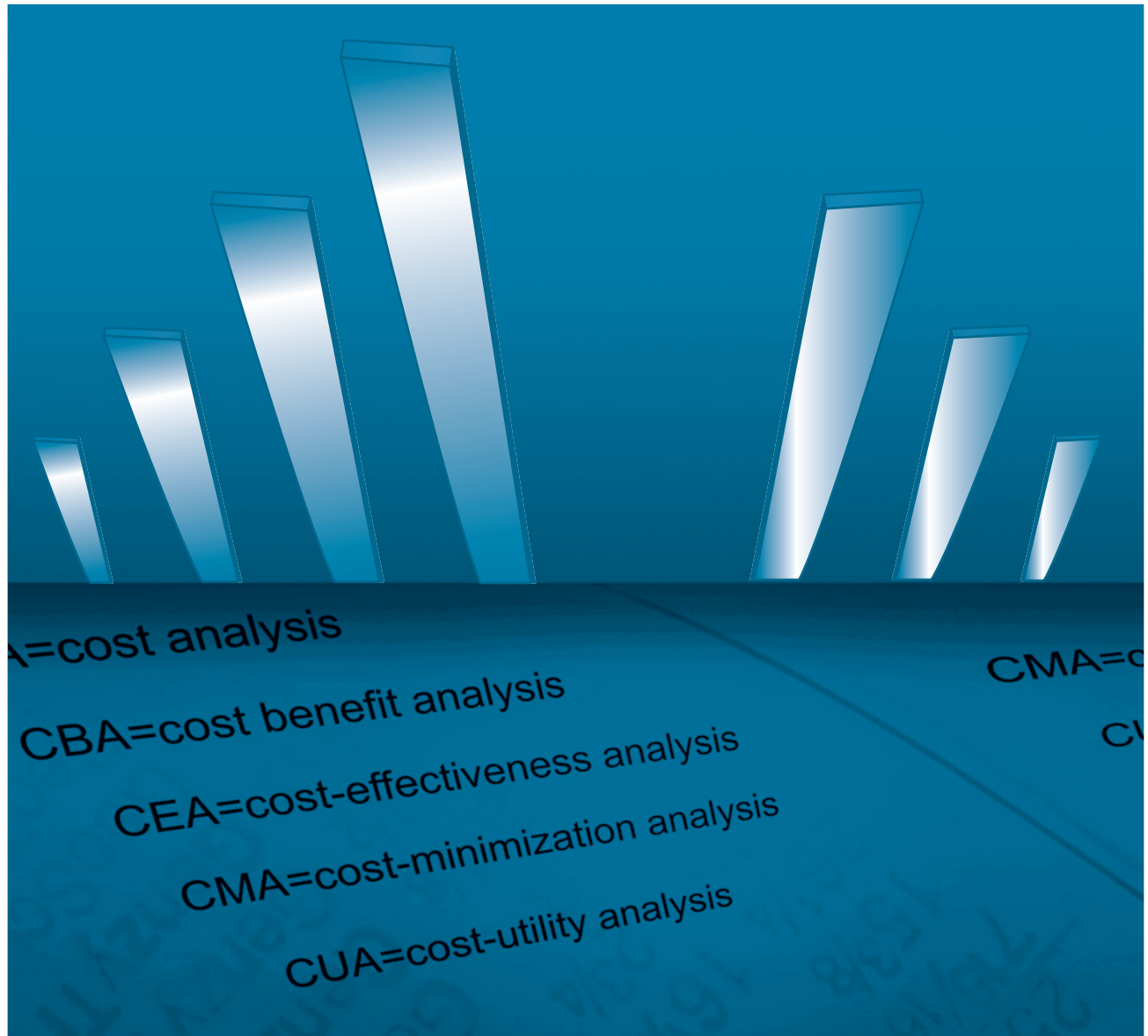


# Фармакоэкономика

Современная Фармакоэкономика и Фармакоэпидемиология



**FARMAKOEKONOMIKA**

Modern Pharmacoeconomic and Pharmacoepidemiology

2020 Vol. 13 No2

[www.pharmacoeconomics.ru](http://www.pharmacoeconomics.ru)

- Доступность генной терапии *in vivo*. Проблемы и решения
- Концепция ценностно-ориентированного здравоохранения
- Стоит ли переосмыслить полученный полвека назад положительный опыт применения хондроитинсульфатов при атеросклерозе?

№2

Том 13

2020



<https://doi.org/10.17749/2070-4909/farmakoekonomika.2020.021>

ISSN 2070-4909 (print)

ISSN 2070-4933 (online)

# Анализ 19,9 млн публикаций базы данных PubMed/MEDLINE методами искусственного интеллекта: подходы к обобщению накопленных данных и феномен “fake news”

Торшин И. Ю.<sup>1,2</sup>, Громова О. А.<sup>1,2</sup>, Стаховская Л. В.<sup>3</sup>, Ванчакова Н. П.<sup>4</sup>, Галустян А. Н.<sup>5</sup>, Кобалава Ж. Д.<sup>6</sup>, Гришина Т. Р.<sup>7</sup>, Громов А. Н.<sup>1</sup>, Иловайская И. А.<sup>8</sup>, Коденцова В. М.<sup>9</sup>, Калачева А. Г.<sup>7</sup>, Лиманова О. А.<sup>7</sup>, Максимов В. А.<sup>10</sup>, Малявская С. И.<sup>11</sup>, Мозговая Е. В.<sup>12</sup>, Тапильская Н. И.<sup>5,12</sup>, Рудаков К. В.<sup>1</sup>, Семенов В. А.<sup>13</sup>

<sup>1</sup> *Федеральный исследовательский центр «Информатика и управление» РАН (ФИЦ ИУ РАН) (ул. Вавилова, д. 44, корп. 2, Москва 119333, Россия)*

<sup>2</sup> *Центр хранения и анализа больших данных, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова» (Ленинские горы, д. 1, Москва 119991, Россия)*

<sup>3</sup> *Федеральное государственное бюджетное учреждение «Федеральный центр цереброваскулярной патологии и инсульта» Министерства здравоохранения Российской Федерации (ул. Островитянова, д. 1, стр. 10, Москва 117342, Россия)*

<sup>4</sup> *Центр психосоматической медицины при клинической больнице № 122 им. Л. Г. Соколова (пр. Культуры, д. 4, Санкт-Петербург 194291, Россия)*

<sup>5</sup> *Государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный педиатрический медицинский университет» Министерства здравоохранения Российской Федерации (ул. Литовская, д. 2, Санкт-Петербург 194100, Россия)*

<sup>6</sup> *Федеральное государственное автономное образовательное учреждение высшего образования «Российский университет дружбы народов» (Медицинский институт) (ул. Миклухо-Маклая, д. 10/3, Москва 117198, Россия)*

<sup>7</sup> *Федеральное государственное бюджетное образовательное учреждение высшего образования «Ивановская государственная медицинская академия» Министерства здравоохранения Российской Федерации (Шереметевский проспект, д. 8, г. Иваново 153012, Россия)*

<sup>8</sup> *Государственное бюджетное учреждение здравоохранения Московской области «Московский областной научно-исследовательский клинический институт им. М. Ф. Владимирского» (ул. Щепкина, д. 61/2, Москва 129110, Россия)*

<sup>9</sup> *Федеральное государственное бюджетное учреждение науки «ФИЦ питания и биотехнологии» (Устьинский проезд, д. 2/14, Москва 109240, Россия)*

<sup>10</sup> *Федеральное государственное бюджетное образовательное учреждение дополнительного профессионального образования «Российская медицинская академия непрерывного профессионального образования» Министерства здравоохранения Российской Федерации (Баррикадная ул., д. 2/1, стр. 1, Москва 125993, Россия)*

<sup>11</sup> *Федеральное государственное бюджетное образовательное учреждение высшего образования «Северный государственный медицинский университет» Министерства здравоохранения Российской Федерации (Троицкий проспект, д. 51, г. Архангельск 163000, Россия)*

<sup>12</sup> Федеральное государственное бюджетное научное учреждение  
«Научно-исследовательский институт акушерства, гинекологии и репродуктологии  
имени Д. О. Отта» (Менделеевская линия, д. 3, Санкт-Петербург 199034, Россия)

<sup>13</sup> Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Кемеровский государственный медицинский университет»  
(ул. Ворошилова, д. 22а, г. Кемерово 650056, Россия)

Для контактов: Громова Ольга Алексеевна, e-mail: unesco.gromova@gmail.com

## РЕЗЮМЕ

**Введение.** Англоязычные базы данных PubMed/MEDLINE и Embase являются ценными информационными ресурсами для нахождения оригинальных публикаций по фундаментальной и клинической медицине. В настоящее время не существует систем искусственного интеллекта, позволяющих оценивать качество этих публикаций.

**Цель.** Разработка и апробация системы для проведения сентимент-анализа (то есть анализа эмоциональной модальности) публикаций по биомедицине.

**Материалы и методы.** Сформулирована методика анализа «больших данных» биомедицинских публикаций, основанная на топологической теории сентимент-анализа медицинских текстов. Разработаны алгоритмы, позволяющие с 90%-й 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).

**Результаты.** Проведен анализ текстов абстрактов 19,9 млн публикаций, зарегистрированных в PubMed/MEDLINE за последние 50 лет (1970–2019). Показано, что публикации с низким сентимент-качеством (значение β-балла текста меньше нуля, что соответствует преобладанию манипулятивных и негативных сентиментов в тексте) составляют всего 18,5% (3,68 из 19,9 млн). Наибольшими значениями β-балла характеризовались публикации по спортивной медицине, системной биологии, нутрициологии, по использованию методов прикладной математики и интеллектуального анализа данных в медицине. Рубрикация всего массива публикаций по 27840 рубрикам (MESH-система PubMed/MEDLINE) указала на повышение β-балла по годам (то есть на положительную динамику сентимент-качества текстов публикаций) для 27090 исследованных рубрик. Наиболее интенсивная положительная динамика найдена для исследований по генетике, физиологии, фармакологии и геронтологии. Выделены 249 рубрик с резко отрицательной динамикой сентимент-качества и с выраженным нарастанием манипулятивных сентиментов, характерных для «желтой» англоязычной прессы. Приведены отдельные оценки международных экспертов, которые подтверждают выявленные закономерности.

**Заключение.** Разработанная система искусственного интеллекта позволяет проводить эффективную оценку сентимент-качества биомедицинских исследований, отфильтровывая потенциально неадекватные публикации, публикуемые под маской «доказательных».

## КЛЮЧЕВЫЕ СЛОВА

Анализ больших данных, фармакоинформатика, искусственный интеллект, доказательная медицина, машинное обучение, алгоритмы оценки качества публикаций, тематическое моделирование.

Статья поступила: 18.12.2019 г.; в доработанном виде: 19.03.2020 г.; принята к печати: 01.06.2020 г.

## Конфликт интересов

Авторы заявляют об отсутствии необходимости раскрытия финансовой поддержки или конфликта интересов в отношении данной публикации. Все авторы сделали эквивалентный вклад в подготовку публикации.

## Финансирование

Работа выполнена при поддержке грантов РФФИ 19-07-00356 17-07-00935 17-07-01419 18-07-01022 18-07-00944 18-07-00929 16-07-01133.

## Для цитирования

Торшин И. Ю., Громова О. А., Стаховская Л. В., Ванчакова Н. П., Галустян А. Н., Кобалава Ж. Д., Гришина Т. Р., Громов А. Н., Иловайская И. А., Коденцова В. М., Калачева А. Г., Лиманова О. А., Максимов В. А., Малявская С. И., Мозговая Е. В., Тапильская Н. И., Рудаков К. В., Семенов В. А. Анализ 19,9 млн публикаций базы данных PubMed/MEDLINE методами искусственного интеллекта: подходы к обобщению накопленных данных и феномен “fake news”. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология.* 2020; 13 (2): 146–163. <https://doi.org/10.17749/2070-4909/farmakoeconomika.2020.021>.

## Analysis of 19.9 million publications from the PubMed/MEDLINE database using artificial intelligence methods: approaches to the generalizations of accumulated data and the phenomenon of “fake news”

Torshin I. Yu.<sup>1,2</sup>, Gromova O. A.<sup>1,2</sup>, Stakhovskaya L. V.<sup>3</sup>, Vanchakova N. P.<sup>4</sup>, Galustyan A. N.<sup>5</sup>, Kobalava Zh. D.<sup>6</sup>, Grishina T. R.<sup>7</sup>, Gromov A. N.<sup>1</sup>, Ilovaiskaya I. A.<sup>8</sup>, Kodentsova V. M.<sup>9</sup>, Kalacheva A. G.<sup>7</sup>, Limanova O. A.<sup>7</sup>, Maksimov V. A.<sup>10</sup>, Malyavskaya S. I.<sup>11</sup>, Mozgovaya E. V.<sup>12</sup>, Tapil'skaya N. I.<sup>5,12</sup>, Rudakov K. V.<sup>1</sup>, Semenov V. A.<sup>13</sup>

<sup>1</sup> Federal Research Center “Informatics and Management of the Russian Academy of Sciences (44-2 Vavilova Str., Moscow 119333, Russia)

<sup>2</sup> Moscow State University (1 Leninskie gory, Moscow 119991, Russia)

<sup>3</sup> Federal Center for Cerebrovascular Pathology and Stroke (1-10 Ostrovityanova Str., Moscow 117997, Russia)

<sup>4</sup> Center for Psychosomatic Medicine at the Clinical Hospital No. 122 named after L. G. Sokolov (4 pr. Kultury, St. Petersburg, 194291, Russia)

<sup>5</sup> Saint-Petersburg State Pediatric Medical University (2 Litovskaya Str., St. Petersburg 194100, Russia)

<sup>6</sup> Peoples' Friendship University of Russia (10/3 Miklukho-Maklaya Str., Moscow 117198, Russia)

<sup>7</sup> Ivanovo State Medical Academy (8 Sheremetevsky prospekt, Ivanovo 153012, Russia)

<sup>8</sup> The State Budgetary Healthcare Institution of Moscow Area Moscovs regional research clinical institute n.a. M.F. Vladimirov (61/2 Shchepkina Str., Moscow 129110, Russia)

<sup>9</sup> Federal Research Center for Nutrition and Biotechnology (2/14 Ustinsky proezd, Moscow 109240, Russia)

<sup>10</sup> Federal State Budgetary Educational Institution of Continuing Professional Education "Russian Medical Academy of Continuing Professional Education" of the Ministry of Health of the Russian Federation (2/1 Building 1 Barrikadnaya Str., Moscow 125993, Russia)

<sup>11</sup> Northern State Medical University (51 Troitskiy Ave., Arkhangelsk 163000, Russia)

<sup>12</sup> The Research Institute of Obstetrics, Gynecology and Reproductology named after D. O. Ott (3 Mendeleev Line, St. Petersburg 199034, Russia)

<sup>13</sup> Kemerovo State Medical University (22a Voroshilova Str., Kemerovo 650056, Russia)

**Corresponding author:** Olga A. Gromova, e-mail: [unesco.gromova@gmail.com](mailto:unesco.gromova@gmail.com)

## SUMMARY

**Introduction.** The English-language databases PubMed/MEDLINE and Embase are valuable information resources for finding original publications in basic and clinical medicine. Currently, there are no artificial intelligence systems to evaluate the quality of these publications.

**Aim.** Development and testing of a system for sentiment analysis (i.e. analysis of emotional modality) of biomedical publications.

**Materials and methods.** The technique of analysis of the "Big data" of biomedical publications was formulated on the basis of the topological theory of sentiment analysis. Algorithms have been developed that allow for the classification of texts from 16 sentiment classes with 90% accuracy (manipulative speech, research without positive results, propaganda, falsification of results, negative personal attitude, aggressive text, negative emotional background, etc.). Based on the algorithms, a scale for assessing the sentiment quality of research ( $\beta$ -score) is proposed.

**Results.** Abstracts of 19.9 million publications registered in PubMed/MEDLINE over the past 50 years (1970–2019) were analyzed. It was shown that publications with low sentiment quality (the value of the  $\beta$ -score of the text is less than zero, which corresponds to the prevalence of manipulative and negative sentiments in the text) comprise only 18.5% (3.68 out of 19.9 million). The greatest values of the  $\beta$ -score were characterized by publications on sports medicine, systems biology, nutrition, on the use of applied mathematics and data mining in medicine. The rubrication of the entire array of publications by 27,840 headings (MESH-system of PubMed/MEDLINE) indicated an increase in the  $\beta$ -score by years (i.e., the positive dynamics of sentiment quality of the texts of publications) for 27,090 of the studied headings. The most intense positive dynamics was found for research in genetics, physiology, pharmacology, and gerontology. 249 headings with sharply negative dynamics of sentiment quality and with a pronounced increase in the manipulative sentiments characteristic of the tabloid press were highlighted. Separate assessments of international experts are presented that confirm the patterns identified.

**Conclusion.** The proposed artificial intelligence system allows a researcher to make an effective assessment of the sentiment quality of biomedical research papers, filtering out potentially inappropriate publications disguised as "evidence-based".

## KEY WORDS

Big data analysis, pharmacoinformatics, artificial intelligence, evidence-based medicine, machine learning, publication quality assessment algorithms, thematic modeling.

**Received:** 18.12.2019; **in the revised form:** 19.03.2020; **accepted:** 01.06.2020.

## Conflict of interests

The authors declare they have nothing to disclose regarding the funding or conflict of interests with respect to this manuscript.

The authors contributed equally to this article.

## Funding

This work was supported by grants from the Russian Foundation for Basic Research 19-07-00356 17-07-00935 17-07-01419 18-07-01022 18-07-00944 18-07-00929 16-07-01133.

## For citation

Torshin I. Yu., Gromova O. A., Stakhovskaya L. V., Vanchakova N. P., Galustyan A. N., Kobalava Zh. D., Grishina T. R., Gromov A. N., Ilovaiskaya I. A., Kodentsova V. M., Kalacheva A. G., Limanova O. A., Maksimov V. A., Malyavskaya S. I., Mozgovaya E. V., Tapiiskaya N. I., Rudakov K. V., Semenov V. A. Analysis of 19.9 million publications of the PubMed/MEDLINE database using artificial intelligence methods: approaches to the generalizations of accumulated data and to the phenomenon of "fake news". *FARMAKOEKONOMIKA. Sovremennaya farmakoekonomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoconomics and Pharmacoepidemiology*. 2020; 13 (2): 146–163 (in Russ.). <https://doi.org/10.17749/2070-4909/farmakoekonomika.2020.021>.



## Основные моменты

## Что уже известно об этой теме?

- ▶ Англоязычные базы данных PubMed/MEDLINE и Embase являются ценными информационными ресурсами для нахождения оригинальных публикаций по фундаментальной и клинической медицине
- ▶ В настоящее время не существует систем искусственного интеллекта, позволяющих оценивать качество этих публикаций

## Что нового дает статья?

- ▶ Сформулирована методика анализа «больших данных» биомедицинских публикаций, основанная на топологической теории сентимент-анализа медицинских текстов. Проведен анализ текстов абстрактов 19,9 млн публикаций, зарегистрированных в PubMed/MEDLINE, за последние 50 лет
- ▶ Публикации с низким сентимент-качеством составляют всего 18,5% (3,68 из 19,9 млн)
- ▶ Наибольшими значениями  $\beta$ -балла характеризовались публикации по спортивной медицине, системной биологии, нутрициологии, по использованию методов прикладной математики и интеллектуального анализа данных в медицине

## Как это может повлиять на клиническую практику в обозримом будущем?

- ▶ Разработанная система искусственного интеллекта позволяет проводить эффективную оценку сентимент-качества биомедицинских исследований, отфильтровывая потенциально неадекватные публикации, публикуемые под маской «доказательных»
- ▶ Выделенные рубрики с резко отрицательной динамикой сентимент-качества за последние 10 лет указывают на то, что следует с осторожностью относиться к так называемым «международным стандартам»
- ▶ Выборки статей из PubMed/MEDLINE, не содержащие лингвистических признаков манипуляции, весьма важны для разработки клинических рекомендаций, учета международного опыта, поиска медицинских разработок с инновационным потенциалом, реализации национальных проектов

## Highlights

## What is already known about this subject?

- ▶ PubMed/MEDLINE and Embase English-language databases that are valuable information resources of original publications in fundamental and clinical medicine
- ▶ There are currently no artificial intelligence systems to evaluate the quality of these publications

## What are the new findings?

- ▶ The technique of analysis of the “Big data” of biomedical publications is formulated based on the topological theory of sentiment analysis of medical texts. Abstracts of 19.9 million publications registered in PubMed/MEDLINE over the past 50 years were analyzed
- ▶ Publications with low sentiment quality account for only 18.5% (3.68 of 19.9 million)
- ▶ The greatest values of the  $\beta$ -score were characterized by publications on sports medicine, systems biology, nutrition, on the use of applied mathematics and data mining in medicine

## How might it impact on clinical practice in the foreseeable future?

- ▶ The developed artificial intelligence system allows for an effective assessment of the sentiment quality of biomedical research, filtering out potentially inappropriate publications published under the guise of “evidence”
- ▶ Highlighted headings with sharply negative dynamics of sentiment quality over the past 10 years indicate that caution should be taken concerning the so-called “International standards”
- ▶ Subsets of articles from PubMed/MEDLINE that do not contain signs of linguistic manipulation are very important for developing clinical guidelines, taking into account international experience, searching for medical developments with innovative potential, and implementing national projects

## ВВЕДЕНИЕ / INTRODUCTION

... Просматривая газеты за несколько последних лет, я убедился, что некоторые методы уголовного мира, причем самые грязные методы, стали безнаказанно применяться в политике...

«Двадцатый век начинается» (1987),

по мотивам «Записок о Шерлоке Холмсе» А. К. Дойла

Англоязычные базы данных (БД) PubMed/MEDLINE [1] и Embase [2] представляют собой исключительно ценные источники информации по исследованиям из области фундаментальной и клинической медицины. Наличие более 30 млн публикаций, зарегистрированных в БД свободного доступа PubMed/MEDLINE, и более 40 млн публикаций в БД Embase указывает на насущную необходимость использования адекватных алгоритмов для поиска и отбора публикаций, необходимых для того или иного исследования. Такого рода алгоритмы особенно важны для масштабного планирования научных исследований и для разработки соответствующих национальных программ в области здравоохранения.

Существующие алгоритмы поиска БД научных публикаций по ключевым словам позволяют достаточно точно выделять подмножества публикаций, релевантных той или иной теме исследований. Однако ни одна из имеющихся поисковых систем не позволяет оценивать качество многих тысяч находимых таким образом публикаций. А ведь решение проблемы оценки научного качества статьи по ее тексту принципиально важно для нахождения объективной научной информации.

Приведем несколько простых примеров. При работе над нашими монографиями [3,4], каждая из которых цитирует по 2500 источников, только по витамину D нам пришлось провести детальный анализ массива из 12700 публикаций, зарегистрированных

в PubMed/MEDLINE, в т.ч. публикаций результатов соответствующих клинических исследований. При попытке экспертного анализа этого массива, что называется «вручную», для экспертов стало очевидным, что более 30% рассмотренных публикаций носят ярко выраженный негативный характер: ведь в них не представлено вообще никаких положительных клинических эффектов от использования витамина D по отношению к любой из исследованных патологий/исходов. Далее мы называем такие публикации «безрезультатными исследованиями».

Поясним вышесказанное. Очевидно, что в полноценном научном исследовании всегда получается набор положительных и отрицательных результатов. В то же время, представление только отрицательных результатов или чрезмерно оптимистичное «выпирание» какого-то одного положительного результата указывает на научную однобокость исследования, вызывает серьезные сомнения в мотивации такого «исследователя» и, вероятно, указывает на определенного рода конфликт интересов.

Пусть для достаточного числа пациентов имеется таблица из 100 показателей состояния здоровья, включая 10 исследуемых «исходов». Такая таблица подразумевает  $100 \times 99 / 2 = 4950$  попарных взаимодействий показателей, в т.ч.  $10 \times 99 / 2 = 495$  взаимодействий, относящихся к исходам исследования. Очевидно, что данные в такую таблицу включаются не случайным образом, а в соответствии с рациональным дизайном биомедицинского исследования. Поэтому т.н. «модельные» выборки (полученные посредством датчика случайных чисел), в которых параметры совершенно не коррелируют друг с другом, исключены по определению.

Нами проведены анализы более 850 таких таблиц реальных данных, включая результаты оригинальных исследований [3,4], и данные из публично доступных источников (<http://physionet.org> и др.). Накопленный экспертный опыт позволяет однозначно

утверждать, что если при анализе таблицы из 100 параметров не найдено ни одного статистического значимого взаимодействия для 4950 попарных взаимодействий, то такая таблица или (1) содержит грубейшие ошибки, допущенные при сборе данных, или (2) данные в таблице были получены с использованием датчика случайных чисел с равномерным распределением, или (3) допущены грубейшие ошибки при анализе данных – игнорирование фундаментальных математических условий применения того или иного статистического критерия, отсутствие стратификационного анализа, арифметические ошибки (см. цитируемые далее работы проф. Дж. Иоаннидиса) и т.п., или (4) возможна фальсификация декларируемых в публикации результатов (предпринимаемая, например, в целях недобросовестной конкуренции).

Как было отмечено выше, «безрезультатными» будем называть такие исследования, в которых не представлено никаких положительных клинических эффектов (то есть представлены только отрицательные результаты, без каких-либо позитивных выводов). Для целей настоящей статьи важен тот факт, что «негативные» или «безрезультатные» исследования отличаются особым стилем изложения, преобладанием определенной лексики и грамматических конструкций. Текст безрезультатных исследований характеризуется избытком конструкций отрицания («did not», «no significant effect» и др.), излишней формализацией языковых конструкций: например, пациенты могут называться не «пациентами» или «пациентками», а некими бесполоыми «участниками», англ. – «participants», вместо указания четких клинических диагнозов в абстракте и в тексте статьи по МКБ-10 представлено избыточное, навязчивое употребление не клинических терминов наподобие «основная точка» («primary outcome») или некие абстрактные «неврологические события» при полном отсутствии понятной клиницисту терминологии (например, «I64 Инсульт», «балл по шкале MMSE», «шкала Рэнкина», «балл по IPSS» и т.п.). Иначе говоря, «безрезультатные» исследования, в некотором роде, написаны «по шаблону».

Особого внимания заслуживает и то, что тексты определенных публикаций, индексируемых в PubMed/MEDLINE, характеризуются преобладанием штампов манипулятивного характера, которые в англоязычной литературе широко известны как «spin words» («слова для накручивания сознания») или «weasel words» (дословно – «хорьковые слова», то есть слова, используемые для подмены обсуждения по существу на психологическое давление на читателя). Сам термин «weasel words» восходит к шекспировской фразе «I can suck melancholy out of a song, as a weasel sucks eggs» («Я могу высосать меланхолию из песни как хорек высасывает яйца» – В. Шекспир, комедия «Как вам это понравится», 2-й акт) и подразумевает слова, которые «высасывают смысл из слов рядом с ними и оставляют только треск пустой скорлупы» [5].

Например, словосочетание «adverse outcomes» (что можно перевести как «неблагоприятные исходы») является «хорьковым словом», типичным не только для безрезультатных исследований, но и для бульварной прессы [6]. Многочисленные примеры, приводимые в монографиях [6,7], наглядно показывают, что данное словосочетание повсеместно используется в желтой прессе исключительно для запугивания читателя, а не для представления какой-либо важной информации. К сожалению, как показывают результаты настоящего исследования, лексика англоязычных таблоидов (то есть бульварных листов, см. полный список на сайте www.earthnewsletters.com) весьма широко используется и в некоторых англоязычных медицинских журналах. Существование упоминаемых выше и многих других языковых «шаблонов» и «штампов» позволяет предположить, что процесс отделения манипулятивных публикаций от нормальных публикаций может быть автоматизирован посредством использования современных методов машинного обучения и интеллектуального анализа данных.

В настоящее время не существует систем искусственного интеллекта, позволяющих оценивать качество этих публикаций. Это определило цель нашей работы.

**Цель** – разработка и апробация системы для проведения сентимент-анализа (анализа эмоциональной модальности) публикаций по биомедицине.

## МАТЕРИАЛЫ И МЕТОДЫ / MATERIALS AND METHODS

### Структура системы сентимент-анализа в рамках машинного обучения

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 полный научной публикации) состоит из признакового описания (то есть 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.

### Оценка разрешимости задач классификации текстов

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

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

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  $\alpha$ -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  $\alpha$  функционал  $X_{\text{инф}}(\alpha)$ , позволяющий оценить информативность  $\alpha$ -го признака относительно классов  $K_1/K_2$  [10]. Тогда множество всех значений признаков  $P$  можно линейно упорядочить в соответствии со значением оценки информативности (то есть по убыванию значений  $D_{\text{инф}}$ ). После упорядочения множества  $P$  большим значениям  $D_{\text{инф}}(\alpha)$  соответствуют меньшие значения ранга информативности  $\alpha$ , так что условие (1) записывается как критерий разрешимости на упорядоченном множестве признаков:

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

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

В качестве  $D_{\text{инф}}$  был применен один из  $D$ -функционалов [11–15], 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, «адювантная терапия»).

Значения  $D_{инф}$  для различных ключевых слов вычислялись следующим образом. Пусть  $\alpha$ -ый термин встречается в  $n_1^\alpha$  абстрактах выборки  $K_1$ , и в  $n_2^\alpha$  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):

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

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

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

Некоторые признаки могут быть избыточны, так как встречаются в устойчивых сочетаниях слов (например, «лекарственная терапия»), семантически связанных терминах или являются синонимами и др. Поэтому при анализе условия (2) важно принимать во внимание парные взаимодействия терминов. Для произвольного класса  $K$  «расстояние» между  $i$ -ым и  $j$ -ым 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) будут фигурировать уже не отдельные слова/термины, а кластера терминов, установленные методом анализа метрических сгущений [9, 11–13].

Вычисление условия (2) на множестве прецедентов и лежит в основе использованного метода классификации текстов. Условие (2) позволяет вычислить множество  $K_0 \subset P$ , то есть отобрать наиболее информативные признаки, гарантирующие разрешимость задачи классификации. Далее, с применением методов порождения синтетических признаков и прогнозирования числовых переменных [16], использующих алгоритмы логических правил, ранжирования, нейронных сетей и регрессионных подходов, над множеством признаков  $K_0$  строятся алгоритмы, позволяющие различать тексты классов  $K_1/K_2$ .

### Сентимент-анализ текстов

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 (текста)  $A$  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» (так как 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 2529638 текстов были выделены 7116 языковых конструкций, являющихся наиболее информативными лингвистическими признаками, позволяющими с достаточной точностью распознавать исследованные классы текстов (среднее значение площади под кривой работы алгоритма – 0,88).

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/MEDLINE, поддержанных некоммерческими организациями-ино-

Таблица 1. Классы текстов с сентиментами, изученные в настоящем исследовании.

Table 1. Classes of sentiment texts analyzed in the present study.

№	Классы текстов	N	AUC	m	β-балл
1	Манипулятивные обороты	12201	0,90	74	β <sup>1</sup> , βманип.
2	Адекватное изложение клинических результатов	15827	0,87	25	β <sup>2</sup> , βклин.
3	Положительные результаты	14394	0,84	13	β <sup>3</sup> , βполж.
4	Безрезультатные исследования	5672	0,96	16	β <sup>4</sup> , βотриц.
5	Пропаганда	14229	0,88	202	β <sup>5</sup> , βпроп.
6	Некоммерческие организации-иноагенты	4525	0,89	38	β <sup>6</sup> , βНКО
7	Подделка результатов	6733	0,83	101	β <sup>7</sup> , βотозв.
8	Негативное личное отношение	18920	0,91	132	β <sup>8</sup> , βнег.
9	Позитивное личное отношение	20992	0,92	143	β <sup>9</sup> , βпоз.
10	Сеяние раздора, агрессии, массовых беспорядков	8830	0,86	199	β <sup>10</sup> , βраздор
11	Язык тоталитарных сект	5221	0,84	783	β <sup>11</sup> , βсекта
12	Позитивный эмоциональный фон	459993	0,83	1790	β <sup>12</sup> , βпоз.эм.
13	Негативный эмоциональный фон	475952	0,93	1117	β <sup>13</sup> , βнег.эм.
14	Конструктивный диалог	459993	0,82	2009	β <sup>14</sup> , βконстр.
15	Неконструктивный диалог	475952	0,89	113	β <sup>15</sup> , βнеконстр.
16	Дипломатический «слэнг»	530204	0,88	361	β <sup>16</sup> , βдипл.
	Итого	2529638	0,88	7116	–

Примечание. 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).

Note. N is the number of texts in the training sampling; AUC – area under the curve as an estimate of the accuracy of the algorithm in cross-validation; m is the total number of identified the most informative features (language structures, sentiments); β-score – designation of the corresponding β-score (see text).



агентами, официально зарегистрированными в госдепартаментах соответствующих стран, информация с сайтов [www.state.gov](http://www.state.gov), [www.ngoadvisor.net](http://www.ngoadvisor.net) др.), тексты из англоязычных ресурсов, приводимых в монографиях [6,7,17–19] и на сайте [www.earthnewspapers.com](http://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 текста каждого исследования  $\beta(\text{текст}) = \beta_+(\text{текст}) - \beta_-(\text{текст})$  вычислялась как разность между значением балльной оценки «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).

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 известного австрийского философа науки К. Р. Поппера [20,21]. Согласно последнему, утверждение относится к области науки тогда и только тогда, когда существует возможность его экспериментального опровержения с использованием доступного в настоящее время научного инструментария. Если эксперт использует критерий фальсифицируемости и помещает в класс  $K_1$  тексты, удовлетворяющие данному критерию (соответственно, в класс  $K_2$  помещаются тексты, не удовлетворяющие этому критерию), то формулы (1–3) позволяют «обучать» алгоритмы для решения задачи классификации «научных» и «псевдонаучных» текстов.

К. Р. Поппер достаточно мягко относился к проблеме псевдонаучных публикаций («... наука часто ошибается, и псевдонаука может случайно натолкнуться на истину») [21]. Тем не менее для здравомыслящего ученого очевидна недопустимость использования излишне эмоциональных и уж тем более манипулятивных и пропагандистских языковых конструкций (сентиментов) в тексте научной статьи. В настоящем исследовании мы считаем, что преобладание такого рода сентиментов косвенно указывает на псевдонаучность и даже лженаучность соответствующего текста – ведь излишняя эмоциональность (1) ассоциирована с нарушениями правил 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 нарушения принципа фальсифицируемости. Поэтому сформулированные выше количественные оценки сентимент-качества текстов ( $\beta_1 \dots \beta_{16}$  в таблице 1,  $\beta_+$  и  $\beta_-$ ) могут интерпретироваться как количественные оценки научности/псевдонаучности текстов в соответствии со следующими критериями:

- Изложение научного результата должно использовать конструктивную лексику ( $\beta_{14}$ ), содержать хотя бы какие-нибудь положительные результаты ( $\beta_3$ ) и проводиться на основе специальной терминологии соответствующей научной области ( $\beta_2$ );
- Излишне депрессивный эмоциональный фон текста ( $\beta_-, \beta_4, \beta_8, \beta_{13}$ ) несовместим с конструктивным изложением результатов ( $\beta_4$ );
- Эмоциональный фон научного текста должен быть нейтральным ( $\beta_{12}, \beta_9 \approx 0$ ) или слегка положительным ( $\beta_+ > 0$ , что интуитивно ясно, так как настоящий исследователь всегда радуется новому открытию);

- Направленная манипуляция эмоциями (характерная, в частности, для ретрагированных исследований,  $\beta_7$ ) абсолютно недопустима;

- Научный текст не может содержать языковых конструкций пропагандистского характера ( $\beta_5$ ) или конструкций, указывающих выраженное «личное отношение», в особенности эмоциональное приятие/неприятие обсуждаемого результата ( $\beta_8$ );

- Изложение научного результата не может содержать манипулятивных приемов изложения ( $\beta_1$ ), в т.ч. используемых при вербовке прозелитов тоталитарных сект ( $\beta_{11}$ ), в работе некоммерческих организаций (НКО), являющихся иноагентами ( $\beta_6$ ), при организации массовых митингов ( $\beta_5$ ) или массовых беспорядков ( $\beta_{10}$ );

- Манипулятивные языковые конструкции, повсеместно используемые политиками или средствами массовой информации ( $\beta_{15}$ ), неприемлемы для научных текстов. Двухичный языковый стандарт дипломатического слэнга также недопустим ( $\beta_{16}$ ).

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 сентиментов 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 [3,4,22–25] и показала свою практическую эффективность.

На основании разработанных алгоритмов проведен анализ массива из 19918308 публикаций, найденных в базе данных PubMed/MEDLINE по запросу «(1970 [Date – Publication]: 2019 [Date – Publication])». Массив включал аннотированные публикации в англоязычных рецензируемых научных журналах с полным текстом абстракта. Анализ этого массива текстов был проведен с использованием описанных выше современных методов машинного обучения, основанных на топологической теории сентимент-анализа (то есть анализа эмоциональной модальности текста), развиваемой в рамках алгебраического подхода к распознаванию и классификации.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ / RESULTS AND DISCUSSION

### Общая оценка массива публикаций в PubMed/MEDLINE

Была изучена встречаемость 16 классов сентиментов (манипулятивные обороты речи, пропаганда, негативное личное отношение и др. (см. табл 1)). В течение всего этого периода количество публикаций с низким показателем сентимент-качества ( $\beta$ -балл текста ниже нуля, что соответствует преобладанию манипулятивных и негативных сентиментов) росло гораздо медленнее, чем общее число публикаций в PubMed/MEDLINE (рис. 1).

Визуальный анализ распределений значений  $\beta$ -балла (рис. 2) показывает, что полученный результат весьма позитивен, ведь большинство текстов публикаций, индексируемых в PubMed/MEDLINE (более 80%), характеризуются положительными значениями  $\beta$ -балла.

Анализ распределения значений  $\beta_+$  и  $\beta_-$  (рис. 3) показывает, что большинство исследований (более 90%, включая некоторые из исследований со значениями  $\beta < 0$ ) характеризуются низкими значениями и балла  $\beta_+$ , и балла  $\beta_-$ . Иначе говоря, большинство исследований в PubMed/MEDLINE практически не содержат манипулятивных или эмоционально перегруженных конструкций, будь то положительные или отрицательные эмоциональные коннотации. Этот факт соответствует приемлемому качеству подавляющего числа публикаций в PubMed/MEDLINE и еще раз показывает, что PubMed/MEDLINE – ценнейший ресурс для поиска информации по фундаментальной медицине, клинической медицине и по смежным областям (науки о живых системах).



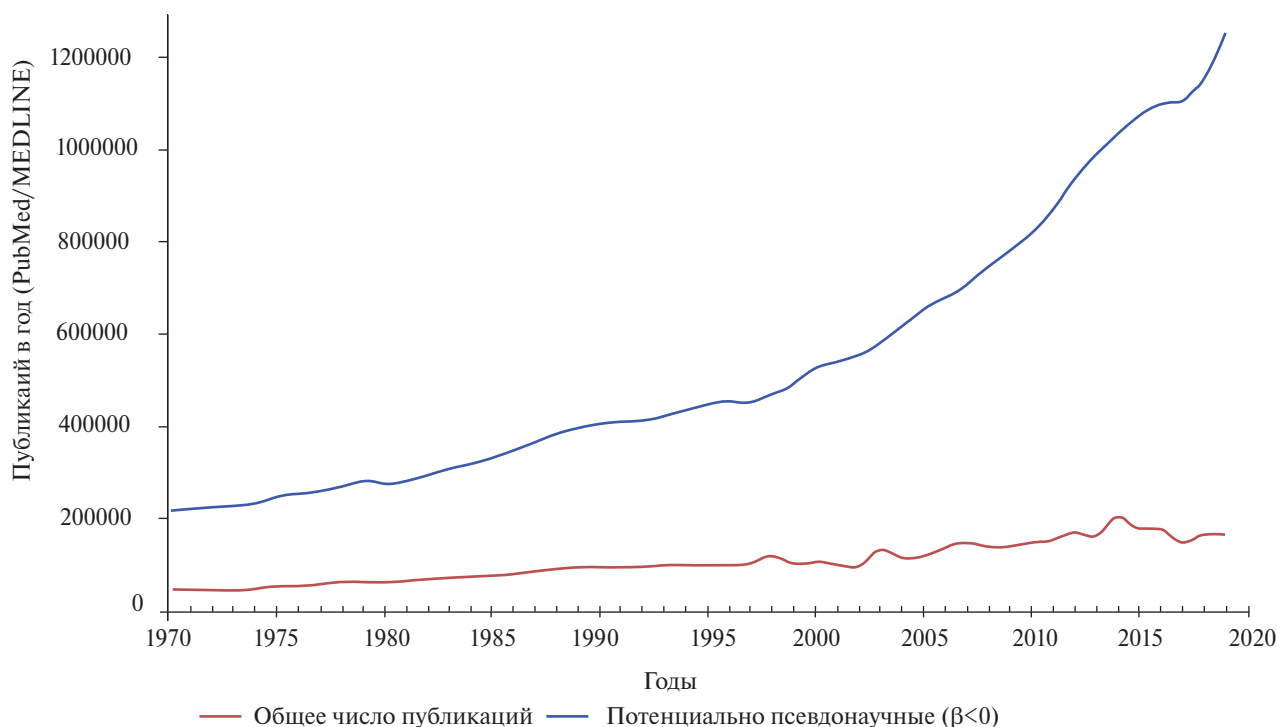


Рисунок 1. Динамика публикаций в PubMed/MEDLINE, в т.ч. публикаций с отрицательным  $\beta$ -баллом.

Figure 1. The dynamics of publications in PubMed/MEDLINE, including publications with a negative  $\beta$ -score.

**Общая характеристика публикаций с отрицательным значением  $\beta$ -балла**

Тем не менее 18,5% публикаций в PubMed/MEDLINE (3,68 млн из 19,92 млн) все же характеризовались отрицательными значениями  $\beta$ -балла (рис. 4). Интересно отметить, что наибольшим вкладом в отрицательное значение  $\beta$ -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 жесткой 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 встречаться и в профессионально написанных статьях по тренировке медицинского персонала или

по психологии. Однако определенные сочетания этих терминов, наряду с избыточно частым их употреблением, распознаются разработанными алгоритмами анализа текстов как языковые шаблоны, характерные для менеджмента адептов тоталитарных сект, 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  $\beta_{10}$  в таблице 1) также нередко встречаются в предположительно «научной» литературе ( $\beta_{10}=2,89$ ): «resistance forces» (силы сопротивления), «political» в сочетании с «repressions» (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 таблицы 2).

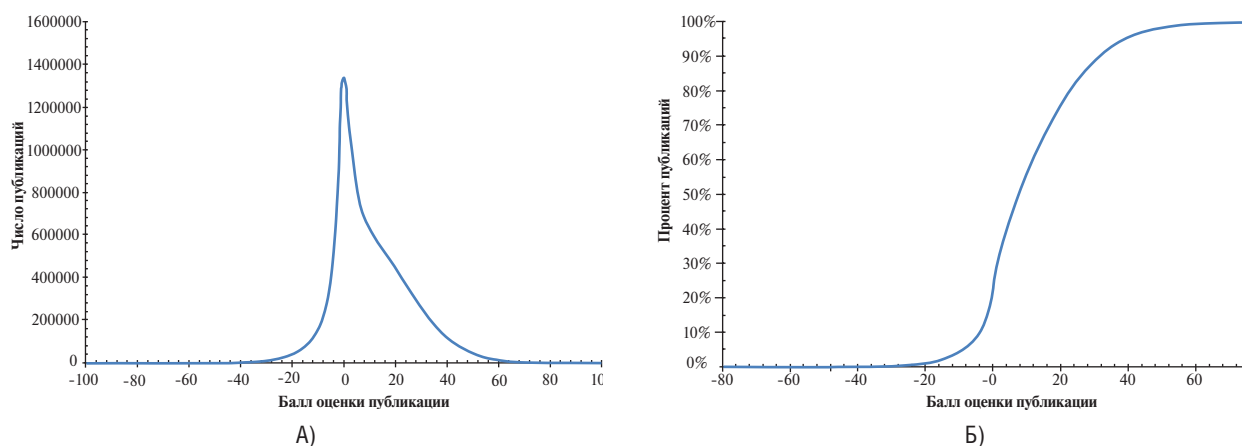


Рисунок 2. Эмпирическая функция распределения (э.ф.р.) значений  $\beta$ -балла по 19,9 млн публикаций в PubMed/MEDLINE: А) Плотность вероятности э.ф.р.; Б) Интегральная форма э.ф.р.

Figure 2. Empirical distribution function (EDF) of  $\beta$ -score values for 19.9 million publications in PubMed/MEDLINE: А) probability density; Б) The integral form of EDF.

Данная интернет-версия статьи была скачана с сайта <http://www.pharmacoeconomics.ru>. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

Негативные сентименты ( $\beta_{13}=4,92$ ) также были весьма типичны для текстов с  $\beta < 0$ . Действительно, чрезмерно частое употребление конструкций на основе таких словосочетаний как «waste» (понапрасну), «is poorly» ([есть] плохо), «the worst» (наихудшее), «poor quality» (очень низкого качества), «a worse» (еще хуже), «zero potential» (нулевой потенциал), «has to avoid» (должно избегать) и др. может целенаправленно настраивать эмоции читателя такого текста на негативное отношении к затрагиваемой в тексте публикации научной теме.

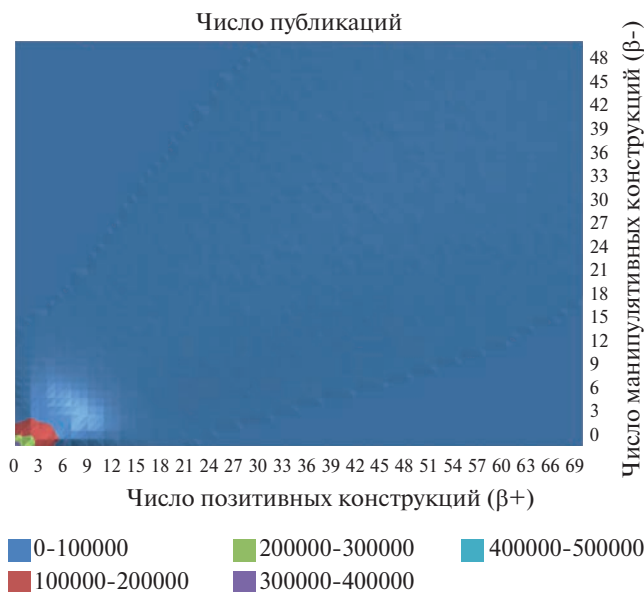
Длины прямоугольников на нижней полосе пропорциональны значениям соответствующих  $\beta$ -баллов (см. табл. 1).

**Распределение по странам происхождения публикаций**

Анализ стран, из которых исходят англоязычные публикации по биомедицине (рис. 5), показал, что наибольшими значениями  $\beta$ -балла характеризовались тексты публикаций из ОАЭ ( $\beta=15,46$ ), Ирана ( $\beta=13,67$ ) и Тайваня ( $\beta=12,14$ ), затем следовали 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.

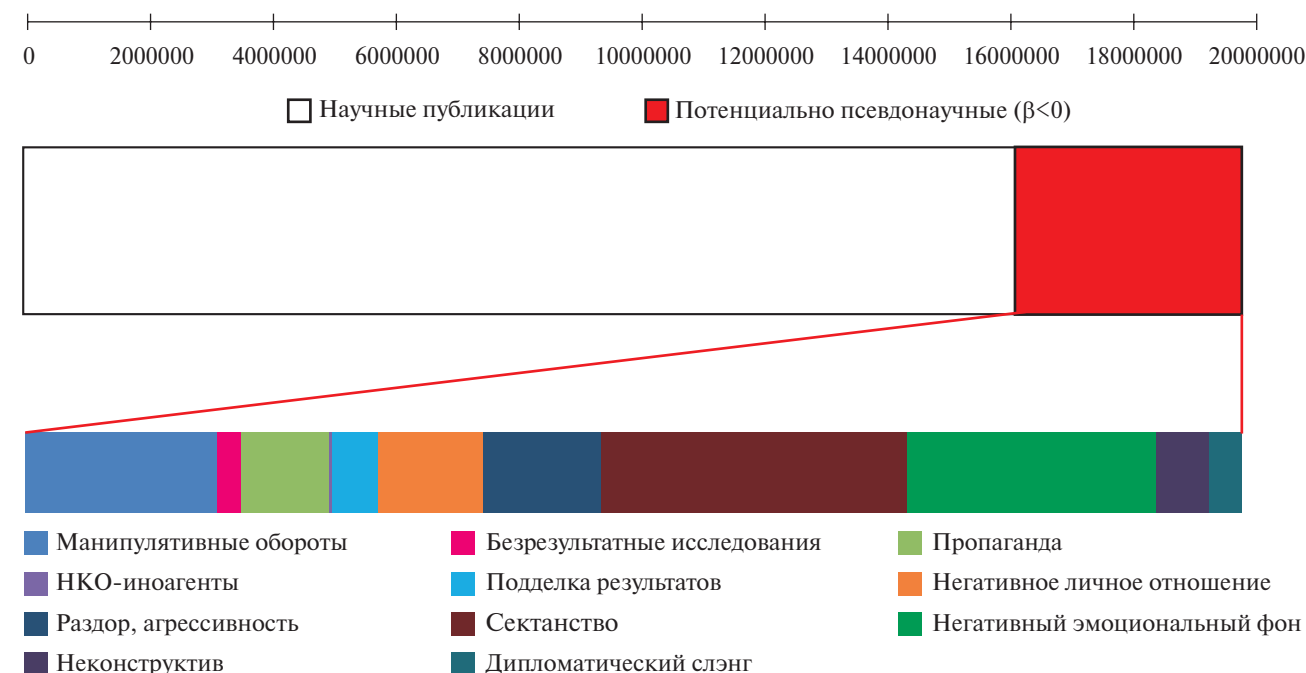
**Распределение публикаций по рубрикам MESH и по научным направлениям**

Рубрикация всего массива публикаций по 27840 рубрикам MESH-системы PubMed/MEDLINE позволила выявить рубрики, наиболее затронутые и наименее затронутые манипулятивными сентиментами. Практически все 3,68 млн текстов с отрицательными



**Рисунок 3.** Распределение значений  $\beta+$  и  $\beta-$  по 19,9 млн публикаций в PubMed/MEDLINE (вид сверху).  
Примечание. Цветовая шкала отображает число публикаций с соответствующими значениями позитивных (балл  $\beta+$ ) и манипулятивных (балл  $\beta-$ ) языковых конструкций. Очевидно преобладание статей с достаточно малыми значениями  $\beta+$  и  $\beta-$  (0-6 баллов).

**Figure 3.** Distribution of  $\beta+$  and  $\beta-$  values over 19.9 million publications in PubMed/MEDLINE (top view).  
Note. The color scale displays the number of publications with the corresponding values of positive (score  $\beta+$ ) and manipulative (score  $\beta-$ ) language constructs. Obviously, there is a predominance of articles with fairly low values of both  $\beta+$  and  $\beta-$  score (0-6 points).



**Рисунок 4.** Вклад различных негативных сентиментов в формирование отрицательных значений  $\beta$ -балла для 3,68 млн из 19,92 млн публикаций в PubMed/MEDLINE.  
Примечание. Длины прямоугольников на нижней полосе пропорциональны значениям соответствующих  $\beta$ -баллов (см. табл. 1).

**Figure 4.** Contribution of various negative sentiments to the formation of negative  $\beta$ -score for 3.68 million of 19.92 million publications in PubMed/MEDLINE.  
Note. The lengths of the rectangles in the lower strip are proportional to the values of the corresponding  $\beta$ -ballows (see table 1).

ми значениями  $\beta$ -балла попали в 3520 из 27840 рубрик PubMed/MEDLINE. Наиболее характерные примеры таких рубрик приведены в таблице 2.

Манипулятивная англоязычная лексика наиболее распространена в публикациях в разделах «экономика» ( $\beta=-17,15$ ) и «юриспру-

денция» ( $\beta=-12,60$ ). 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» ( $\beta=-3,61$ ), легализацию эвтаназии ( $\beta=-1,53$ ) и т.п. Чрезвычайно интересно от-

Данная интернет-версия статьи была скачана с сайта <http://www.pharmacoeconomics.ru>. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: info@irbis-1.ru.

**Таблица 2.** Примеры рубрик PubMed/MEDLINE с наиболее отрицательными и наиболее положительными значениями  $\beta$ -балла по всем публикациям в данной рубрике за последние 50 лет (1970–2019).

**Table 2.** Examples of PubMed/MEDLINE headings with the most negative and most positive  $\beta$ -score values for all publications in this heading over the past 50 years (1970–2019).

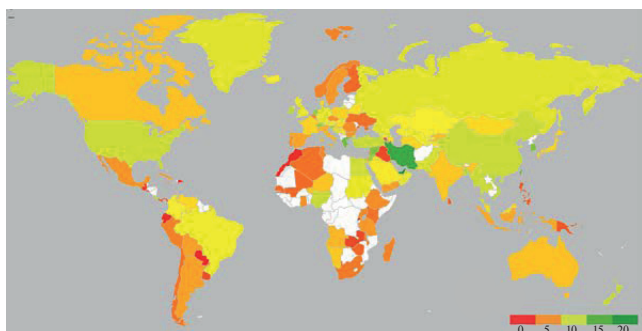
Рубрика PubMed/MEDLINE	Статей	$\beta$ -балл	Рубрика PubMed/MEDLINE	Статей	$\beta$ -балл
MH – *Economics	7855	– 17,15	MH – *Resistance Training	2282	28,94
MH – *Jurisprudence	13221	– 12,60	MH – Muscle Strength/*physiology	4561	28,81
MH – Health Care Costs/*statistics & numerical data	4225	– 6,41	MH – Systems Biology/*methods	2249	26,03
MH – Mental Health Services/*economics	871	– 4,30	MH – *Molecular Dynamics Simulation	8896	25,77
MH – *Drug Costs	2671	– 3,86	MH – Learning/*physiology	10886	24,70
MH – *Policy Making	4876	– 3,79	MH – *Metabolomics	2106	23,06
SO – Fam Plann Perspect	1389	– 3,61	SO – Artif Intell Med	1309	23,05
MH – *Insanity Defense	918	– 2,73	MH – Athletic Performance/*physiology	3640	22,82
MH – Health Policy/*legislation & jurisprudence	3943	– 2,59	MH – Maternal Behavior/*physiology	937	22,79
MH – Commitment of Mentally Ill/*legislation & jurisprudence	2055	– 2,29	MH – Dietary Supplements/analysis	1095	22,69
MH – Homosexuality, Male/*statistics & numerical data	1163	– 2,22	MH – Maze Learning/*physiology	2007	22,17
MH – *Medication Errors	2358	– 2,07	MH – Drug Resistance, Multiple, Bacterial/*genetics	1909	22,12
MH – Condoms/*statistics & numerical data	2582	– 2,04	OT – NAFLD	1148	21,83
SO – Cochrane Database Syst Rev	12405	– 1,86	OT – BDNF	1905	21,79
MH – Mental Competency/*legislation & jurisprudence	1250	– 1,70	MH – Proteomics/*methods	14267	21,41
MH – Developing Countries/*economics	916	– 1,58	MH – Muscle, Skeletal/*physiology	16346	20,98
MH – Suicide, Assisted/*legislation & jurisprudence	1206	– 1,53	MH – *Finite Element Analysis	3408	20,86
MH – *Population Control	1326	– 1,48	MH – Pattern Recognition, Automated/*methods	14042	20,84
MH – *Abortion, Legal	2394	– 1,33	MH – Gait/*physiology	8099	20,78
MH – HIV Infections/*transmission	2754	– 1,23	MH – *Gene Expression Profiling	21965	20,71
MH – Violence/*legislation & jurisprudence	108	– 1,28	MH – Tandem Mass Spectrometry/*methods	11254	20,70
MH – *Conflict of Interest	3680	– 1,03	MH – Physical Endurance/*physiology	5745	20,70
MH – Costs and Cost Analysis	34779	– 0,94	MH – Computational Biology/*methods	17691	20,51
MH – Child Abuse/*legislation & jurisprudence	1214	– 0,88	MH – *Principal Component Analysis	986	20,32
MH – *Lobbying	2123	– 0,83	MH – *Droughts	2520	20,10
MH – *Fraud	1495	– 0,69	MH – Artificial Intelligence	5325	19,59
MH – *Scientific Misconduct	2196	– 0,67	MH – Nonlinear Dynamics	5409	19,41
SO – N Y Times Web	2170	– 0,64	MH – Caregivers/*psychology	11592	19,24
MH – Marketing of Health Services/*methods	1882	– 0,62	MH – Spectroscopy, Fourier Transform Infrared/*methods	2802	19,24
MH – Civil Rights/*legislation & jurisprudence	1724	– 0,61	MH – *Neural Networks (Computer)	14038	19,24
MH – *Sterilization, Reproductive	1674	– 0,59	MH – Energy Intake/*physiology	2905	19,11
MH – *Peer Review, Research	1597	– 0,58	MH – *Problem-Based Learning	2226	19,09
MH – *Abortion, Induced	7870	– 0,54	SO – Proteomics	5841	19,06
MH – Fees, Dental	1030	– 0,53	MH – Thermodynamics	22139	19,05
MH – Evidence-Based Medicine/*standards	1392	– 0,52	MH – Soccer/*physiology	1543	19,04
OT – *Family Planning Programs	2348	– 0,44	MH – Knee Joint/physiology	2528	18,96
MH – *Euthanasia	3079	– 0,40	MH – Genome-Wide Association Study/*methods	1825	18,85
MH – *Embryo Research	878	– 0,31	SO – FEBS J	5728	18,83
MH – State Medicine/standards	1088	– 0,30	MH – Gastrointestinal Tract/*microbiology	2647	18,82
MH – *Psychotropic Drugs	1690	– 0,27	MH – Insulin Resistance	7251	18,79
Итого	149420	– 2,23	Итого	255052	21,27

метить, что даже в случае статей из чисто публицистических изданий, каким-то образом попавших в PubMed/MEDLINE (например, «Нью Йорк Таймс», рубрика «SO – N Y Times Web»), среднее значение  $\beta$ -балла было намного выше ( $\beta = -0,64$ ), чем в случае отдельных «научных» журналов (табл. 3).

В то же время наибольшими положительными значениями  $\beta$ -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,

Данная интернет-версия статьи была скачана с сайта <http://www.pharmacoeconomics.ru>. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: [info@irbis-1.ru](mailto:info@irbis-1.ru).



**Рисунок 5.** Распределение значений  $\beta$ -балла по странам (среднее за последние 50 лет, 1970–2019).  
Примечание. Цветовая шкала внизу рисунка кодирует значение  $\beta$ -балла (красный – 0–5 и т.д.).

**Figure 5.** The distribution of  $\beta$ -score values by countries (average over the past 50 years, 1970–2019).  
Note. The color bar at the bottom of the figure encodes the value of the  $\beta$ -score (red – 0–5, etc.).

Automated/\* methods», «Neural Networks (Computer)», «Problem-Based Learning», «Artificial Intelligence»).

### Распределение публикаций по годам

Представляет интерес рассмотреть, как изменялась динамика  $\beta$ -балла публикаций по годам по каждой из исследованных 27840 рубрик PubMed/MEDLINE. Для этого временные ряды в координатах «год- $\beta$ -балла» для каждой из рубрик были представлены в виде символьных последовательностей в рамках подхода символьной динамики [26]. Парные расстояния между полученными символьными последовательностями высчитались с использованием метрики Левенштейна [27] и была проведена кластеризация на основе метода анализа метрических сгущений [11–13]. В результате проведения кластеризации положительная динамика сентимент-качества исследований (то есть систематическое повышение  $\beta$ -балла по годам) была установлена для 27090 из 27840 исследованных рубрик (97,3%). Наиболее интенсивная положительная динамика найдена для текстов исследований по генетике, физиологии, фармакологии и геронтологии (рис. 6А). Выделены 249 рубрик с резко отрицательной динамикой качества и с выраженным нарастанием манипулятивных сентиментов, подобных встречающихся в «желтой прессе» западных стран (отдельные примеры этих рубрик приведены на рисунке 6Б).

### «Пираты» в доказательной медицине?

Наиболее характерной и весьма неожиданной оказалась динамика рубрики PubMed/MEDLINE «SO – Cochrane Database Syst Rev». В самом деле, до 2008 г.  $\beta$ -балл публикаций под этой рубрикой колебался вблизи вполне приемлемых значений +1,5...+2,2. После 2008 г. началось постоянное снижение значений  $\beta$ -балла, так что в 2013 г. средний  $\beta$ -балл публикаций в этой рубрике приобрел отрицательное значение. После 2013 г. началось ещё более стремительное падение значений  $\beta$ -балла (на 5,1 балла/год), так что к концу 2019 г. среднее значение  $\beta$ -балла по этой рубрике составило –32,0. В то же время почти все остальные рубрики, имеющие непосредственное отношение к доказательной медицине («Evidence-Based Medicine/\* methods», «Evidence-Based Medicine/\* education», просто «\*Evidence-Based Medicine» и др.), характеризовались отчетливой положительной динамикой  $\beta$ -балла (рис. 6Б). Исключение составила рубрика «Evidence-Based Medicine/\* standards» (стандарты доказательной медицины), в которой также отмечена отрицательная динамика.

Столь резкое изменение сентимент-качества публикаций из рубрики «SO – Cochrane Database Syst Rev» и отрицательная динамика в рубрике «Evidence-Based Medicine/\* standards» за последние шесть лет полностью соответствуют концепции «рейдерского захвата доказательной медицины» (англ. hijack) т.н. «пиратами»

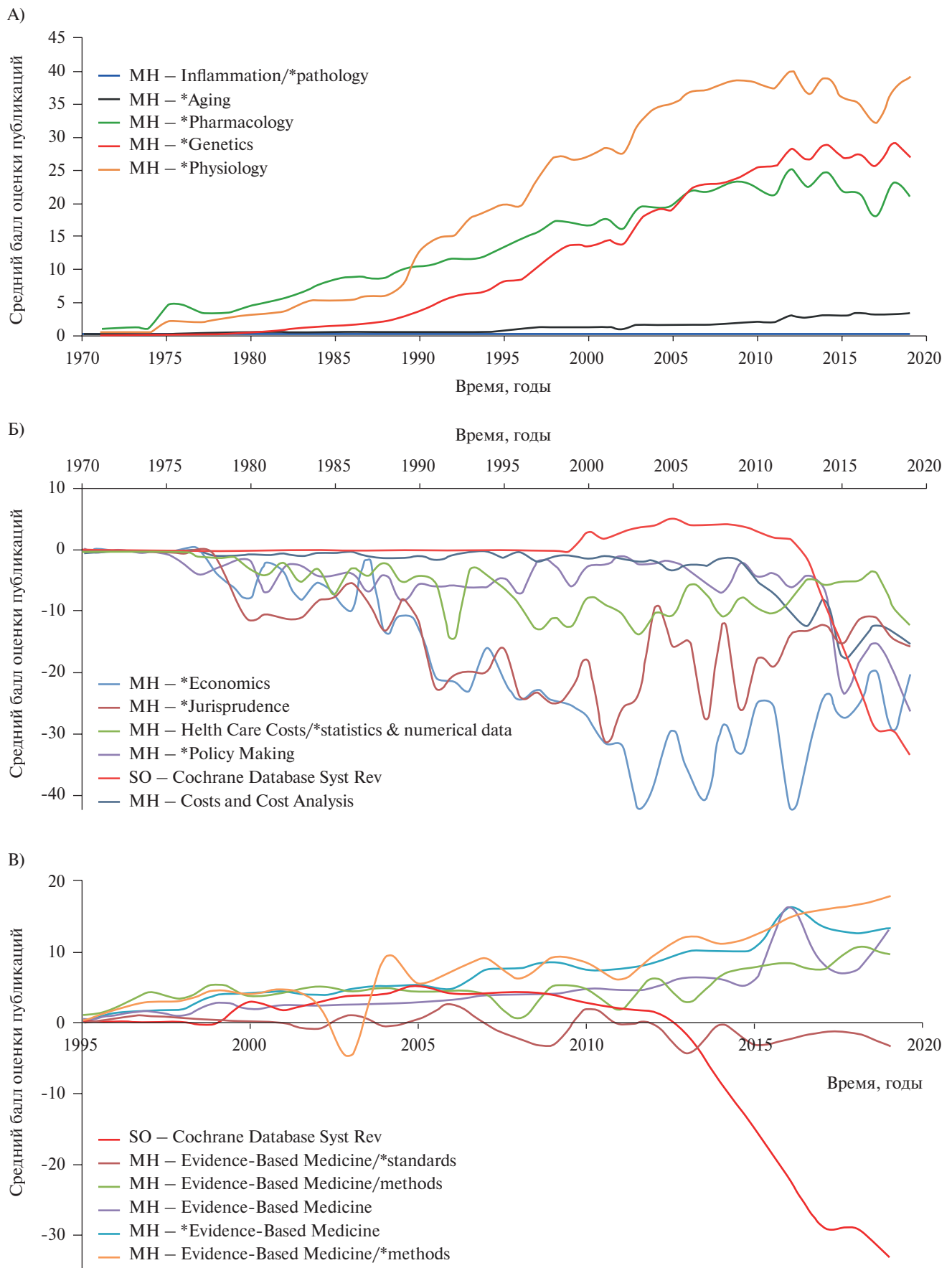
**Таблица 3.** Примеры отдельных англоязычных научных журналов с наиболее отрицательными и наиболее положительными значениями  $\beta$ -балла по всем публикациям в за последние 50 лет (1970–2019).

**Table 3.** Examples of English-language scientific journals with the most negative and the most positive values of  $\beta$ -score averaged over all available publications during the last 50 years (1970–2019).

Журнал	Число	$\beta$ -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	1357	–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	1389	–3,61
SO – Lancet HIV	311	–2,74
SO – Bioethics	1394	–2,07
SO – Cochrane Database Syst Rev	12405	–1,86
SO – AIDS Treat News	808	–1,69
SO – AIDS Policy Law	4225	–1,65
SO – Evid Based Dent	842	–1,10
SO – Issues Reprod Genet Eng	54	–0,94
SO – Harv Health Lett	1635	–0,84
SO – Autism Res	587	21,14
SO – Gigascience	237	21,27
SO – J Proteome Res	6573	21,54
SO – Database (Oxford)	633	21,71
SO – Clin Epigenetics	390	21,85
SO – Wiley Interdiscip Rev RNA	429	21,87
SO – Ecology	3693	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	1225	22,82
SO – Eur J Sport Sci	686	22,84
SO – Nutr Diabetes	250	22,89
SO – Sports Med	2396	23,84
SO – Front Plant Sci	5955	23,85
SO – BMC Biotechnol	1207	23,99
SO – J Comput Phys	59	24,49
SO – BMC Microbiol	3104	24,92
SO – PLoS Comput Biol	4966	25,19
SO – Metabolomics	310	27,39
SO – BMC Genomics	9555	28,11
SO – Microbiome	321	30,80
SO – BMC Syst Biol	1714	35,97

(термин работы [28]). Проф. Дж. П. Иоаннидис с соавторами отмечают в работе [28], 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»





**Рисунок 6.** Примеры динамики  $\beta$ -балла по различным категориям PubMed/MEDLINE: А) Примеры положительной динамики, Б) примеры отрицательной динамики, В) примеры из области доказательной медицины.

**Figure 6.** Examples of  $\beta$ -score dynamics for various PubMed/MEDLINE categories: А) examples of positive dynamics, Б) examples of negative dynamics, В) examples from the field of evidence-based medicine.

Данная интернет-версия статьи была скачана с сайта <http://www.pharmacoeconomics.ru>. Не предназначено для использования в коммерческих целях. Информацию о репринтах можно получить в редакции. Тел.: +7 (495) 649-54-95; эл. почта: [info@irbis-1.ru](mailto:info@irbis-1.ru).

[28]. По результатам настоящего исследования можно предположить, что обсуждаемый «рейдерский захват», возможно, происходил в 2013–2014 гг. (см. рис. 6B).

Описанные проф. Иоанндисом «медицинские пираты», как правило, не имеют отношения к реальной клинической практике, но обладают существенным медийным ресурсом для навязывания своих мнений сотням тысяч практикующих врачей во всем мире [29]. Поэтому ведущие аналитики в области доказательной медицины обоснованно утверждают, что в настоящее время «метаанализы по типу кохрейновских не являются сколько-нибудь доказательными» [30] хотя бы вследствие повсеместного использования устаревших и, более того, ошибочных методов анализа данных. Принимая во внимание то, что в Австралии, Канаде, ЕС и США официально зарегистрирована торговая марка «Кохрейн» [31] и что текстовые продукты интеллектуальной деятельности под данной торговой маркой пользуются популярностью среди врачей, упоминаемые выше «пираты» действительно захватили весьма ценный медийный ресурс. При этом резко упал уровень дискурса: в отличие от сдержанного и более или менее положительного дискурса до 2008 г. ( $\beta=1,5-2,2$ ) уровень дискурса «упал» до откровенно манипулятивного нагнетания негативных эмоций, агрессии, сектантской нетерпимости и т.п. ( $\beta=-32,0$  в 2019 г.).

Типичным примером «пиратского» продукта, произведенного после этого «захвата», является опубликованная в 2018 г. статья под кодовым номером «CD003177». Данная статья, детально проанализированная нами ранее [32], является своего рода «абсолютным лидером» среди публикаций с отрицательным  $\beta$ -баллом ( $\beta=-42$ , в PubMed/MEDLINE таких статей менее 0,5%, см. распределение на рисунке 2). Для данного текста весьма высоки количества сентиментов манипулятивного характера ( $\beta_1=55$ ); сентиментов, отражающих негативное личное отношение ( $\beta_8=35$ ); сентиментов, формирующих негативный эмоциональный фон ( $\beta_{13}=37$ ); сентиментов, провоцирующих раздор/агрессию ( $\beta_{10}=23$ ), и сентиментов, соответствующих использованию «сектантской» лексики ( $\beta_{11}=32$ ). Даже взятые по отдельности эти значения существенно превышают суммарный отрицательный балл ( $\beta$ ) практически всех остальных текстов в PubMed/MEDLINE (99,98% из 19,9 млн исследованных публикаций – см. рисунок 3B).

Детальный экспертный анализ текста «CD003177» указал на грубейшие нарушения, допущенные при сборе и анализе данных, включая: (1) отсутствие четкой клинической логики, (2) от-

сутствие анализа клинической неоднородности когорт и (3) отсутствие знания простейших основ фармакологии и фармации (например, различий между лекарственным препаратом, БАД и пищевым продуктом) [32]. Сделанные в работе [32] 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 (см. выше), вычисленными для данной публикации.

### Оценка массива публикаций по витамину D

Описанные проблемы со «стандартами доказательной медицины» (упомянутая выше рубрика «Evidence-Based Medicine/\*standards») отнюдь не ограничены обсуждаемыми выше примерами. Используемая в настоящей статье методология сентимент-анализа медицинских текстов позволяет устанавливать весьма интересные взаимосвязи между областями, которые, казалось бы, совершенно не связаны друг с другом.

Например, в ходе работы над третьей редакцией нашей монографии по фундаментальной и клинической медицине витамина D<sub>3</sub> [3] мы вычисляли  $\beta$ -балл для всех публикаций по исследованиям применения витамина D<sub>3</sub> в клинической практике. Заметим, что профилактика дефицита витамина D<sub>3</sub> является экономически выгодным способом снижения нагрузки населения коморбидными хроническими патологиями. Помимо многочисленных экспертных оценок, приводимых нами в монографии, в крупномасштабном исследовании (n=7217) была прямо продемонстрирована корреляция между более низким уровнем 25(OH)D<sub>3</sub> в крови (<20 нг/мл) и более высокими затратами на содержание пациента в стационаре (P=0,001) [33].

Тем не менее разработанная нами система алгоритмов продолжает регистрировать вал «безрезультатных» исследований витамина D<sub>3</sub> ( $\beta<0$ ,  $\beta_4>10$ ), характеризующихся характерными «шаблонами». Это и термины жесткого отрицания клинических эффектов витамина D<sub>3</sub> («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 слова «все» («all»). Заметим, что использование данного слова должно быть весьма осторожным: ведь парадокс критиянида Эпименида известен уже более 2500 лет: «Все критяне – лжецы». Поэтому аномально частое употребление слова

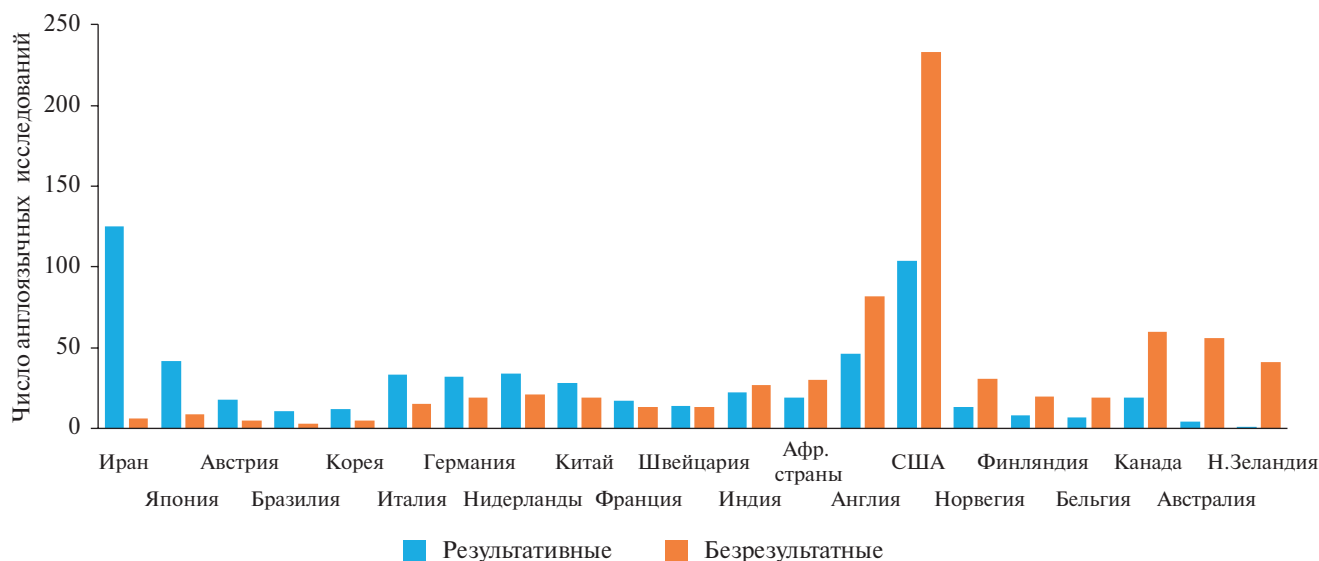


Рисунок 7. Результативность англоязычных клинических исследований витамина D<sub>3</sub> в различных странах.

Примечание. Страны упорядочены по убыванию отношения числа результативных и безрезультатных исследований.

Figure 7. The effectiveness of English-language clinical trials of vitamin D<sub>3</sub> in various countries.

Note. Countries are ordered by decreasing ratio of the number of productive and inconclusive studies.

«все» указывает на необоснованное стремление к обобщению (как правило, без каких-либо логических 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.

Анализ  $\beta$ -баллов этих «безрезультатных» исследований витамина  $D_3$  не только позволил выявить соответствующие рубрики PubMed/MEDLINE, но и установить ряд весьма поучительных курьезов. Например, в Австралии и в Новой Зеландии результативность исследований эффектов витамина  $D_3$  близка к нулю: только одно из 41 исследований, проведенных в Новой Зеландии, было результативно (в Австралии – только четыре из 56). Поскольку эти две страны являются абсолютными рекордсменами по низкой результативности исследований витамина  $D_3$  (рис. 7), то можно предположить, что на австралийском континенте законы физики, химии и биологии действуют по-другому. Однако, в соответствии с принципом Оккама, гораздо более вероятным объяснением столь выдающегося отсутствия результативности клинических исследований витамина  $D_3$  в Австралии и Новой Зеландии может играть чрезвычайно высокий уровень лоббистской деятельности молокозаводчиков этих двух стран (мировое лидерство по производству сухого молока). Мы обсудили эту возможную взаимосвязь ранее [34]. Кстати, крайне любопытно отметить, что языковые штампы «необходимы дальнейшие исследования» и «источник финансовой поддержки» являются характерными особенностями именно безрезультатных исследований по витамину  $D_3$ .

## ЗАКЛЮЧЕНИЕ / CONCLUSION

В работе сформулирована методика анализа «больших данных» биомедицинских публикаций, основанная на топологической теории сентимент-анализа медицинских текстов (то есть анализа эмоциональной модальности). Разработаны алгоритмы, позволяющие с 90%-й 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 ( $\beta$ -балл) и проведен анализ текстов абстрактов 19,9 млн публикаций из PubMed/MEDLINE. Показано, что публикации с низким сентимент-качеством ( $\beta$ -балл текста меньше 0, что соответствует преобладанию манипулятивных и негативных сентиментов) составляют всего 18,5% (3,7 из 19,9 млн). Рубрификация всего массива публикаций по 27840 рубрикам (MESH-система PubMed/MEDLINE) указала на повышение  $\beta$ -балла по годам (то есть на положительную динамику качества исследований) для 27090 исследованных рубрик. Наиболее интенсивная положительная динамика найдена для исследований по генетике, физиологии, фармакологии и геронтологии.

### Благодарности

Авторы выражают благодарность за консультирование по теме статьи акад. РАН В.Н. Серову, акад. РАН Г.Т. Сухих, акад. РАН Е.И. Гусеву, акад. РАН Ю.И. Журавлеву, проф. Н.К. Тетрашвили, проф. А.А. Баранову, проф. А.И. Федину, доц. к.м.н. А.К. Хаджидису.

Особая благодарность – руководителю проекта «Электронная библиотека» Российской государственной библиотеки, руководителю отдела УФКС РГБ к.м.н. Н.В. Авдеевой.

Также были выделены 249 рубрик с резко отрицательной динамикой качества и с выраженным нарастанием манипулятивных и негативных сентиментов. 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 т.п. Крайне интересно отметить, что абсолютным лидером по резкому ухудшению сентимент-качества публикаций являлись публикации из рубрик, связанных с повсеместно навязываемыми т.н. «стандартами доказательной медицины» (падение  $\beta$ -балла более чем на пять баллов/год за последние шесть лет – больше, чем во всех остальных 248 рубриках). Этому падению соответствует серия недавних скандалов вокруг публикационной активности соответствующего издания [32,34]. Интересным направлением дальнейших исследований является разработка алгоритмов «искусственного интеллекта» для оценки сентимент-качества статей по доказательной медицине.

На наш взгляд, использование систем алгоритмов, подобных разработанной системе, 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 г. В рамках этого закона установлено, что действующие в настоящее время протоколы лечения будут применяться до их пересмотра не позднее 31 декабря 2021 г. Предполагается, что клинические рекомендации будут разработаны или обновлены медицинскими профессиональными НКО по отдельным группам заболеваний [35].

Выборки статей из PubMed/MEDLINE, не содержащие лингвистических признаков манипуляции, весьма важны для разработки клинических рекомендаций [35,36], дальнейшего совершенствования деятельности агентств по оценке технологий здравоохранения (ОТЗ) разных стран [38–40], организации агентств по оценке технологий здравоохранения [37–39], повышения эффективности поиска медицинских разработок с инновационным потенциалом [40], в т.ч. инновационных лекарственных препаратов [41] и для реализации проектов по другим важным направлениям в сфере российского здравоохранения.

Таким образом, разработанный комплекс алгоритмов позволяет проводить эффективную оценку сентимент-качества текстов биомедицинских исследований, отфильтровывая потенциально псевдонаучные публикации, публикуемые под масками якобы «доказательных исследований», «международных рекомендаций», «рекомендуемых стандартов» и т.п. Идентификация манипулятивных, эмоционально негативных и прочих нежелательных сентиментов в публикациях PubMed/MEDLINE важна для: (1) предотвращения «накачки» медийной среды российских врачей негативными эмоциями; (2) остановки «продвижения» вышеупомянутых деструктивных тем; (3) идентификации реальных доказательных исследований и отделения их от штампованных «fake news», микрирующих под доказательность.

### Acknowledgements

The authors are grateful for the advice on the topic of acad. RAS V.N. Serov, Acad. RAS G.T. Dry, Acad. RAS E.I. Gusev, Acad. RAS Yu.I. Zhuravlev, prof. N.K. Tetraashvili, prof. A.A. Baranov, prof. A.I. Fedina, Assoc. Ph.D. A.K. Hajidis.

Special thanks to the head of the Electronic Library project of the Russian State Library, the head of the UFKS Department of the RSL, Ph.D. N.V. Avdeeva.

## ЛИТЕРАТУРА:

- Canese K., Weis S. PubMed: The Bibliographic Database. 2002 Oct 9 [Updated 2013 Mar 20]. In: The NCBI Handbook. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); [Электронный ресурс] URL: <https://www.ncbi.nlm.nih.gov/books/NBK153385/>. Дата обращения: 12.12.2019.
- Li L., Smith H.E., Atun R., Tudor Car L. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database Syst Rev.* 2019; MR000041. DOI: <https://dx.doi.org/10.1002/14651858.MR000041.pub2>.
- Громова О.А., Торшин И.Ю. Витамин D – смена парадигмы. М. 2017; 750 с.
- Громова О.А., Торшин И.Ю. Микронутриенты и репродуктивное здоровье. Руководство. М. 2019; 672 с.
- 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; 208 p.
- Watson D. Watson's Dictionary of Weasel Words, Contemporary Cliches, Cant and Management Jargon. Knopf, 1st Ed. 2004; 357 p.
- Torshin I.Y., Rudakov K.V. 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.
- Torshin I. Yu., Rudakov K.V. 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.
- Torshin I. Y. 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.
- Torshin I. Yu., Rudakov K.V. On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2015; 25 (4): 577–587.
- Torshin I.Y., Rudakov K.V. 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.
- Torshin I. Yu., Rudakov K.V. 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.
- Torshin I.Y., Rudakov K.V. 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.
- Torshin I.Y., Rudakov K.V. 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.
- Torshin I. Yu., Rudakov K.V. On the Procedures of Generation of Numerical Features Over Partitions of Sets of Objects in the Problem of Predicting Numerical Target Variables. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2019; 29 (4): 654–667. DOI: <https://dx.doi.org/10.1134/S1054661819040175>.
- Чернышев В.М. Меч Обоюдоострый. Конспект по Сектоведению. М. 2011; 138 с.
- Дворкин А.Л. Сектоведение: Тоталитарные секты. Опыт систематического исследования. 3-е изд., перераб. и доп. Н. Новгород. 2014; 816 с.
- Okter A. Mastermind: The Truth of the British Deep State. *Arashtirma Publishing.* 2017; 698 pp.
- Котеров А.Н. Критерии причинности в медико-биологических дисциплинах: история, сущность и радиационный аспект. Сообщение 1. постановка проблемы, понятие о причинах и причинности, ложные ассоциации. *Радиационная биология, радиозология.* 2019; 59 (1): 5–36. DOI: <https://dx.doi.org/10.1134/S0869803119010065>.
- Поппер К.Р. Предположения и опровержения: рост научного знания. М. 2004; 638 с.
- Громова О.А., Торшин И.Ю., Тетрашвили Н.К., Тапильская Н.И. Систематический анализ эффектов молибдена: здоровье беременной и плода. *Вопросы гинекологии, акушерства и перинатологии.* 2019; 18 (4): 83–94. DOI: <https://dx.doi.org/10.20953/1726-1678-2019-4-83-94>.
- Громова О.А., Торшин И.Ю., Тетрашвили Н.К., Галустьян А.Н., Курицына Н.А. О перспективах использования комбинаций фолиевой кислоты и активных фолатов для нутрициальной поддержки беременности. *Акушерство и гинекология.* 2019; 4: 87–94. DOI: <https://dx.doi.org/10.18565/aig.2019.4.87-94>.
- Торшин И.Ю., Лила А.М., Громова О.А., Наумов А.В., Громов А.Н. Об антикоагулянтных и антиагрегантных свойствах молекулы глюкозамина сульфата. *Современная ревматология.* 2019; 13 (3): 135–141. DOI: <https://dx.doi.org/10.14412/1996-7012-2019-3-135-141>.
- Громова О.А., Торшин И.Ю., Максимов В.А., Громов А.Н., Рудаков К.В. Систематический анализ исследований лактитола. *Экспериментальная и клиническая гастроэнтерология.* 2019; (2): 131–142. DOI: <https://doi.org/10.31146/1682-8658-ecg-162-2-131-142>.
- Арнольд В., Ильяшенко Ю., Аносов Д. и др. Динамические системы – 1. *Итоги науки и техн. Сер. Соврем. пробл. мат. Фундам. Направления.* М.: ВИНТИ. 260 с.
- Левенштейн В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов. *Доклады Академии Наук СССР.* 1965 4 163 (4): 845–848.
- Ioannidis J.P.A. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. *J Clin Epidemiol.* 2017 Apr; 84: 11–13. DOI: <https://dx.doi.org/10.1016/j.jclinepi.2017.02.001>.
- Ioannidis J.P. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol.* 2016 May; 73: 82–6. DOI: <https://dx.doi.org/10.1016/j.jclinepi.2016.02.012>. PMID: 26934549.
- Møller M.H., Ioannidis J.P.A., 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: <https://dx.doi.org/10.1007/s00134-017-5039-y>.
- Cochrane is a registered trademark in Australia, Canada, the European Community and the USA. 2017-09-19. [Электронный ресурс] URL: <http://trademarks.justia.com/791/85/cochrane-79185910.html>. Дата обращения: 12.12.2019.
- Торшин И.Ю., Громова О.А., Кобалава Ж.Д. О репрессиях  $\omega$ -3 полиненасыщенных жирных кислот адептами доказательной медицины. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология.* 2019; 12 (2): 91–114. DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.91-114>.
- Hannemann A., Wallaschofski H., Nauck M., Marschall P., Flessa S., Grabe H.J., Schmidt C.O., Baumeister S.E. Vitamin D and health care costs: Results from two independent population-based cohort studies. *Clin Nutr.* 2018 Dec; 37 (6 Pt A): 2149–2155. DOI: <https://dx.doi.org/10.1016/j.clnu.2017.10.014>.



34. 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] URL: <https://stimul.online/articles/science-and-technology/iskusstvennyu-intellekt-na-strazhe-zdorovya-beseda-vtoraya/>. Дата обращения: 12.12.2019.

35. Блинов Д. В., Акарачкова Е. С., Орлова А. С., Крюков Е. В., Корабельников Д. И. Новая концепция разработки клинических рекомендаций в России. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология*. 2019; 12 (2): 125–144. DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.125-144>.

36. 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 DOI: <https://doi.org/10.17749/2070-4909.2019.12.1.34-41>.

37. Хачатрян Г. Р., Омеляновский В. В., Мельникова Л. С., Ратушняк С. С. Международный опыт организации и финансового обеспечения агентств по оценке технологий здравоохранения. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология*. 2019. Т. 12. № 2. С. 146–154. DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.146-154>.

## REFERENCES:

1. Canese K., Weis S. PubMed: The Bibliographic Database. 2002 Oct 9 [Updated 2013 Mar 20]. In: The NCBI Handbook. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); [Electronic resource] URL: <https://www.ncbi.nlm.nih.gov/books/NBK153385/>. Accessed: 12.12.2019.

2. Li L., Smith H. E., Atun R., Tudor Car L. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database Syst Rev*. 2019; MR000041. DOI: <https://dx.doi.org/10.1002/14651858.MR000041.pub2>.

3. Gromova O. A., Torshin I. Yu. Vitamin D – a paradigm shift. Moscow. 2017; 750.

4. Gromova O. A., Torshin I. Yu. Micronutrients and reproductive health. Guide. Moscow. 2019; 672 с.

5. Stewart Chaplin. The Stained Glass Political Platform. The Century Magazine. USA. 1900.

6. Summers E. Weasel Words: 200 Words You Shouldn't Trust: 200 Words You Can't Trust. Chambers (Ed.), Slang & Idiom Dictionaries. 2009; 208 p.

7. Watson D. Watson's Dictionary of Weasel Words, Contemporary Cliches, Cant and Management Jargon. Knopf, 1st Ed. 2004; 357 p.

8. Torshin I. Y., Rudakov K. V. 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.

9. Torshin I. Yu., Rudakov K. V. 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.

10. Torshin I. Y. 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.

11. Torshin I. Yu., Rudakov K. V. On the theoretical basis of the metric analysis of poorly formalized problems of recognition and classification. *Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*. 2015; 25 (4): 577–587.

38. Лазарева М. Л., Тюрина И. В. Финансовая статистическая отчетность медицинских организаций: существующие недостатки и направления оптимизации. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология*. 2018; 11 (4): 61–66 DOI: <https://doi.org/10.17749/2070-4909.2018.11.4.061-066>.

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): 3–7 DOI: <https://doi.org/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 DOI: <https://doi.org/10.17749/2070-4909.2019.12.1.27-33>.

41. Мусина Н. З., Федяева В. К., Омеляновский В. В., Хачатрян Г. Р., Герасимова К. В., Лемешко В. А., Кончиц С. П. Обзор существующих зарубежных подходов к определению и оценке инновационности лекарственных препаратов. *ФАРМАКОЭКОНОМИКА. Современная Фармакоэкономика и Фармакоэпидемиология*. 2017; 10 (3): 66–74 DOI: <https://doi.org/10.17749/2070-4909.2017.10.3.066-074>.

12. Torshin I. Y., Rudakov K. V. 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.

13. Torshin I. Yu., Rudakov K. V. 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.

14. Torshin I. Y., Rudakov K. V. 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.

15. Torshin I. Y., Rudakov K. V. 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.

16. Torshin I. Yu., Rudakov K. V. On the Procedures of Generation of Numerical Features Over Partitions of Sets of Objects in the Problem of Predicting Numerical Target Variables. *Pattern Recognition and Image Analysis*. 2019; 29 (4): 654–667. DOI: <https://dx.doi.org/10.1134/S1054661819040175>.

17. Chernyshev V. M. Sword Double-edged. Abstract on Sectology. Moscow. 2011; 138 s. (in Russ)

18. Dvorkin A. L. Sectology: Totalitarian sects. Experience in systematic research. 3rd ed., Revised. and add. N. Novgorod. 2014; 816 s. (in Russ)

19. Okter A. Mastermind: The Truth of the British Deep State. *Arashtirma Publishing*. 2017; 698 pp.

20. Koterov A. N. Causal Criteria in Medical and Biological Disciplines: History, Essence and Radiation Aspect. Report 1. *Problem Statement, Conception of Causes and Causation, False Associations. Radiatsionnaya biologiya, radioekologiya* (in Russ). 2019; 59 (1): 5–36. DOI: <https://dx.doi.org/10.1134/S0869803119010065>.

21. Popper K. R. Assumptions and rebuttals: the growth of scientific knowledge. Moscow. 2004.

22. Gromova O. A., Torshin I. Yu., Tetrushvili N. K., Tapil'skaya N. I. A systematic analysis of the effects of molybdenum: the health of the

- pregnant woman and the fetus. *Voprosy ginekologii, akusherstva i perinatologii*. 2019; 18 (4): 83–94 (in Russ). DOI: <https://dx.doi.org/10.20953/1726-1678-2019-4-83-94>.
23. Gromova O. A., Torshin I. Yu., Tetrushvili N. K., Galustyan A. N., Kuritsyna N. A. On prospects for using combinations of folic acid and active folates for the nutritional support of pregnancy. *Akusherstvo i ginekologiya*. 2019; 4: 87–94 (in Russ). DOI: <https://dx.doi.org/10.18565/aig.2019.4.87-94>
24. Torshin I. Y., Lila A. M., Gromova O. A., Naumov A. V., Gromov A. N. On the anticoagulant and antiaggregatory properties of a glucosamine sulfate molecule. *Modern Rheumatology Journal*. 2019; 13 (3): 135–141. (In Russ.) DOI: <https://doi.org/10.14412/1996-7012-2019-3-135-141>.
25. Gromova O. A., Torshin I. Y., Maximov V. A., Gromov A. N., Rudakov K. V. Systematic analysis of lactitol studies. *Experimental and Clinical Gastroenterology*. 2019;(2):131-1 42. (In Russ.) DOI: <https://doi.org/10.31146/1682-8658-ecg-162-2-131-142>.
26. Arnold V., Ilyashenko Yu., Anosov D. et al. Dynamical systems – 1. Results of science and technology. *Ser. Lyng. prob. mat. Fundam. Directions*. Moscow. 260 s. (in Russ)
27. Levenshtein V. I. Binary codes with correction of loss, insertion and substitution of characters. *Doklady Akademii Nauk SSSR*. 1965; 163 (4): 845–848 (in Russ).
28. Ioannidis J. P. A. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. *J Clin Epidemiol*. 2017 Apr; 84: 11–13. DOI: <https://dx.doi.org/10.1016/j.jclinepi.2017.02.001>.
29. Ioannidis J. P. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. 2016 May; 73: 82–6. DOI: <https://dx.doi.org/10.1016/j.jclinepi.2016.02.012>.
30. Møller M. H., Ioannidis J. P. A., 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: <https://dx.doi.org/10.1007/s00134-017-5039-y>.
31. Cochrane is a registered trademark in Australia, Canada, the European Community and the USA. 2017-09-19. [Electronic resource] URL: [trademarks.justia.com/791/85/cochrane-79185910.html](http://trademarks.justia.com/791/85/cochrane-79185910.html). Accessed: 12.12.2019.
32. Torshin I. Y., Gromova O. A., Kobalava Z. D. Concerning the “repression” of  $\omega$ -3 polyunsaturated fatty acids by adepts of evidence-based medicine. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2019; 12 (2): 91–114 (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.91-114>.
33. Hannemann A., Wallaschofski H., Nauck M., Marschall P., Flessa S., Grabe H. J., Schmidt C. O., Baumeister S. E. Vitamin D and health care costs: Results from two independent population-based cohort studies. *Clin Nutr*. 2018 Dec; 37 (6 Pt A): 2149–2155. DOI: <https://dx.doi.org/10.1016/j.clnu.2017.10.014>.
34. Mekhanik A. G. Artificial intelligence on guard of health. The second conversation with O. A. Thundering and I. Yu. Torshin. *Stimul: Zhurnal ob innovatsiyakh v Rossii*. 30.10.2019. (in Russ) [Electronic resource] URL: <https://stimul.online/articles/science-and-technology/iskusstvennyy-intellekt-na-strazhe-zdorovya-beseda-vtoraya/>. Accessed: 12.12.2019.
35. Blinov D. V., Akarachkova E. S., Orlova A. S., Kryukov E. V., Korabelnikov D. I. New framework for the development of clinical guidelines in Russia. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2019; 12 (2): 125–144 (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.125-144>.
36. Zhuravleva N. I., Shubina L. C., Sukhorukikh O. A. The use of the level of evidence and grade of recommendations scales in developing clinical guidelines in the Russian Federation. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2019; 12 (1): 34–41 (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2019.12.1.34-41>.
37. Khachatryan G. R., Omelyanovskiy V. V., Melnikova L. S., Ratushnyak S. S. Organizational structure and funding of health technology assessment agencies around the world. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2019; 12 (2): 146–154 (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2019.12.2.146-154>
38. Lazareva M. L., Tyurina I. V. Financial statistical reporting by medical organizations: shortcomings and areas of optimization. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2018; 11 (4): 61–66. (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2018.11.4.061-066>.
39. Omelyanovsky V. V., Fedyayeva V. K., Musina N. Z. The concept of multi-criteria analysis of decision-making in the current system of health technology assessment in Russia. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2018; 11 (3): 3–7 (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2018.11.3-003-007>
40. Khrustalev M. B., Maksimova A. A. Effective search for potentially innovative scientific results in medicine. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2019; 12 (1): 27–33. (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2019.12.1.27-33>
41. Musina N. Z., Fedyayeva V. K., Omel'yanovskii V. V., Khachatryan G. R., Gerasimova K. V., Lemeshko V. A., Konchits K. P. Review of the current approaches to the assessment of the drug innovative potential worldwide. *FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2017; 10 (3): 66–74. (In Russ.) DOI: <https://doi.org/10.17749/2070-4909.2017.10.3.066-074>.

#### Сведения об авторах:

Торшин Иван Юрьевич – к. ф.-м. н., к. х. н., с. н. с., Институт фармакоинформатики, ФИЦ «Информатика и Управление» РАН. Scopus Author ID: 7003300274; Author ID: 54104; ORCID ID: <https://orcid.org/0000-0002-2659-7998>; WoS ResearcherID: C-7683-2018; РИНЦ SPIN-код: 1375-1114.

Громова Ольга Алексеевна – д. м. н., профессор, в. н. с., научный руководитель Института фармакоинформатики, ФИЦ «Информатика и Управление» РАН; в. н. с. Центра хранения и анализа больших данных, МГУ. Author ID: 94901; Scopus Author ID: 7003589812; ORCID ID: <https://orcid.org/0000-0002-7663-710X>; WoS ResearcherID: J-4946-2017. РИНЦ SPIN-код: 6317-9833. E-mail: [unesco.gromova@gmail.com](mailto:unesco.gromova@gmail.com).

Стаховская Людмила Витальевна – д. м. н., профессор, директор ФГБУ «Федеральный центр цереброваскулярной патологии и инсульта». ORCID ID: <https://orcid.org/0000-0001-6325-923>.

Ванчакова Нина Павловна – д. м. н., профессор, психиатр, Центр Психосоматической Медицины при клинической больнице № 122 им. Л. Г. Соколова ORCID ID: <https://orcid.org/0000-0003-1997-0202>.

Галустян Анна Николаевна – к. м. н., доцент, зав. кафедрой фармакологии с курсом клинической фармакологии и фармакоэкономики, ГБОУ ВПО СПб-ГПМУ Минздрава России. ORCID ID: <https://orcid.org/0000-0001-9679-632X>.

Кобалава Жанна Давидовна – д. м. н., профессор, заведующая кафедрой внутренних болезней с курсом кардиологии и функциональной диагностики МИ ФГАОУ РУДН; заведующая кафедрой внутренних болезней, кардиологии и клинической фармакологии ФПК МР МИ ФГАОУ РУДН, МИ ФГАОУ РУДН.

Гришина Татьяна Романовна – д. м. н., проф., зав. кафедрой фармакологии ФГБОУ ВО ИвГМА Минздрава России. ORCID ID: <https://orcid.org/0000-0002-1665-1188>. Author ID: 113019.

Громов Андрей Николаевич – инженер-исследователь ФИЦ ИУ РАН. AuthorID: 15082; Scopus Author ID: 7102053964; ORCID ID: <https://orcid.org/0000-0001-7507-191X>; WoS ResearcherID: C-7476-2018; РИНЦ SPIN-код: 8034-7910.

Иловайская Ирэна Адольфовна – врач-эндокринолог высшей категории, д. м. н., доцент, старший научный сотрудник отделения терапевтической эндокринологии, ГБУЗ МО МОНИКИ им. М. Ф. Владимирского. ResearcherID: I-1159-2014; ORCID ID: <https://orcid.org/0000-0003-3261-7366>. Scopus Author ID: 6506067338.

*Коденцова Вера Митрофановна* – д.б.н., профессор, г.н.с., лаборатория витаминов и микроэлементов, ФГБУН «ФИЦ питания и биотехнологии».  
*Калачева Алла Геннадьевна* – к.м.н., доцент кафедры фармакологии ФГБОУ ВО ИвГМА Минздрава России. ORCID ID: <https://orcid.org/0000-0001-6144-5781>.  
*Лиманова Ольга Адольфовна* – к.м.н., доцент кафедры фармакологии ФГБОУ ВО ИвГМА Минздрава РФ. ORCID ID: <https://orcid.org/0000-0002-2342-3036>.  
*Максимов Валерий Алексеевич* – д.м.н., профессор кафедры диетологии и нутрициологии, гастроэнтеролог, РМАПО. ORCID ID: <https://orcid.org/0000-0003-4120-1071>.

*Малявская Светлана Ивановна* – д.м.н., профессор, проректор по научной работе ФГБОУ ВО СГМУ (г. Архангельск) Минздрава России. ORCID ID: <http://orcid.org/0000-0003-2521-0824>; eLIBRARY ID: 6257-4400.

*Мозговая Елена Витальевна* – д.м.н., доцент, руководитель акушерского отдела с перинатологией ФГБНУ «Научно-исследовательский институт акушерства, гинекологии и репродуктологии им. Д. О. Отта» РАН; профессор кафедры акушерства, гинекологии и репродуктологии медицинского факультета СПбГУ. WoS ResearcherID: L-1432-2017; Author ID Scopus: 24822403200; Author ID: 386830.

*Тапильская Наталья Игоревна* – д.м.н., профессор, ведущий научный сотрудник отделения вспомогательных репродуктивных технологий, ФГБНУ «Научно-исследовательский институт акушерства, гинекологии и репродуктологии им. Д. О. Отта» РАН. Author ID Scopus: 23013489000; WoS ResearcherID: A-7504-2016; ORCID ID: <https://orcid.org/0000-0001-5309-0087>; ID map of science: 00052162; РИНЦ SPIN-код: 3605-0413.

*Рудаков Константин Владимирович* – акад. РАН, зам. директора Федерального Исследовательского Центра Информатика и Управление РАН, зав. кафедрой интеллектуальные системы МФТИ. Scopus Author ID: 6603540895.

*Семенов Владимир Александрович* – д.м.н., профессор, ФГБОУ ВО «Кемеровский Государственный Медицинский Университет» Минздрава России.

#### About the authors:

*Ivan Yu. Torshin* – MD, PhD, Senior Researcher, Federal Research Center “Informatics and Management”, Russian Academy of Sciences; Big Data Storage and Analysis Center, Moscow State University. Scopus Author ID: 7003300274; Author ID: 54104; ORCID ID: <https://orcid.org/0000-0002-2659-7998>; WoS ResearcherID: C-7683-2018; RSCI SPIN-code: 1375-1114.

*Olga A. Gromova* – MD, Dr Sci Med, Professor, Senior Researcher, Scientific Director of the Federal Research Center “Informatics and Management”, Russian Academy of Sciences; Leading Researcher, Center for Big Data Analysis, Moscow State University; Author ID: 94901; Scopus Author ID: 7003589812; ORCID ID: <https://orcid.org/0000-0002-7663-710X>; WoS ResearcherID: J-4946-2017. RSCI SPIN-code: 6317-9833. E-mail: [unesco.gromova@gmail.com](mailto:unesco.gromova@gmail.com).

*Lyudmila V. Stakhovskaia* – MD, Dr Sci Med, Professor, director, Federal Center for Cerebrovascular Pathology and Stroke. ORCID ID: <https://orcid.org/0000-0001-6325-923>.

*Nina P. Vanchakova* – MD, Dr Sci Med (medical psychology and psychiatry), Professor, Psychiatrist, Center for Psychosomatic Medicine at the Clinical Hospital No. 122 named after L. G. Sokolov. ORCID ID: <https://orcid.org/0000-0003-1997-0202>.

*Anna N. Galustyan* – MD, PhD, Associate Professor, Head of the Department of Pharmacology, Saint-Petersburg National Pediatric Medical University. ORCID ID: <https://orcid.org/0000-0001-9679-632X>.

*Zhanna D. Kobalava* – MD, PhD, Professor, Head of the Department of Internal Medicine, Cardiology and Clinical Pharmacology, Peoples' Friendship University.

*Tatiana R. Grishina* – MD, Dr Sci Med, Head of the Department of pharmacology, Ivanovo State Medical Academy. ORCID ID: <https://orcid.org/0000-0002-1665-1188>. Author ID: 113019.

*Andrey N. Gromov* – research engineer, Federal Research Center “Informatics and Management”, Russian Academy of Sciences. AuthorID: 15082; Scopus Author ID: 7102053964; ORCID ID: <https://orcid.org/0000-0001-7507-191X>; WoS ResearcherID: C-7476-2018; RSCI SPIN-code: 8034-7910.

*Irena A. Illovayskaya* – MD, Dr Sci Med, endocrinologist of the highest category, associate professor, senior researcher at the Department of Therapeutic Endocrinology, Moscow Regional Research and Clinical Institute (MONIKI). ResearcherID: I-1159-2014; ORCID ID: <https://orcid.org/0000-0003-3261-7366>. Scopus Author ID: 6506067338.

*Vera M. Kodentsova* – Dr Sci Biol, Professor, Senior Researcher, Laboratory of Vitamins and Microelements, Federal Research Center for Nutrition and Biotechnology.

*Alla G. Kalacheva* – MD, PhD, Associate Professor of the Department of Pharmacology and Clinical Pharmacology, Ivanovo State Medical Academy. ORCID ID: <https://orcid.org/0000-0001-6144-5781>.

*Olga A. Limanova* – MD, PhD, Associate Professor of the Department of Pharmacology, Ivanovo State Medical Academy. ORCID ID: <https://orcid.org/0000-0002-2342-3036>.

*Valery A. Maksimov* – MD, Dr Sci Med, gastroenterologist, Professor of the Department of Dietetics and Nutritionology, Russian Medical Academy of Continuing Professional Education. ORCID ID: <https://orcid.org/0000-0003-4120-1071>.

*Svetlana I. Malyvskaya* – MD, Dr Sci Med, Professor, Head of Sciences, Northern State Medical University. ORCID ID: <http://orcid.org/0000-0003-2521-0824>; eLIBRARY ID: 6257-4400.

*Elena V. Mozgovaya* – MD, Dr Sci Med, Associate Professor, Head of the Obstetric Department with Perinatology, The Research Institute of Obstetrics, Gynecology and Reproductology named after D. O. Ott; Professor, Department of Obstetrics, Gynecology and Reproductology, Faculty of Medicine, Saint-Petersburg State Pediatric Medical University. WoS ResearcherID: L-1432-2017; Author ID Scopus: 24822403200; Author ID: 386830.

*Natalia I. Tapiiskaya* – MD, Dr Sci Med, Professor, Leading Researcher of the Department of Assisted Reproductive Technologies, The Research Institute of Obstetrics, Gynecology and Reproductology named after D. O. Ott. Author ID Scopus: 23013489000; WoS ResearcherID: A-7504-2016; ORCID ID: <https://orcid.org/0000-0001-5309-0087>; ID map of science: 00052162; RSCI SPIN-code: 3605-0413.

*Konstantin V. Rudakov* – Academician of the Russian Academy of Sciences, Scientific Director of the Federal Research Center for Informatics and Management RAS, head of Department of Intellectual Systems MIPT. Scopus Author ID: 6603540895.

*Vladimir A. Semenov* – MD, Dr Sci Med, Professor, Kemerovo State Medical University.