авторов, «ни к чему не приводит» (рис. 10). Если посмотреть, кто же публикует эти, так сказать, «исследования», то почти в каждом случае авторы имеют ярко выраженный конфликт интересов.

Есть весьма курьёзные примеры конфликта интересов, которые выглядят просто как моральный урок (рис. 11). Например, в Австралии и в Новой Зеландии результативность исследований эффектов витамина D близка к нулю: только одно из сорока исследований в Новой Зеландии показало, что витамин D по-

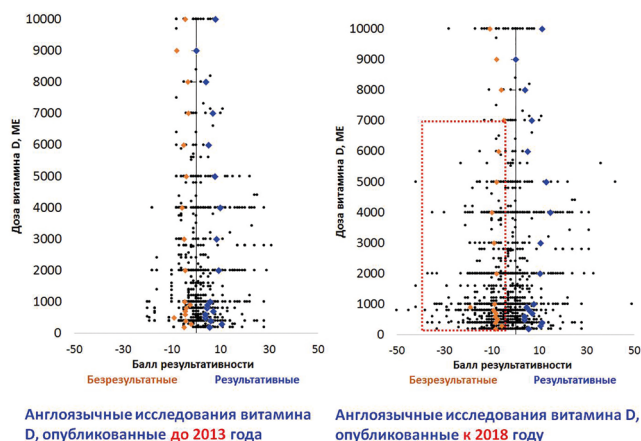


Рис. 10. Динамика β -балла клинических исследований эффектов применения препаратов витамина D: стратификация по дозам колекальциферола

лезен (в Австралии — четыре из пятидесяти шести). Поскольку эти две страны — абсолютные лидеры по безрезультатности исследований витамина D, то, наверное, на австралийском континенте действуют какие-то другие законы физики, химии и биологии. Уже не говоря о сильнейшем лобби молокозаводчиков (мировое лидерство по производству сухого молока). Мы обсудили эту взаимосвязь ранее [33]. Кстати, крайне любопытно отметить, что языковые штампы «необходимы дальнейшие исследования» и «источник финансовой поддержки...» являются характерными особенностями именно безрезультатных исследований по витамину D₃.

К сожалению, медийная «гибридная война» против микронутриентов усугубляется ещё и вследствие гипостопсового невежества, а также нежелания разбираться в чём-либо. А зачем заниматься анализом больших данных? А зачем делать какой-то там сентимент-анализ публикаций, которые печатаются в «очевидно престижных журналах», из «списка Web of Science», причём с «высоким рейтингом цитирования»? Зачем столь детально разбирать результаты клинических исследований? Конечно, гораздо проще неким актёрам средней руки, которые наряжаются выдающимися

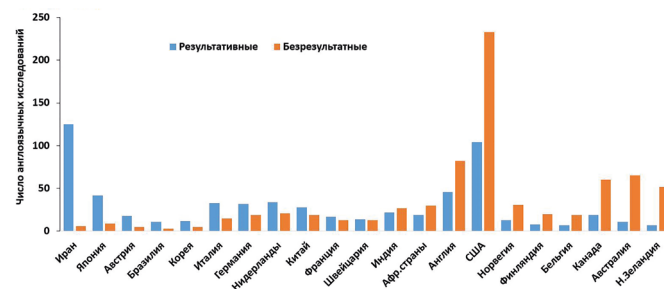


Рис. 11. Результативность англоязычных клинических исследований витамина D₃ по различным странам

Примечание: Страны упорядочены по убыванию индекса результативности (О.Ш. числа результативных и безрезультатных исследований)

докторами, категорично заявлять на телевидении, что «витамины не нужны», «доказательная медицина доказала, что витамины опасны», «витамины приносят выгоду только их производителям» и т. п.

Тем не менее, предлагаемая в настоящей работе методология sentiment-анализа биомедицинских исследований позволяет если не противодействовать, то, по крайней мере, идентифицировать медийные атаки против микронутриентов, тем самым нивелируя их эффективность. Очевидно, что адекватный исследователь не будет доверять «штампованным» метаанализам с престижной маркой и с очевидными элементами эмоциональной манипуляции.

Предлагаемая методология sentiment-анализа может быть эффективно использована для разработки методических пособий и национальных рекомендаций по медицине. С использованием данной методологии sentiment-анализа были разработаны, в частности, российские национальные рекомендации по витамину D [34].

Другой пример: нами был проведён анализ текстов англоязычных публикаций, цитируемых в некоторой неопубликованной методичке. Из 169 ссылок на статьи, цитируемые в анализируемом тексте методички, публикации с отрицательным β -баллом составили 40 (т. е. 24 %). При этом были выявлены не менее одиозные примеры наиболее манипулятивных публикаций, чем упоминаемый выше текст «CD003177» [27, 28]. В частности, значение β -балла для текстов «CD004909» и «CD009154» составило $\beta = -95$ (публикаций со столь низкими значениями β -баллов менее 0,01 % среди 30 млн публикаций в PUBMED). Большинство цитируемых публикаций по использованию микронутриентов при беременности являлись манипулятивными (значения β -балла в диапазоне $-40 \dots -8$). Анализ эмпирических функций распределения значений β -баллов,

полученных при анализе текста рассматриваемой методички, указывает на статистически достоверное ($p = 0,004$ по тесту Колмогорова–Смирнова) смещение цитируемых публикаций в область манипулятивных (рис. 12). В настоящее время текст методического пособия проходит существенную переработку.

О медийный войнах, недопинге и скандале на пустом месте

*"Громче всех кричит
«ДЕРЖИ ВОРА!» сам вор..."*

Русская народная поговорка

Говоря о враждебных действиях в медийном пространстве, нельзя не упомянуть о нашумевшем скандале, связанном с якобы употреблением российскими спортсменами якобы допинговых средств. Даже с точки зрения профессионального фармаколога, весь это скандал, очевидным образом, является чёрным пиаром российского спорта – ведь в качестве допинга назывались такие средства как мельдоний (антигипоксанта), переливание крови и т. п. Применение описанной выше методики sentiment-анализа позволило выявить ряд псевдонаучных и высокоманипулятивных публикаций, которые имеют отношение к спортивной медицине.

Несмотря на отмеченную ранее позитивную динамику бета-балла по спортивной медицине, разработанные алгоритмы позволяют идентифицировать публикации в данной области, включающие неправомерные обобщения и откровенную манипуляцию простейшими фактами. Например, алгоритм указал на одну из публикаций неизвестного ныне г-на Родченкова как характеризующуюся чрезвычайно высоким уровнем лингвистической манипулятивности (β -балл = -15), нетипичным для публикаций по спортивной медицине (менее 0,04 % публикаций по спортивной медицине характеризовались таким значением β -балла).

Не затрагивая достаточно короткий текст самой публикации [35], уже сам абстракт публикации заставляет насторожиться читателя, не готового к столь неуместной манипуляции сознанием: здесь и подмена понятий, и нарушения логики, и очевидная презумпция виновности российских спортсменов, и пренебрежение фундаментальными основами математической статистики и др.

Абстракт текста [35] начинается с утверждения, что «гомологическое переливание крови – это запрещённый метод манипуляций с кровью». Однако гомологическое переливание крови (профессионалы, вообще-то, говорят, «инфузионно-трансфузионная терапия») – этот общепринятый метод лечения! В статье же авторы имеют в виду т. н. «переливание крови от спортсмена к спортсмену» (не по медицинским показаниям, о чём авторы, опять же, умолчали) –

Эмпирическая функция распределения значений бета-баллов цитируемых публикаций

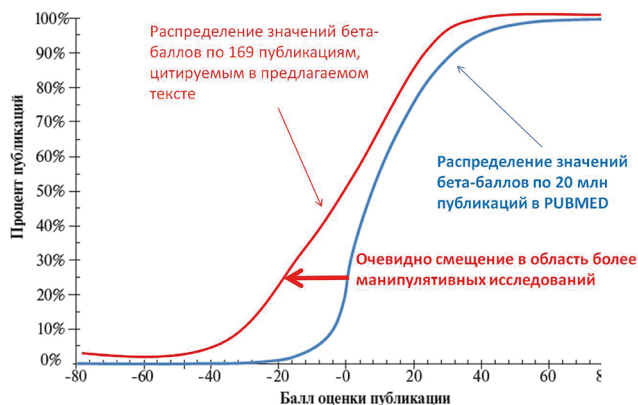


Рис. 12. Анализ эмпирических функций распределения значений β -баллов, полученных при анализе текста некоторого методического пособия

не вполне легальной псевдомедицинской процедуре, якобы улучшающей снабжение тканей кислородом. Нельзя не подчеркнуть, что последнее утверждение не имеет никакой доказательной базы, т. к. не было проведено и не было опубликовано никаких рандомизированных исследований по переливаниям крови спортсменам-олимпийцам с последующей оценкой спортивных результатов.

Далее авторы текста [35] утверждают, что «за последние шесть лет не было зарегистрировано никаких негативных аналитических данных, касающихся гомологичных переливаний крови». Однако в тексте тут же формулируется презумпция виновности российских спортсменов, т. к. в следующем предложении появляется утверждение того, что переливания-то, мол, происходят, но российские спортсмены якобы «проводят тщательный отбор потенциальных доноров». Затем, авторы формулируют предположение, что "любой спортсмен может быть потенциальным донором для партнеров по команде с тем же фактором RhD и группой крови АВ" и утверждают, что при анализе фенотипа эритроцитов у 535 российских спортсменов было обнаружено «несколько» (не указано сколько) пар спортсменов с одинаковым фенотипом.

Таким образом, только текст абстракта содержит существенное количество манипулятивных штампов, связанных с преднамеренными или с непреднамеренными нарушениями логики. Текст статьи позволяет сделать ещё большее количество «открытий». Заметим, что в статье подробнейшим образом приведены совершенно лишние данные по сочетаниям признаков RhD/AB по разным видам спорта (в 2 таблицах статьи, состоящей из 2 страниц) и практически полностью сокрыты реально важные данные о якобы исследованных маркёрах фенотипов эритроцитов (С, с, Е, е, Jka, Jkb, Fya, Fyb).

Несмотря на очевидное сокрытие важнейших данных по маркёрам эритроцитов авторы [35] смело делают вывод о том, что, дескать, «на основе частоты встречаемости антигенов эритроцитов была рассчитана теоретическая вероятность обнаружения донора в команде с точно таким же фенотипом». При этом никакая вероятность не была рассчитана (авторам следовало бы изучить математическую статистику перед тем, как использовать данный термин), а была сосчитана некая волюнтаристски определённая частота встречаемости идентичного фенотипа (причём с ошибками в приводимой формуле). После этого авторы утверждают, что «фактическое число случаев, когда два человека имеют один и тот же фенотип в одном и том же виде спорта, было в пять раз выше теоретической вероятности». Данное утверждение вообще не выдерживает никакой критики, т. к. авторы даже не удосужились оценить статистическую достоверность полученного ими различия в частотах (потому и написали «несколько пар спортсменов», т. к. приведение конкретного числа спортсменов сразу

же поднимает вопрос о статистической достоверности основного вывода «исследования»), уже не говоря о том, что авторы подменили строго определяемое математическое понятие «вероятность» понятием «частота».

Заметим, что при попытке оценить статистическую достоверность, например, по критерию Пирсона, различия в частотах могут достигать и 4, и 5, и 6 раз, но не быть статистически достоверными. Текст и абстракта, и самой статьи [35] перегружен англоязычными лингвистическими маркёрами лингвистической манипуляции наподобие «was in fact», «prohibited method», «no adverse», «this kind of», «would be a more», «would fail» и многими другими, так что β_- -балл статьи составил 51 (практически, абсолютный рекорд для статей по спортивной медицине – менее 0,007 % статей характеризовались такими значениями β_- -балла!). Остаётся только заметить, что если бы эта и несколько других явно манипулятивных публикаций того же автора были идентифицированы ранее, то, вероятно, можно было бы избежать целой серии медийных атак против российских спортсменов, которые активно продолжают до сих пор, несмотря на полное отсутствие каких-либо материальных доказательств якобы совершенных «нарушений».

О манипулятивных публикациях, направленных на дискредитацию психиатрии как науки

«Если человеческое равенство надо навсегда сделать невозможным, если высшие, как мы их называем, хотят сохранить своё место навеки, тогда господствующим душевным состоянием должно стать управляемое безумие».

Дж. Оруэлл, антиутопия «1984».

Эта фраза известного британского писателя вызывает различные ассоциации. В России почему-то некоторые «креаклы» думают, что Дж. Оруэлл таким образом критиковал советский коммунизм и т. п. Никому из этих «креативных» меньшинств и в голову, почему-то не приходит, что Джордж Оруэлл (настоящее имя – Эрик Артур Блэр), вообще говоря, британский писатель и, как всякий психически адекватный писатель, он пишет о своей стране.

О каком «безумии» мог говорить Оруэлл? Безумная реклама из телевизора, имеющая многие признаки т. н. «нейролингвистического программирования»? Безумие телепередач, где гигантскими буквами пропагандируется «АЛЧНОСТЬ»? Бытовое безумие обыкновенного фашизма, охарактеризованного гениальным М. Роммом? Да, эти разновидности безумия очень опасны. Нас однако интересуют здесь чисто психиатрические аспекты проекции смысла цитируемой выше фразы английского писателя, причём в применении к анализу больших текстовых данных.

В самом деле, если некие облечённые властью персонажи хотят сеять то, что реально сертифицируется врачами-психиатрами как «безумие», им нужно решить несколько задач, в т. ч.:

- Легализация всех возможных извращений, которые считались таковыми даже в племенах дикарей-людоедов. Почти сделано, т. к. легализация таковых — официальная политика почти всех западных стран и наказывают уже не за извращения, а за противостояние им. Чтобы это понять, достаточно открыть любой британский или американский таблоид.

- Легализация химической зависимости (сначала т. н. «лёгких наркотиков»). Близко к завершению, см. ситуацию в западной Европе, на британских островах и в отдельных штатах США.

- Широчайшее распространение и пропаганда поведения высокого риска для жизни (в т. ч. суицида). В активной фазе — см. многочисленные медийные материалы о том, как ради «эпатажного селфи» сотни молодых людей во всём мире расстаются с жизнью.

- Устранение авторитета медицины вообще, полная дискредитация т. н. «доказательной медицины». Вполне успешно продвигается, см. предыдущие разделы статьи.

- Размытие установленных критериев психиатрических патологий, продвижение деструктивной идеологии «все психически здоровы, но с некоторыми отклонениями».

В последнем случае, т. е. в разрушении психиатрии как науки, активно принимают участие некоторые «психиатры». Разработанный нами метод топологического сентимент-анализа позволяет выявлять реальные случаи таких ультра-манипулятивных публикаций, не имеющих ничего общего с психиатрией, но, тем не менее, публикуемых в профессиональных журналах по психиатрии. В результате, алгоритм выявил поднимающуюся волну статей (несколько десятков статей в год), в которых, и в самом деле, психиатрия развенчивается как наука по принципу «генетика — лженаука», «кибернетика — буржуазная наука» и проч.

Рассмотрим, например, весьма яркий пример в виде публикации [36] и соответствующего медийного материала. Британские учёные (буквально — университет Ливерпуля) доказали, что, оказывается:

- Все психиатрические диагнозы в руководстве «Diagnostic and Statistical Manual (DSM)», что называется, «на одно лицо» и существенно не отличаются друг от друга. Заметим, что словечко «все» исходит от авторов текста [36].

- Все психиатрические диагнозы маскируют, оканчивается, какие-то неспецифические «травмы» и «враждебные события» («adverse events» — классическое «хорьковое слово», см. выше).

- Ни один диагноз ничего не говорит о какой-то абстрактной «индивидуальности» пациента и как лечить его/её.

- Рабочие критерии психиатров, повсеместно используемые в практике, научно бессмысленны («scientifically meaningless»).

При этом уровни манипулятивных и пропагандистских сентиментов, выявленные методами топологического анализа больших данных в данной публикации просто зашкаливают — в 10 раз выше, чем в среднестатистической публикации по психиатрии. Вот пример сразу нескольких манипулятивных конструкций всего лишь в одном предложении: руководство «Diagnostic and Statistical Manual (DSM)» оказывается представляет собой «диагностическое наклеивание ярлыков» («diagnostic labelling») и является «неумной категорной системой» (disingenuous categorical system) [36]. В общем, ситуация в полном соответствии с известным рассказом Э.А. По («Система доктора Смоля и профессора Перро») и с известным голливудским фильмом 2014 года, очевидно снятым по мотивам этого рассказа.

Заключение

В работе сформулирована методика анализа «больших данных» биомедицинских публикаций, основанная на топологической теории сентимент-анализа медицинских текстов (т. е. анализа эмоциональной модальности). Разработаны алгоритмы, позволяющие с 90 % точностью классифицировать тексты по 16 классам сентиментов (манипулятивные обороты речи, исследования без положительных результатов, пропаганда, подделка результатов, негативное личное отношение, агрессивность текста, негативный эмоциональный фон и др.). На основе алгоритмов предложена балльная шкала оценки сентимент-качества исследований (β -балл) и проведён анализ текстов абстрактов 19,9 млн публикаций из PUBMED. Показано, что публикации с низким сентимент-качеством (β -балл текста меньше 0, что соответствует преобладанию манипулятивных и негативных сентиментов) составляют 18,5 % (3,7 из 19,9 млн). Рубрикация всего массива публикаций по 27 840 рубрикам (MESH-система PUBMED) указала на повышение β -балла по годам (т. е. на положительную динамику качества исследований) для 27 090 исследованных рубрик. Наиболее интенсивная положительная динамика найдена для исследований по использованию методов прикладной математики и интеллектуального анализ данных в медицине.

Выделены 249 рубрик с резко отрицательной динамикой качества и с выраженным нарастанием манипулятивных сентиментов, подобных встречающимся в «жёлтой прессе» западных стран. Для публикаций из данных рубрик характерны отрицательные значения β -балла, т. е. преобладание в текстах публикаций манипулятивных языковых конструкций над содержательными. К данным рубрикам относятся публикации по экономическим и юридическим аспектам

биомедицины, по разработке т. н. «международных стандартов» для контроля за рождаемостью, тексты по экономике развивающихся стран, по психотропным препаратам, по легализации эвтаназии, стоимости стоматологических услуг в западных странах и др. Крайне интересно отметить, что абсолютным лидером по резкому ухудшению сентимент-качества публикаций являлись 12 405 публикаций из рубрик, имеющих отношение к доказательной медицине (падение β -балла более чем на 5 баллов/год за последние 6 лет – больше, чем во всех остальных 248 рубриках).

Таким образом, разработанный комплекс алгоритмов позволяет проводить эффективную оценку сентимент-качества биомедицинских исследований, отфильтровывая потенциально псевдонаучные публикации, публикуемых под масками якобы «доказательных исследований», «международных рекомендаций», «рекомендуемых стандартов», «мнения международного сообщества» и т. п. На наш взгляд, использование подобных систем алгоритмов может быть полезно в рамках реализации федерального закона № 489-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» по вопросам клинических рекомендаций» от 25.12.2018. В рамках этого закона постановлено, что действующие в настоящее время протоколы лечения будут применяться до их пересмотра не позднее 31 декабря 2021 г. Предполагается, что клинические рекомендации будут медицинскими профессиональными НКО по отдельным группам заболеваний [37].

Очевидно, что идентификация негативных, манипулятивных и прочих сентиментов в публикациях PUBMED затруднит для НКО, финансируемых из-за рубежа, проведение «накачки» врачей негативными эмоциями, педалирование соответствующих деструк-

тивных тем (см. левую колонку табл. 2) и выдачу штампованных «fake news» за якобы «доказательные» исследования. Выборки статей из PUBMED, не содержащие лингвистических признаков манипуляции, весьма важны для разработки клинических рекомендаций [38], учёта международного опыта по организации агентств по оценке технологий здравоохранения [39], повышения эффективности поиска медицинских работок с инновационным потенциалом [40], в т. ч. инновационных лекарственных препаратов и для реализации других важных направлений российского здравоохранения.

Также предполагаем, что сформулированный подход к анализу текстов поможет врачам выработать: 1) культуру грамотного игнорирования агрессивных меньшинств, которые навязывают свою «кочку зрения» через те или иные средства массовой информации; 2) навык сопротивления бездумному подчинению тоталитарной секте «доказательщиков»; 3) разумное противостояние интернет-среде, в которой работают платные профессионалы, поддерживающие постоянное раздражение в россиянах.

ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ

Конфликт интересов. Исследование не имело спонсорской поддержки. Авторы несут полную ответственность за предоставление окончательной версии рукописи в печать.

Участие авторов. Все авторы принимали участие в разработке концепции статьи и написании рукописи. Окончательная версия рукописи была одобрена всеми авторами.

СВЕДЕНИЯ ОБ АВТОРАХ

Максимов Валерий Алексеевич

ORCID ID: 0000-0003-4120-1071

д. м. н., профессор кафедры диетологии и нутрициологии ФГБОУ ДПО РМАНПО Минздрава России, Москва, Россия

Торшин Иван Юрьевич

ORCID ID 0000-0002-2659-7998

SPIN-код: 1375-1114

к. х. н., с. н. с. ФИЦ ИУ РАН, Москва, Россия; ЦХАБД НОЦ ЦЭ МГУ имени М.В. Ломоносова, Москва, Россия

Maximov Valery A.

ORCID ID: 0000-0003-4120-1071

D. Sci. in Medicine, Professor Department of dietetics and nutrition FSBEI FPE RMACPE MON Russia, Moscow, Russia

Torshin Ivan Yu.

ORCID ID 0000-0002-2659-7998

SPIN code: 1375-1114

PhD in Chemical, Senior researcher FRC CSC RAS, Moscow, Russia; Big Data Storage and Analysis Center Lomonosov MSU, Moscow, Russia

Громова Ольга Алексеевна
Автор, ответственный за переписку
 e-mail: unesco.gromova@gmail.com
 ORCID ID: 0000-0002-7663-710X
 SPIN-код: 6317-9833
 д. м. н, профессор, в. н. с., научный руководитель
 Института фармакоинформатики при ФИЦ
 ИУ РАН, Москва, Россия; ЦХАБД НОЦ ЦЭ
 МГУ имени М.В. Ломоносова, Москва, Россия

Gromova Olga A.
Corresponding author
 e-mail: unesco.gromova@gmail.com
 ORCID ID: 0000-0002-7663-710X
 SPIN code: 6317-9833
 D. Sci. in Medicine, Professor FRC CSC RAS,
 Moscow, Russia; Big Data Storage and Analysis
 Center Lomonosov MSU, Moscow, Russia

Галустян Анна Николаевна
 ORCID ID: 0000-0001-9679-632X
 к. м. н., доцент, заведующая кафедрой
 фармакологии с курсом клинической
 фармакологии и фармакоэкономики ФГБОУ
 ВО СПбГПМУ Минздрава России, Санкт-
 Петербург, Россия

Galustyan Anna N.
 ORCID ID: 0000-0001-9679-632X
 PhD in Medicine, Assistant of professor, Head of
 the Department of Pharmacology with the course
 of Clinical Pharmacology and Pharmacoeconom-
 ics FSBI HPE «SPbSPMU» MOH Russia, Saint-
 Petersburg, Russia

Гоголева Ирина Викторовна
 ORCID ID: 0000-0002-2178-8285
 к. м. н, доцент кафедры фармакологии ФГБОУ
 ВО ИвГМА Минздрава России, Иваново,
 Россия

Gogoleva Irina V.
 ORCID ID: 0000-0002-2178-8285
 PhD in Medicine, Associate Professor of the
 Department of Pharmacology FSBEI HE IvSMA
 MOH Russia, Ivanovo, Russia

Гришина Татьяна Романовна
 ORCID ID: 0000-0002-1665-1188
 SPIN-код: 1241-0701
 д. м. н., профессор, заведующая кафедрой
 фармакологии ФГБОУ ВО ИвГМА Минздрава
 России, Иваново, Россия

Grishina Tatiana R.
 ORCID ID: 0000-0002-1665-1188
 SPIN code: 1241-0701
 D. Sci. in Medicine, Professor, Head of Depart-
 ment of pharmacology FSBEI HE IvSMA MOH
 Russia, Ivanovo, Russia

Громов Андрей Николаевич
 ORCID ID: 0000-0001-7507-191X
 SPIN-код: 8034-7910
 инженер-исследователь ФИЦ ИУ РАН,
 Москва, Россия; ЦХАБД НОЦ ЦЭ МГУ имени
 М.В. Ломоносова, Москва, Россия

Gromov Alexander N.
 ORCID ID: 0000-0001-7507-191X
 SPIN code: 8034-7910
 FRC CSC RAS, Moscow, Russia;
 Big Data Storage and Analysis Center Lomonosov
 MSU, Moscow, Russia

Калачева Алла Геннадьевна
 ORCID ID: 0000-0001-6144-5781
 к. м. н., доцент кафедры фармакологии ФГБОУ
 ВО ИвГМА Минздрава России, Иваново,
 Россия

Kalacheva Alla G.
 ORCID ID: 0000-0001-6144-5781
 PhD in Medicine, Associate Professor of the
 Department of Pharmacology FSBEI HE IvSMA
 MOH Russia, Ivanovo, Russia

Кобалава Жанна Давидовна
 ORCID ID: 0000-0003-1126-4282
 SPIN-код: 9828-5409
 д. м. н., профессор, член корреспондент РАН,
 заведующая кафедрой внутренних болезней
 с курсом кардиологии и функциональной
 диагностики имени академика Моисеева В.С.
 РУДН, Москва, Россия

Kobalava Zhanna D.
 ORCID ID: 0000-0003-1126-4282
 SPIN code: 9828-5409
 D. Sci. in Medicine, Professor, corresponding
 member of the RAS, head of the Department of
 internal diseases with the course of cardiology and
 functional diagnostics named after academician
 Moiseev V.S. PFUR, Moscow, Russia

Коденцова Вера Митрофановна

д. б. н., профессор, г. н. с. лаборатории витаминов и микроэлементов ФГБУН «ФИЦ питания и биотехнологии», Москва, Россия

Kodentsova Vera M.

D. Sci. in Biological, Professor, chief Researcher of the Laboratory of Vitamins and Trace Elements Federal Research Center of Nutrition and Biotechnology, Moscow, Russia

Лиманова Ольга Адольфовна

ORCID ID: 0000-0002-2342-3036
к. м. н., доцент кафедры фармакологии и клинической фармакологии ФГБОУ ВО ИвГМА Минздрава России, Иваново, Россия

Limanova Olga A.

ORCID ID: 0000-0002-2342-3036
PhD in Medicine, Associate Professor of the Department of Pharmacology and Clinical Pharmacology FSBEI HE IvSMA MOH Russia, Ivanovo, Russia

Малявская Светлана Ивановна

ORCID ID: 0000-0003-2521-0824
SPIN-код: 6257-4400
д. м. н., профессор, проректор по научной работе ФГБОУ ВО СГМУ (г. Архангельск) Минздрава России, Архангельск, Россия

Malyvskaya Svetlana I.

ORCID ID: 0000-0003-2521-0824
SPIN code: 6257-4400
D. Sci. in Medicine, Professor, Head of Sciences FSBEI HE NSMU (Arkhangelsk) MOH Russia, Arkhangelsk, Russia

Рудаков Константин Владимирович

д. ф.-м. н., академик РАН, заместитель директора ФИЦ ИУ РАН, Москва, Россия; заведующий кафедрой интеллектуальные системы МФТИ, Москва, Россия

Rudakov Konstantin V.

D. Sci. in Physico-Mathematical, Academician of the Russian Academy of Sciences, Deputy Director FRC CSC RAS, Moscow, Russia; Head of the Department of Intelligent Systems MIPT, Moscow, Russia

Сардарян Иван Суменович

к. м. н., доцент кафедры фармакологии с курсом клинической фармакологии и фармакоэкономики ФГБОУ ВО СПбГПМУ Минздрава России, Санкт-Петербург, Россия

Sardaryan Ivan S.

PhD in Medicine, Associate Professor of the Department of Pharmacology with a course in Clinical Pharmacology and Pharmacoeconomics FSBI HPE «SPbSPMU» MOH Russia, Saint-Petersburg, Russia

Сорокин Александр Игоревич

к. ф.-м. н., с. н. с. ЦХАБД НОЦ ЦЭ МГУ имени М.В. Ломоносова, Москва, Россия

Sorokin Alexander I.

PhD in Physico-Mathematical, Senior Research Officer Big Data Storage and Analysis Center Lomonosov MSU, Moscow, Russia

Стаховская Людмила Витальевна

ORCID ID: 0000-0001-6325-9237
д. м. н., доцент, профессор кафедры неврологии, нейрохирургии и медицинской генетики ЛФ, директор ФГБУ «ФЦЦПИ» Минздрава России, Москва, Россия

Stakhovskaya Ludmila V.

ORCID ID: 0000-0001-6325-9237
D. Sci. in Medicine, Associate Professor, Professor of the Department of Neurology, Neurosurgery and Medical Genetics, LF Director FSBS «FCCPS» MOH Russia, Moscow, Russia

Тапильская Наталья Игоревна

ORCID ID: 0000-0001-5309-0087
SPIN-код: 3605-0413
д. м. н., профессор, в. н. с. отделения вспомогательных репродуктивных технологий ФГБНУ «НИИ АГиР им. Д.О. Отта», Санкт-Петербург, Россия

Tapilskaya Natalia I.

ORCID ID: 0000-0001-5309-0087
SPIN code: 3605-0413
D. Sci. in Medicine, Professor, leading researcher of the Department of Assisted Reproductive Technologies FSBSI «The Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott», Saint-Petersburg, Russia

Тетруашвили Нана Картлосовна

д. м. н., заведующая 2-м отделением акушерским патологии беременности ФГБУ НЦАГиП им. академика В.И. Кулакова Минздрава России, Москва, Россия

Федотова Любовь Эдуардовна

ORCID ID: 0000-0002-0778-1562

к. м. н., доцент кафедры фармакологии и клинической фармакологии ФГБОУ ВО ИвГМА Минздрава России, Иваново, Россия

Хаджидис Александр Кириакович

к. м. н., доцент кафедры аллергологии с курсом к клинической фармакологии ФГБОУ ВО СПбГПМУ Минздрава России, Санкт-Петербург, Россия

Tetruashvili Nana K.

D. Sci. in Medicine, Head of the 2nd department of obstetrics pathology of pregnancy FSBI «National Medical Research Center for Obstetrics, Gynecology and Perinatology named after Academician V.I. Kulakov» MOH Russia, Moscow, Russia

Fedotova Lyubov E.

ORCID ID: 0000-0002-0778-1562

PhD in Medicine, Associate Professor of the Department of Pharmacology and Clinical Pharmacology FSBEI HE IvSMA MOH Russia, Ivanovo, Russia

Hadzhidis Alexander K.

PhD in Medicine, Assistant of professor of the Department of Allergology with a course in Clinical Pharmacology FSBI HPE «SPbSPMU» MOH Russia, Saint-Petersburg, Russia

Литература / References

- Sanders S, Del Mar C. Clever searching for evidence. *BMJ*. 2005;330(7501):1162-3. DOI: 10.1136/bmj.330.7501.1162.
- Ramos K, Linscheid R, Schafer S. Real-time information-seeking behavior of residency physicians. *Fam Med*. 2003;35(4):257–260.
- Torshin IYu. Bioinformatics in the post-genomic era: physiology and medicine. Nova Biomedical Books, NY, USA (2007). ISBN 1-60021-048-1.
- Громова О.А., Торшин И.Ю. Витамин D. Смена парадигмы, 2-е издание, переработанное и дополненное / Под ред. Е. И. Гусев, И. Н. Захарова). Москва : ГЭОТАР-Медиа; 2021. — 736 с. [Gromova OA, Torshin IYu. Vitamin D. Smena paradigm, 2-e izdanie, pererabotannoe i dopolnennoe / Ed by EI Gusev, IN Zakharova. Moscow: GEOTAR-Media; 2021. (In Russ).]. ISBN 978-5-9704-5787-0.
- Stewart Chaplin. “The Stained Glass Political Platform”. The Century Magazine, USA, 1900.
- Summers E. Weasel Words: 200 Words You Shouldn't Trust: 200 Words You Can't Trust. Chambers (Ed.), Slang & Idiom Dictionaries, 2009. ISBN-13: 978-0550104762.
- Watson D. Watson's Dictionary of Weasel Words, Contemporary Cliches, Cant and Management Jargon. Knopf; 1st Ed. 2004. ISBN-10: 1740513215, ISBN-13: 978-1740513210.
- Torshin IYu, Rudakov KV. Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 1: factorization approach. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2017;27(1):16–28. DOI: 10.1134/S1054661817010151.
- Torshin IYu, Rudakov KV. Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 2: metric approach within the framework of the theory of classification of feature values. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2017;27(2):184–199. DOI: 10.1134/S1054661817020110.
- Torshin IYu. Optimal dictionaries of the final information on the basis of the solvability criterion and their applications in bioinformatics. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2013;23(2):319–327. DOI: 10.1134/S1054661813020156.
- Torshin IYu, Rudakov KV. On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification. *Pattern Recognition and Image Analysis*. 2015;25(4):577–587. DOI: 10.1134/S1054661815040252.
- Torshin IYu, Rudakov KV. On metric spaces arising during formalization of problems of recognition and classification. Part 1: properties of compactness. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2016;26(2):274–284. DOI: 10.1134/S1054661816020255.
- Torshin IYu, Rudakov KV. On metric spaces arising during formalization of problems of recognition and classification. Part 2: density properties. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2016;26(3):483–496. DOI: 10.1134/S1054661816030202.
- Torshin IYu, Rudakov KV. On the application of the combinatorial theory of solvability to the analysis of chemographs. Part 1: fundamentals of modern chemical bonding theory and the concept of the chemograph. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2014;24(1):11–23. DOI: 10.1134/S1054661814010209.
- Torshin IYu, Rudakov KV. On the application of the combinatorial theory of solvability to the analysis of chemographs. Part 2: local completeness of invariants of chemographs in view of the combinatorial theory of solvability. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2014;24(2):196–208. DOI: 10.1134/S1054661814020151.
- Torshin IYu, Rudakov KV. On the Procedures of Generation of Numerical Features Over Partitions of Sets of Objects in the Problem of Predicting Numerical Target Variables. *Pattern Recognition and Image Analysis*. 2019;29(4):654–667. DOI: 10.1134/S1054661819040175.
- Чернышев В. М. Меч Обоюдоострый. Конспект по Сектоведению. — М.: Нобель Пресс; 2011; — 138 с. [Chernyshev VM. Mech Oboyudoostryj. Konspekt po Sektovedeniyu. Moscow: Nobel Press; 2011. (In Russ).]. ISBN 9785424134425.
- Дворкин А.Л. Сектоведение: Тоталитарные секты. Опыт систематического исследования. 3-е изд., перераб. и доп. — Нижний Новгород: Христианская библиотека; 2014; 816 с. [Dworkin A.L. Sektovedenie: Totalitarnye sekty. Opyt sistematicheskogo issledovaniya. 3-e izd., pererab. i dop. Nizhny Novgorod: Hristianskaya biblioteka; 2014. (In Russ).]. ISBN 978-5-905472-03-9, 978-5-905472-21-3.
- Okter A. Mastermind: The Truth of the British Deep State. Arashtirma Publishing. 2017; 698 pp.
- Поппер К.Р. Предположения и опровержения: Рост научного знания: Пер. с англ. / К.Р. Поппер. — М.: ООО «Издательство АСТ»: ЗАО НПЛ «Ермак»; 2004. — 638 с. [Popper K. R. Conjectures and Refutations. The Growth of Scientific Knowledge. Moscow: ООО «Izdatel'stvo AST»: ЗАО NPL «Ermak»; 2004. (In Russ).]. ISBN 5-17-012641-7; ISBN 5-9577-0652-3.
- Арнольд В., Ильяшенко Ю., Аносов Д. и др. Динамические системы — 1. Итоги науки и техн. Сер. Современ. пробл. мат. Фундам. направления. — М.: ВИНТИ. — 260 с. [Arnold V, Ilyashenko Yu, Anosov D, et al. Dinamicheskie sistemy — 1. Itogi nauki i tekhn. Ser. Sovrem. probl. mat. Fundam. napravleniya. Moscow: VINITI. 260 p. (In Russ).].
- Левенштейн В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Доклады Академий Наук СССР. 1965;163(4):845–848. [Levenshtein VI. Dvoichnye kody s ispravleniem vypadenij, vstavok i zameshchenij simvolov. Doklady Akademij Nauk USSR. 1965;163(4):845–848. (In Russ).].

23. Ioannidis JPA. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. *J Clin Epidemiol*. 2017 Apr; 84:11-13. DOI: 10.1016/j.jclinepi.2017.02.001.
24. Ioannidis JP. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. 2016 May; 73:82–86. DOI: 10.1016/j.jclinepi.2016.02.012.
25. Möller MH, Ioannidis JPA, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Med*. 2018 Apr; 44(4):518-520. DOI: 10.1007/s00134-017-5039-y.
26. Cochrane is a registered trademark in Australia, Canada, the European Community and the USA. 2017-09-19. [Internet] URL: trademarks.justia.com/791/85/cochrane-79185910.html. Дата обращения: 12.12.2019.
27. Торшин И.Ю., Громова О.А., Кобалава Ж.Д. «Об ошибках метаанализов сердечно-сосудистых эффектов омега-3 ПНЖК. Часть 1. Фармакологические и клинические аспекты доказательности в эпоху постгеномных исследований, искусственного интеллекта и анализа больших данных» // *Эффективная фармакотерапия*. 2019;15(9):26–34. [Torshin IYu, Gromova OA, Kobalava ZhD. About errors in meta-analyses of cardiovascular effects of omega-3 PUFA. Part 1. Pharmacological and clinical aspects of validity in the era of post-genomic research, artificial intelligence and big data analysis. *Effektivnaya farmakoterapiya*. 2019;15(9):26-34. (In Russ).] DOI: 10.33978/2307-3586-2019-15-9-26-34.
28. Торшин И.Ю., Громова О.А., Кобалава Ж.Д. «Об ошибках метаанализов сердечно-сосудистых эффектов омега-3 ПНЖК. Часть 2. Интеллектуальный анализ данных и метаанализ клинически однородных исследований» // *Эффективная фармакотерапия*. 2019;15(9):36–43. [Torshin IYu, Gromova OA, Kobalava ZhD. About errors in meta-analyses of cardiovascular effects of omega-3 PUFA. Part 2. Intellectual analysis and meta-analysis of clinically homogeneous studies. *Effektivnaya farmakoterapiya*. 2019;15(9):36–43. (In Russ).] DOI: 10.33978/2307-3586-2019-15-9-36-43.
29. Лиманова О.А., Торшин И.Ю., Калачева А.Г. и др. Обеспеченность микронутриентами и женское здоровье: интеллектуальный анализ клинко-эпидемиологических данных // *Вопросы гинекологии, акушерства и перинатологии*. 2014;13(2):5–15. [Limanova OA, Torshin IYu, Sardaryan IS et al. Micronutrient provision and women's health: intellectual analysis of clinicoepidemiological data. *Gynecology, obstetrics and perinatology*. 2014;13(2):5–15. (In Russ).]
30. Громова О.А., Торшин И.Ю., Громов А.Н. и др. Интеллектуальный анализ данных по течению и исходу беременности: роли различных витаминно-минеральных комплексов // *Медицинский алфавит*. 2018;1(6):10–23. [Gromova OA, Torshin IYu, Gromov AN et al. Data mining in course and outcome of pregnancy: role of vitamin and mineral complexes. *Medical alphabet*. 2018;1(6):10–23. (In Russ).]
31. Громова О.А., Торшин И.Ю. Витамин D. Смена парадигмы / Под ред. Е. И. Гусева, И. Н. Захаровой. — Москва: ГЭОТАР-Медиа; 2017. — 568 с. : ил. ; 25 см. — Библиогр. в конце гл. — 3000 (1-й з-д 500) экз. [Gromova OA, Torshin IYu. Vitamin D. Smena paradigmy / Ed by EI Gusev, IN Zakharova. Moscow: GEOTAR-Media; 2017. (In Russ).] ISBN 978-5-9704-4058-2.
32. Hannemann A, Wallaschofski H, Nauck M et al. Vitamin D and health care costs: Results from two independent population-based cohort studies. *Clin Nutr*. 2018 Dec;37(6 Pt A):21490–2155. DOI: 10.1016/j.clnu.2017.10.014.
33. Механик А.Г. Искусственный интеллект на страже здоровья. Беседа вторая с О.А. Громовой и И.Ю. Торшиным // *Стимул: Журнал об инновациях в России*. 30 Октября 2019. [Электронный ресурс]. [Mechanic AG. Iskusstvennyj intellekt na strazhe zdorov'ya. Beseda vtoraya s OA Gromova i IYu Torshin. *Stimul: Zhurnal ob innovaciyah v Rossii*. 2019;15(9):36–43. [Internet]. [cited October 30, 2019]. (In Russ).] URL: https://stimul.online/articles/science-and-technology/iskusstvennyy-intellekt-na-strazhe-zdorovya-beseda-vtoraya/. Дата обращения: 23.06.2020.
34. Национальная программа «Недостаточность витамина D у детей и подростков Российской Федерации: Современные подходы к коррекции». Москва, 2018. Союз педиатров России. ФГАУ «Национальный медицинский исследовательский центр здоровья детей» Минздрава России. ФГБОУ ДПО «Российская медицинская академия непрерывного профессионального образования» Минздрава России. ФГБУН «Федеральный исследовательский центр питания, биотехнологии и безопасности пищи». [Nacional'naya programma «Nedostatochnost' vitamina D u detej i podrostkov Rossijskoj Federacii: Sovremennye podhody k korekcii». Moscow, 2018. Soyuz pediatrov Rossii. FGAU «Nacional'nyj medicinskij issledovatel'skij centr zdorov'ya detej» Minzdrava Rossii. FGBOU DPO «Rossijskaya medicinskaya akademiya nepreryvnogo professional'nogo obrazovaniya» Minzdrava Rossii. FGBUN «Federal'nyj issledovatel'skij centr pitaniya, biotekhnologii i bezopasnosti pishchi». (In Russ).]
35. Krotov G, Nikitina M, Rodchenkov G. Possible cause of lack of positive samples on homologous blood transfusion. *Drug Test Anal*. 2014 Nov-Dec;6(11-12):1160–1162. DOI: 10.1002/dta.1736.
36. Allsopp K et al. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Res*. 2019 Sep;279:15-22. DOI: 10.1016/j.psychres.2019.07.005.
37. Федин А.И. Президент РФ подписал закон о клинических рекомендациях. Невроньюс: новости Неврологии. 2019;1(51):1. [Fedin AI. Prezident RF podpisal zakon o klinicheskikh rekomendaciyah. *Nevron'yus: novosti Nevrologii*. 2019;1(51):1. (In Russ).]
38. Журавлева Н.И., Шубина Л.С., Сухоруких О.А. Обзор методик оценки достоверности научных доказательств и убедительности рекомендаций, применяемых при разработке клинических рекомендаций в Российской Федерации // *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2019;12(1):34–41. [Zhuravleva NI, Shubina LC, Sukhorukikh OA. The use of the level of evidence and grade of recommendations scales in developing clinical guidelines in the Russian Federation. *Modern Pharmacoeconomics and Pharmacoepidemiology [FARMAKOEKONOMIKA. Sovremennaya farmakoeconomika i farmakoepidemiologiya]*. 2019;12(1):34–41. (In Russ).] DOI: 10.17749/2070-4909.2019.12.1.34-41.
39. Омеляновский В.В., Федяева В.К., Мусина Н.З. Концепция многокритериального анализа принятия решений в текущей системе оценки технологий в здравоохранении России // *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2018;11(3):003–007. [Omelyanovsky VV, Fedyeva VK, Musina NZ. The concept of multicriteria analysis of decision-making in the current system of health technology assessment in Russia. *Farmakoeconomika. Modern pharmacoeconomics and pharmacoepidemiology. [Farmakoeconomika. Sovremennaya farmakoeconomika i farmakoepidemiologiya]*. 2018;11(3):003–007. (In Russ).] DOI: 10.17749/2070-4909.2018.11.3-003-007.
40. Хрусталева М.Б., Максимова А.А. Эффективный поиск научных разработок с инновационным потенциалом в медицине // *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2019;12(1):27–33. [Khrustalev MB, Maksimova AA. Effective search for potentially innovative scientific results in medicine. *FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology [Farmakoeconomika. Sovremennaya farmakoeconomika i farmakoepidemiologiya]*. 2019;12(1):27–33. DOI: 10.17749/2070-4909.2019.12.1.27–33.