

ГУСЕВ А.В.,

к.т.н., ООО «К-Скай», Петрозаводск, Россия; ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, Москва, Россия,
e-mail: agusev@webiomed.ai

ГОЛЬДИНА Т.А.,

АО «Санофи Россия», Москва, Россия

АНАЛИЗ ДАННЫХ РЕАЛЬНОЙ КЛИНИЧЕСКОЙ ПРАКТИКИ, ИЗВЛЕЧЕННЫХ ИЗ ЭЛЕКТРОННЫХ МЕДИЦИНСКИХ КАРТ В ПЛАТФОРМЕ WEBIOMED

DOI: 10.25881/18110193_2024_3_44

Аннотация. Реализованные в 2011–2021 гг. государственные программы в сфере информатизации здравоохранения привели к тому, что свыше 91% государственных и муниципальных медицинских организаций внедрили различные медицинские информационные системы. Это позволило начать переход на ведение электронных медицинских карт (ЭМК). Извлечение данных реальной клинической практики (ДРКП) из накопленных ЭМК и последующий анализ этих данных открывает новые и перспективные возможности для развития отечественного здравоохранения. Целью работы стал анализ ДРКП, извлеченных из обезличенных ЭМК.

Материалами работы стала база данных платформы прогностной аналитики Webiomed, в которой на момент исследования были накоплены обезличенные ЭМК свыше 29 млн пациентов, включая 229 млн различных медицинских документов. Поставщиками данных для платформы стали 856 медицинских организаций из 28 субъектов РФ. Функциональные возможности платформы Webiomed позволяют обрабатывать неструктурированные медицинские документы, извлекать с помощью технологий искусственного интеллекта из них данные, пригодные для анализа.

Результаты. В настоящей работе представлены: анализ медицинских организаций как поставщиков данных для платформы Webiomed, анализ структуры и состава ЭМК, анализ популяции пациентов. Анализ проведен на момент выгрузки данных 16.10.2023. Из накопленных обезличенных ЭМК удалось извлечь и систематизировать по разным видам более 4 млрд 558 млн структурированных признаков. Платформа содержит 147 886 190 случаев заболеваний, классифицированных по справочнику МКБ-10. 8 393 403 пациента (28,71% от общего числа) имеют в ЭМК медицинскую информацию. Доля «пустых» ЭМК (которые не содержали ни одной медицинской записи) составила 71,29%. Однако ЭМК 3 448 797 пациентов имеют более чем 10 медицинских документов. В структуре ЭМК преобладают протоколы врачебных осмотров, протоколы лабораторных исследований, электронные рецепты, инструментальные исследования. 4 456 263 пациентов имеют глубину сбора данных свыше 3х лет.

Заключение. Полученные результаты свидетельствуют, что извлечение и обработка ДРКП из обезличенных ЭМК действительно позволяют создавать большие наборы структурированных данных. В настоящее время, по нашим сведениям, платформа Webiomed содержит самую крупную в России базу ДРКП, извлеченных из ЭМК. Представленный в данной статье материал – это первый анализ ЭМК и извлеченных из них признаков, осуществленный и опубликованный в России. Обеспечение качества работы с ДРКП на всех этапах, начиная от разработки структуры ЭМК и ввода данных до формирования цифровых двойников, является важнейшим условием их применения для решения различных задач в системе здравоохранения и фармацевтической индустрии.

Ключевые слова: электронная медицинская карта; данные реальной клинической практики; платформа Webiomed; искусственный интеллект.

Для цитирования: Гусев А.В., Гольдина Т.А. Анализ данных реальной клинической практики, извлеченных из электронных медицинских карт в платформе Webiomed. Врач и информационные технологии. 2024; 3: 44-61. doi: 10.25881/18110193_2024_3_44.

GUSEV A.V.,

PhD, K-SkAI LLC, Petrozavodsk, Russia; Federal Research Institute
for Health Organization and Informatics, Moscow, Russia,
e-mail: agusev@webiomed.ai

GOLDINA T.A.,

Representative office of the joint-stock company «Sanofi Russia», Moscow, Russia

ANALYSIS OF REAL-WORLD DATA EXTRACTED FROM ELECTRONIC MEDICAL RECORDS IN THE WEBIOMED PLATFORM

DOI: 10.25881/18110193_2024_3_44

Abstract. State programs in the field of healthcare informatization implemented in 2011–2021 resulted in over 91% of state and municipal medical organizations implementing various medical information systems. This made it possible to start the transition to electronic medical records (EMR). The extraction of real-world data (RWD) from the accumulated EMRs and subsequent analysis of these data opens new and promising opportunities for the development of domestic healthcare.

Aim: to analyze the RWD extracted from anonymized EMRs.

Materials and methods: we used Webiomed's predictive analytics platform database, which at the time of the study had accumulated anonymized EMRs of over 29 million patients, including 229 million different medical documents. Data providers for the platform were 856 medical organizations from 28 regions of the Russian Federation. The functionality of the Webiomed platform allows processing unstructured medical documents, extracting data from them suitable for analysis using artificial intelligence technologies.

Results. This paper presents: analysis of medical organizations as data providers for Webiomed platform, analysis of EMC structure and composition, analysis of patient population. The analysis was performed at the moment of data upload on 16.10.2023. More than 4 billion 558 million structured attributes were extracted from the accumulated anonymized EMRs and systematized by different types. The platform contains 147,886,190 cases of diseases classified according to the ICD-10. 8,393,403 patients (28.71% of the total) have medical information in the EMR. The proportion of “empty” EMCs (which did not contain any medical record) was 71.29%. However, EMRs of 3,448,797 patients had more than 10 medical records. The structure of EMRs is dominated by protocols of medical examinations, protocols of laboratory tests, electronic prescriptions, and instrumental examinations. 4,456,263 patients have a depth of data collection of more than 3 years.

Conclusion. The results show that the extraction and processing of RWD from anonymized EMRs do allow the creation of large sets of structured data. Currently, to the best of our knowledge, the Webiomed platform contains the largest database of RWD extracted from EMRs in Russia. The material presented in this paper is the first analysis of EMRs and extracted features performed and published in Russia. Ensuring the quality of work with RWD at all stages, from the development of EMR structure and data input to the formation of digital twins, is the most important condition for their application to solve various tasks in the healthcare system and pharmaceutical industry.

Keywords: Electronic Medical Record; Real World Data; Webiomed platform; Artificial Intelligence.

For citation: Gusev A.V., Goldina T.A. Analysis of Real-World Data extracted from electronic medical records in the Webiomed platform. Medical doctor and information technology. 2024; 3: 44-61. doi: 10.25881/18110193_2024_3_44.

ВВЕДЕНИЕ

Развитие цифрового здравоохранения по всему миру, в том числе в России, позволило повсеместно внедрить в медицинских организациях (МО) различные медицинские информационные системы (МИС), основная задача которых состоит в ведении медицинскими работниками электронных медицинских карт (ЭМК). Реализованные в России 2011–2021 гг. государственные программы в сфере информатизации здравоохранения привели к тому, что свыше 91% государственных МО накапливают в своих базах данных ЭМК, причем объем собираемой информации все время увеличивается [1, 2].

ЭМК представляют собой один из основных источников данных реальной клинической практики (ДРКП, Real World Data, RWD). Анализ ДРКП позволяет извлекать ценные и всесторонние знания из имеющегося практического опыта, в том числе получать сведения о реальной распространенности заболеваний и факторов риска, об эффективности лекарственных средств и т. д. [3–5].

ЭМК – довольно сложный источник данных в силу различных объективных причин и проблем, таких как преимущественно неструктурированный способ хранения записей, недостаточный контроль ввода данных в интерфейсах ЭМК, децентрализованная архитектура, отсутствие поддержки единых стандартов кодирования медицинских записей и т. д. [4]. Yamamoto K. et al. (2012) пришли к выводу о необходимости методичной и сложной работы с неструктурированными медицинскими записями с точки зрения обеспечения качества получаемых из них данных [6].

В этой связи для извлечения из ЭМК ДРКП с приемлемым для дальнейшего анализа и исследований качеством необходимо предпринимать ряд сложных технических и организационных мер, в частности – обеспечивать централизованный сбор записей из распределенных ЭМК в различных МО, включая поликлиники и стационары [7, 8].

Активное развитие технологий искусственного интеллекта (ИИ), включая машинное обучение, позволяют автоматизировать обработку ЭМК, получая из них пригодные для анализа данные [9]. Zhao Y. et al (2021) использовали извлечение ДРКП из неструктурированных ЭМК с

помощью технологий обработки естественного языка (Natural Language Processing, NLP), показав что такой подход с одной стороны позволяет автоматизировать формирование наборов данных даже из неструктурированных записей, но в тоже время приводит к ряду проблем при их обработке, включая наличие пропусков и несоответствия [10]. Это влечет за собой необходимость в очистке и ручном пересмотре данных. Laar et al. (2020) предложен алгоритм по сбору результатов лечения наркозависимости также на основе интеллектуального изучения ЭМК и сбору ДРКП, где подчеркивается важность последовательного и многоэтапного анализа и сбора данных [11].

Целью работы стал анализ ДРКП, извлеченных из обезличенных ЭМК, накопленных в платформе Webiomed.

МАТЕРИАЛЫ И МЕТОДЫ

Сбор данных осуществлялся компанией К-Скай в разрабатываемой платформе прогнозной аналитики Webiomed (<https://webiomed.ru/>), обладающей следующими нормативно-правовыми статусами:

- включена в реестр российского программного обеспечения 15 ноября 2018 г. (реестровая запись №4980 от 03.12.2018);
- является первой российской разработкой в области ИИ для здравоохранения, зарегистрированной Росздравнадзором как медицинское изделие (ПУ № РЗН 2020/9958 от 03.04.2024);
- является на момент публикации единственной ИИ-системой, зарегистрированной в качестве «Иной ИС» по постановлению правительства №447.

Кроме того, компания К-Скай обладает международным сертификатом соответствия Системы Менеджмента Качества (СМК) стандарту ISO 13485:2016 «Изделия медицинские. Системы менеджмента качества. Требования для целей регулирования».

Регистрация в качестве «Иная ИС» дает право осуществлять автоматический информационный обмен как с Единой Государственной Информационной Системой в сфере Здравоохранения (ЕГИСЗ), так и с государственными информационными системами в сфере здравоохранения субъектов РФ. Для сбора данных на

платформе были подписаны соглашения с соответствующими операторами персональных медицинских данных, имеющими добровольные информированные согласия пациентов на сбор и обработку их персональных медицинских данных. Соглашения включали предоставление К-Скай права проводить научно-исследовательские работы на основе собранной информации.

Материалом настоящего исследования стали обезличенные ЭМК 29 230 375 пациентов, накопленные в платформе Webiomed в период с 01.05.2020 по 16.10.2023. Поставщиками данных для платформы стали 856 МО из 28 субъектов Российской Федерации (8 федеральных округов). В платформе было накоплено 127 207 369 структурированных электронных медицинских документов (СЭМД).

Сбор и обработка данных с использованием платформы Webiomed проходит в 2 этапа (рис. 1).

На первом этапе в условиях защищенного информационного контура операторов персональных данных осуществляется автоматическое обезличивание записей, накопленных в ЭМК пациентов. Обезличивание может выполняться как на стороне МИС, так и с помощью отдельного интеграционного приложения Webiomed.Connect [12].

Далее из обезличенных записей формируются СЭМД, имеющие единую структуру и формат хранения данных на основе стандарта архитектуры клинических документов «Health Level Seven Clinical Document Architecture. Release 2.0» (HL7 CDA R2.0). Несмотря на использование в термине слова «структурированный», большая часть клинической информации в СЭМД представлена в неструктурированном виде – простыми текстовыми полями (жалобы, данные объективного осмотра, описания и заключения диагностических исследований и тд).

На втором этапе сформированные СЭМД в автоматическом режиме по защищенным каналам связи передаются в платформу Webiomed.

Для извлечения из полученных СЭМД признаков, пригодных для автоматического анализа и интерпретации, используется технология natural language processing (NLP) [4]. После извлечения признаков осуществляется их автоматическая очистка и подготовка для последующего использования: выявляются дублированные значения, данные объединяются в единые эпизоды оказания медицинской помощи (законченные случаи), с помощью правил форматно-логического контроля выявляются и исключаются ошибочные записи, аномалии и т.д. [4].

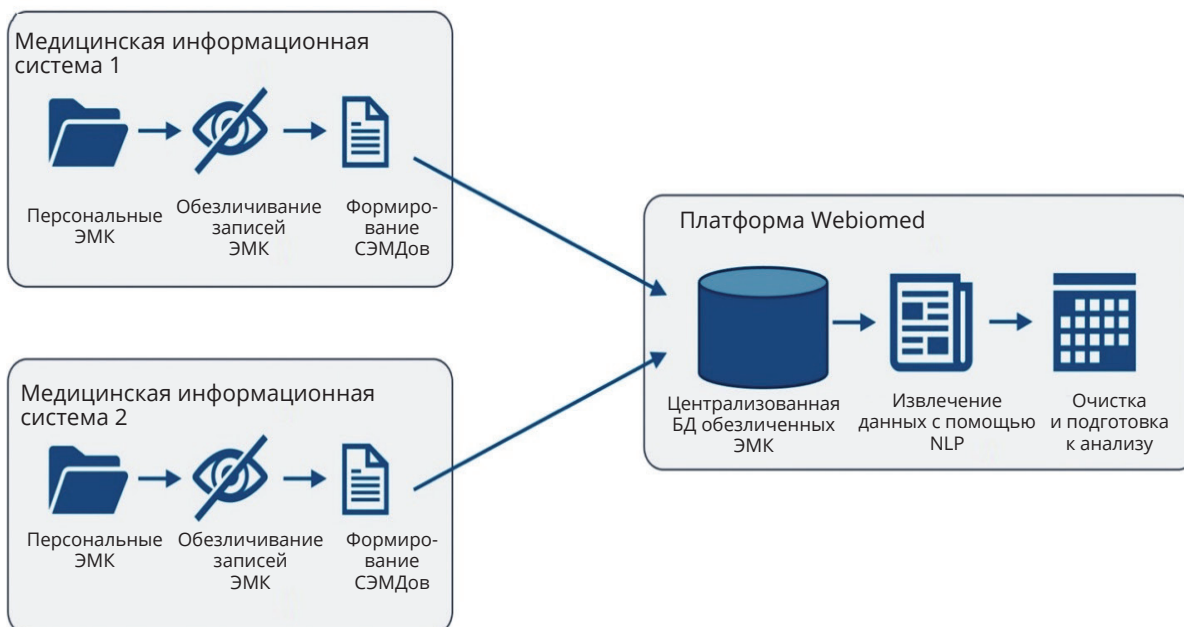


Рисунок 1 — Последовательность сбора и извлечения ДРКП из обезличенных ЭМК при использовании платформы Webiomed.

В результате извлечения, проверки и обработки признаков осуществляется формирование единого цифрового профиля пациента, который включает структурированные данные по всем случаям обращения за медицинской помощью, единый лист окончательных диагнозов (ЛОД), значения извлеченных признаков, включая симптомы из жалоб пациента, объективные данные врачебных осмотров, результаты обследования, данные о назначенных лекарственных препаратах, результаты диспансеризации, данные о диспансерном наблюдении и т.д.

Сформированные таким образом цифровые профили пациентов хранятся в нормированной базе данных, поддерживающей язык SQL. Такая методология позволяет формировать различные запросы к полученным данным и создавать структурированные наборы ДРКП по определенным критериям. Подготовленные цифровые профили используются для дальнейшего анализа, в том числе в научных исследованиях, машинном обучении, скрининге пациентов и др. [4].

РЕЗУЛЬТАТЫ

В России государственное финансирование проектов информатизации здравоохранения началось в 2011 г. К 2014 г. был зафиксирован заметный прирост создания ЭМК в медицинских информационных системах, причем на 2014–2019 гг. пришлась самая активная фаза внедрения МИС в МО. Начиная с 2019 г. динамика добавления новых ЭМК в МИС вышла на относительно стабильный уровень. Кроме того, в 2017–2018 гг. определена и утверждена

действующая на данный момент система нормативно-правового и технического регулирования [2].

Таким образом, развитие информационных технологий в целом и развитие информатизации здравоохранения способствовали активному накоплению ДРКП в ЭМК и, как следствие, накоплению данных в платформе Webiomed. Прирост данных, безусловно, происходит каждый месяц. Именно поэтому результаты исследования, представленного в настоящей статье, по сути, представляют собой только срез в определенный момент времени развития платформы Webiomed.

АНАЛИЗ МЕДИЦИНСКИХ ОРГАНИЗАЦИЙ

Распределение МО по субъектам РФ на момент проведения выгрузки данных (16.10.2023) из платформы Webiomed представлено в таблице 1.

Анализ видов МО, данные которых представлены в платформе Webiomed, был проведен в соответствии с Единой номенклатурой государственных и муниципальных учреждений здравоохранения, утвержденной Приказом № 627 Минздравсоцразвития РФ от 07.10.2005 (таблица 2).

Подключение МО и, соответственно, накопление данных пациентов происходило постепенно. Так, в 2019 г. на платформе Webiomed накопились данные 311 006 пациентов, в 2020 г. добавились данные 1 975 228 пациентов, в 2021 и 2022 гг. – 833 828 и 13 501 959 пациентов, соответственно. За первые 3 квартала 2023 г. включено на платформу 12 607 242 пациентов. Включение МО и накопление данных продолжает происходить и после проведения данного анализа.

Таблица 1 — Поставщики данных для платформы Webiomed

Округ	Число МО	Число пациентов	Число документов
Центральный	103	2 691 050	5 880 126
Северо-Западный	22	447 482	15 031 703
Южный	58	1 185 868	2 721 265
Северо-Кавказский	30	100 491	565 205
Приволжский	356	13 393 658	150 151 102
Уральский	89	2 429 113	28 615 465
Сибирский	83	5 132 781	17 170 677
Дальневосточный	53	1 219 822	7 529 288
Крымский	62	2 630 110	1 352 315
Итого	856	29 230 375	229 017 146

Таблица 2 — Число пациентов и число документов, данные которых представлены в платформе Webiomed

Вид МО	Число МО	Число пациентов	Число документов
Больничные учреждения, включая центральные районные больницы, городские стационары, госпитали и т.д.	576	26 492 447	197 261 591
Диспансеры всех видов	71	45 718	375 406
Амбулаторно-поликлинические учреждения, включая стоматологические поликлиники, детские поликлиники и т.д.	89	2 039 121	13 462 348
Медицинские центры	82	450 440	17 157 060
Учреждения скорой медицинской помощи	21	164 678	691
Учреждения охраны материнства и детства, включая родильные дома, женские консультации, перинатальные центры и т.д.	14	11 230	81 108
Санаторно-курортные учреждения	3	26 741	678 942
Итого	856	29 230 375	229 017 146

Важно подчеркнуть, что данные пациентов, добавленных на платформу в период с 2019 г. и до настоящего времени, регулярно обновляются. Это обеспечивает необходимую глубину данных для различных задач анализа. Так, например, если данные пациента впервые попали на платформу в 2020 г., то в последующие годы данные этого пациента обновляются по мере обращаемости пациента за медицинской помощью.

АНАЛИЗ ПОПУЛЯЦИИ

Выгруженный из платформы Webiomed набор ДРКП включает 229 017 146 документов о 29 230 375 пациентах, из них мужчины составили 47,16% (13 786 080 человек),

женщины – 52,84% (15 444 295 человек). Данные о возрастном составе представлены на рисунке 2.

Указано число полных лет на момент анализа данных платформы, что означает, например, если на момент исследования пациенту было 6 лет 11 месяцев, то он попадет в группу «0–6 лет», если 30 лет и 1 месяц – в группу «30–39 лет».

Информация по пациентам с определенными заболеваниями, имеющимися в базе, является одной из важнейших характеристик, т.к. часто проекты/исследования на основе баз данных реализуются в рамках конкретных заболеваний или терапевтических областей. Распределение пациентов по заболеваниям представлено в таблице 3.

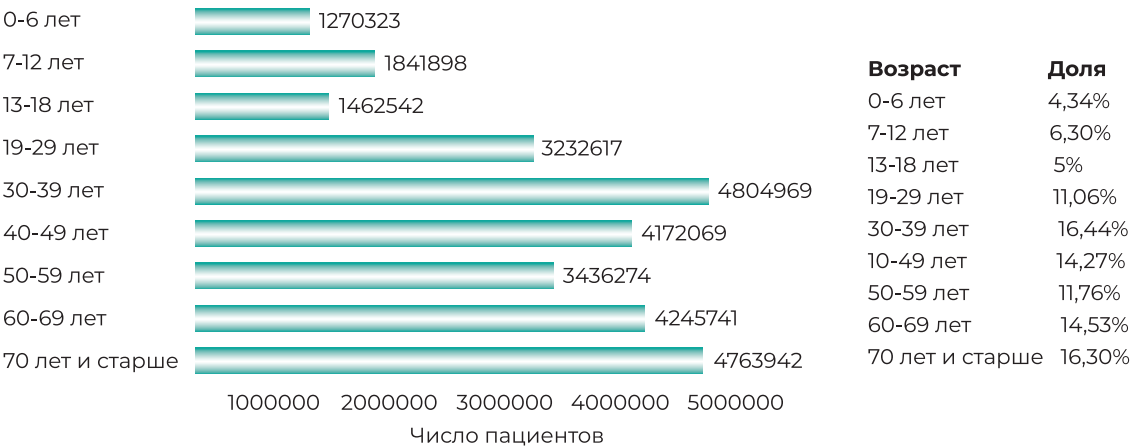


Рисунок 2 — Распределение пациентов по возрастному составу. Указано число полных лет на момент анализа данных платформы, что означает, например, если на момент исследования пациенту было 6 лет 11 месяцев, то он попадет в группу «0–6 лет», если 30 лет и 1 месяц – в группу «30–39 лет».

Таблица 3 — Классы заболеваний по МКБ-10, представленные в платформе Webiomed

Класс МКБ-10	Кол-во случаев
I Некоторые инфекционные и паразитарные болезни (A00-B99)	3 376 071
II Новообразования (C00-D48)	2 317 938
III Болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм (D50-D89)	671 440
IV Болезни эндокринной системы, расстройства питания и нарушения обмена веществ (E00-E90)	3 092 105
V Психические расстройства и расстройства поведения (F00-F99)	1 441 710
VI Болезни нервной системы (G00-G99)	6 216 542
VII Болезни глаза и его придаточного аппарата (H00-H59)	5 316 371
VIII Болезни уха и сосцевидного отростка (H60-H95)	2 505 292
IX Болезни системы кровообращения (I00-I99)	17 203 039
X Болезни органов дыхания (J00-J99)	13 839 440
XI Болезни органов пищеварения (K00-K93)	13 806 006
XII Болезни кожи и подкожной клетчатки (L00-L99)	3 951 349
XIII Болезни костно-мышечной системы и соединительной ткани (M00-M99)	8 158 773
XIV Болезни мочеполовой системы (N00-N99)	11 050 764
XV Беременность, роды и послеродовой период (O00-O99)	2 739 352
XVI Отдельные состояния, возникающие в перинатальном периоде (P00-P96)	282 072
XVII Врожденные аномалии [пороки развития], деформации и хромосомные нарушения (Q00-Q99)	335 531
XVIII Симптомы, признаки и отклонения от нормы, выявленные при клинических и лабораторных исследованиях, не классифицированные в других рубриках (R00-R99)	1 503 912
XIX Травмы, отравления и некоторые другие последствия воздействия внешних причин (S00-T98)	4 986 460
XX Внешние причины заболеваемости и смертности (V01-Y98)	35 302
XXI Факторы, влияющие на состояние здоровья населения и обращения в учреждения здравоохранения (Z00-Z99)	43 038 359
XXII Коды для особых целей (U00-U85)	2 018 362
Итого	147 886 190

Для того, чтобы понять, каким образом формируется этот список, необходимо обратиться к структуре ЭМК. Вне зависимости от того, какая МИС используется в МО или какой способ хранения информации в ЭМК реализовал разработчик, цифровой профиль пациента в платформе Webiomed всегда будет содержать ЛОД. Данный раздел цифрового профиля собирает информацию о зарегистрированных окончательных диагнозах из всех первичных медицинских документов, из законченных случаев лечения (амбулаторных, стационарных, карт диспансеризации/профосмотра) и из врачебных осмотров и представляет собой историю заболеваний, когда-либо выявленных у пациента.

Если у пациента одно и тоже хроническое заболевание было закодировано несколько раз (например, в начале в поликлинике, потом в стационаре, а после пациент попал на реабилитацию в санаторий), то в платформе Webiomed в ЛОД будет обеспечена только 1 запись о таком диагнозе. Кроме того, если в рамках одной нозологии менялись коды хронических диагнозов после точки (например, было I10.1, стало I10.4 или даже I10), то в платформе Webiomed это также кодируется как один диагноз (в случае приведенного примера – диагноз I10). Однако острые заболевания не объединяются в 1 запись в ЛОД: они будут зафиксированы столько раз, сколько было случаев обращения и лечения.

Например, если пациент болел COVID-19 в 2020 г. и в 2023 г. – то это будет 2 отдельных записи.

Таким образом, у 1 пациента в ЛОДе могут быть десятки записей об острых и хронических заболеваниях. Именно этим и обуславливается пятикратная разница между числом пациентов и числом диагнозов в платформе Webiomed.

АНАЛИЗ СТРУКТУРЫ И СОСТАВА ЭМК

Из 29 230 375 пациентов в платформе Webiomed 8 393 403 имеют медицинскую информацию, что составляет 28,71% от общего числа пациентов. Доля «пустых» ЭМК (которые не содержали ни одной клинически значимой записи) составила 71,29%. Это так называемые «пустые» регистрации.

К ЭМК, содержащим минимальное наполнение, были отнесены карты, в которых было расположено до 10 любых медицинских записей. Доля таких ЭМК составила 16,92%. К ЭМК, классифицированным нами как имеющие достаточное наполнение, мы отнесли карты с более чем 10 медицинскими документами, включая информацию о законченных случаях обращения в МО. Доля таких карт составила 11,8%, то есть на платформе Webiomed есть более чем 10 медицинских документов у 3 448 797 пациентов. Более детальная информация представлена в таблице 4.

Число амбулаторных законченных случаев составляет 79 847 397, историй болезней – 3 084 097, карт диспансеризации / профосмотра – 2 209 398, карт диспансерного наблюдения – 149 218.

Первые ЭМК, которые были зарегистрированы в МО, предоставивших материалы для платформы Webiomed, датируются концом 1990-х годов. Существенный рост создания ЭМК в МО начался в 2014 г. и стал замедляться, начиная с 2019 г. (рис. 3).

Год рассчитывается по дате регистрации документа в ЭМК, показывает сколько суммарно в данном году было зарегистрировано документов. В 2023 г. указано число регистраций на момент проведения анализа.

Динамика регистрации медицинских документов в ЭМК коррелирует с длительностью записей. Самый длительный период наблюдения, который зафиксирован в платформе Webiomed, – 41 год. Число пациентов с данными свыше 10 лет наблюдения составляет 534 664, от 5 до 10 лет наблюдения – 2 085 271, от 3 до 5 лет наблюдения – 1 836 328, до 3 лет наблюдения – 3 937 176 человек.

Общее число документов, представленных на платформе Webiomed, составляет 229 017 146. Среди документов преобладают протоколы врачебных осмотров (52,8%), протоколы лабораторных исследований (24,6%), электронные рецепты (9,23%) и протоколы инструментальных исследований (7,7%) (таблица 5).

АНАЛИЗ ПРИЗНАКОВ, ИЗВЛЕЧЕННЫХ ИЗ ЭМК

В ходе исследования платформой Webiomed было обработано свыше 22 901 714 СЭМД, содержащих ДРКП, из которых было извлечено 819 137 323 признаков. При этом общее число признаков, накопленных в платформе,

Таблица 4 — Распределение пациентов по наличию и числу записей в ЭМК

Показатель	Число пациентов	Доля случаев от общего числа
Число пациентов без медицинских документов («пустые» регистрации)	20 836 936	71,29
Число пациентов, по которым есть медицинская информация	8 393 403	28,71
в т.ч. число пациентов с 1 – 10 медицинскими документами	4 944 606	16,92
в т.ч. число пациентов с более чем 10 медицинскими документами	3 448 797	11,8
Число пациентов, по которым есть случаи обращения в МО (законченный случай, история болезни)	6 496 001	22,22
Число пациентов, по которым есть случаи обращения в МО (хотя бы один законченный случай или история болезни) и есть хотя бы 1 медицинский документ	5 745 152	19,65

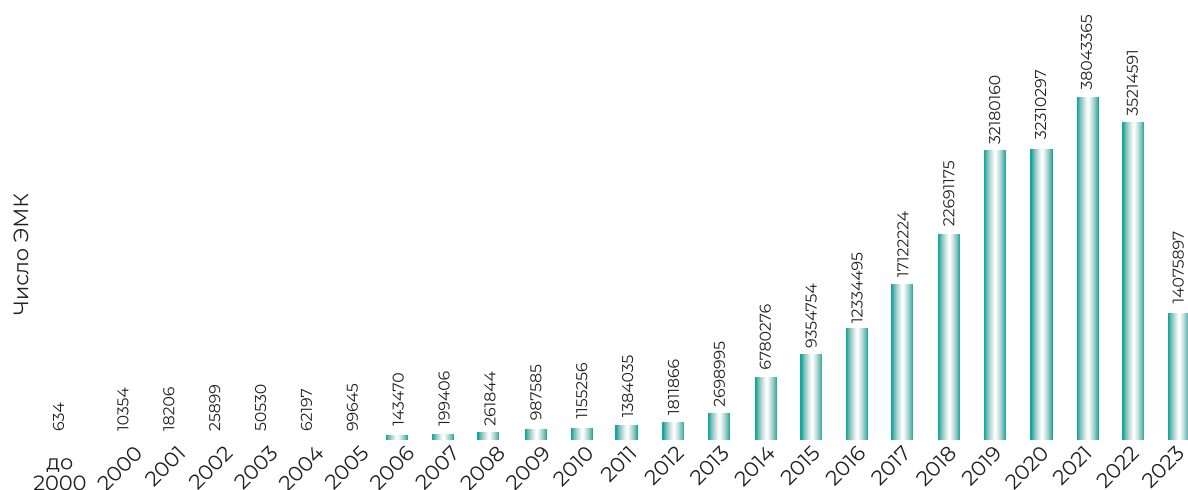


Рисунок 3 — Динамика регистрации медицинских документов в ЭМК.
Год рассчитывается по дате регистрации документа в ЭМК, показывает сколько суммарно в данном году было зарегистрировано документов. В 2023 г. указано число регистраций на момент проведения анализа.

Таблица 5 — Типы медицинских документов в структуре ЭМК платформы Webiomed

Тип документа	Число документов
Протокол врачебного осмотра	121 013 627
Протокол лабораторного исследования	56 443 576
Электронный рецепт	21 134 902
Инструментальное исследование	17 599 537
Листок временной нетрудоспособности	3 448 915
Эпикриз по законченному случаю амбулаторный	3 177 075
Эпикриз в стационаре выписной	2 059 960
Результат диспансеризации/профосмотра	1 374 521
Анкета для граждан в возрасте до 65 лет (версия 3)	610 130
Медицинское свидетельство о смерти	591 204
Протокол операции	355 113
Карта вызова скорой медицинской помощи	277 768
Анкета для граждан в возрасте до 65/75 лет (версии 1, 2)	270 232
Анкета для профосмотра	252 080
Анкета диспансеризации	111 789
Анкета для граждан в возрасте 65/75 лет и старше (версии 1, 2, 3)	32 135
Карта диспансерного наблюдения по гипертонической болезни	21 797
Акт медицинского освидетельствования на состояние опьянения	17 404
Анкета скрининга здоровья	6 861
Медицинское свидетельство о перинатальной смерти	3 888
Итого	229 017 146

составляет 4 558 372 791 (3 739 235 468 признаков получено в готовом виде от МИС) (таблица 6).

Обязательной задачей процесса извлечения является форматно-логический контроль

каждого признака [4]. Система соотносит значение признака не только с правилами контроля для данного конкретного признака, но и с учетом других данных о пациенте. Например,

Таблица 6 — Общая характеристика извлеченных платформой Webiomed признаков

Показатель	Значение	Доля	В т.ч. ошибочных	Доля
Общее число признаков в платформе	4 558 372 791	-	191 305 181	4,20%
из них число признаков, полученных в готовом виде от МИС	3 739 235 468	82,03%	175 195 916	4,69%
из них число признаков, извлеченных платформой	819 137 323	17,97%	16 109 265	1,97%
Число признаков, извлеченных платформой				
Число признаков, извлеченных с помощью NLP	638 280 136	77,92%	5 438 966	0,85%
Число признаков, извлеченных с помощью маппинга (в т.ч. из лабораторных исследований)	177 926 558	21,72%	10 670 299	6,00%
Число признаков, извлеченных программно (из т.ч. документа «Анкета диспансеризации»)	2 930 629	0,36%	0	0%
Дополнительные данные о признаках, полученных в готовом виде от МИС				
Признаки с демографической информацией	394 379 146	10,55%	0	0%
Признаки, характеризующие обращаемость	1 404 274 342	37,56%	0	0%
Признаки, характеризующие заболеваемость	1 790 155 884	47,87%	150 422 060	8,40%
Клинические признаки, переданные в сигнальной информации	150 426 096	4,02%	24 773 856	16,47%

система извлекла что рост пациента «160 км.», проверила по справочнику и выяснила, что диапазон допустимых значений роста включает 50 см – 250 см. Система определяет значение «160 км.» как ошибочное, сохраняет его в базу данных платформы, не отмечает как недоступное для последующего использования.

Для извлечения признаков платформой Webiomed были использованы 3 способа: NLP-модели машинного обучения, маппинг (англ. mapping) и программное извлечение. Подобные виды извлечения являются достаточно трудоемкими как технически, так и организационно. Выбор технологии определяется уровнем структурированности данных, а также самими данными.

При извлечении признаков с помощью технологии NLP создается комплекс моделей на основе машинного обучения, которые способны извлекать из неструктурированного текста машинно-понимаемые признаки [4].

Извлечение с помощью маппинга используется тогда, когда информация в ЭМК хранится в частично-структурированном виде. Часто технология маппинга используется для извлечения признаков из лабораторных тестов, где исходно информация представлена в виде таблицы (название и значение лабораторного теста, его код по Федеральному справочнику лабораторных исследований (ФСЛИ)). В этом случае извлекается пара "код ФСЛИ = значение", далее по коду ФСЛИ в справочнике маппинга ищется уже код

признака и таким образом в результате в базе данных Webiomed сохраняется значение извлеченного признака.

В программном извлечении, как легко можно понять из названия, для определения значения признака используется специально разработанный программный код. Например, при помощи технологии NLP из неструктурированных данных можно извлечь признаки о росте и весе, а затем уже с помощью программного кода можно рассчитать ИМТ и сохранить его в базе данных. Таким образом, в ходе этих последовательных процессов из врачебного протокола будет извлечено 3 признака, 2 из которых указал врач, а третий был программно рассчитан уже самой платформой.

Наиболее часто извлекаемыми признаками являются: частота сердечных сокращений (58 036 477 значений в базе данных), систолическое (50 048 668) и диастолическое (48 889 677) артериальное давление, частота дыхательных движений (49 459 531), температура тела (43 639 113), а также многие лабораторные показатели. В частности, информация об общем холестерине извлечена 8 101 288 раз, гемоглобине – 7 404 256 раз, креатинине – 6 370 174 раз. Отдельную группу признаков составляют симптомы заболеваний, например, головокружение зафиксировано 5 810 970 раз, одышка – 5 641 646 раз. К редко встречающимся признакам (выявлено менее 10 раз) могут относиться конкретные комбинации лекарственных препаратов, например, «Ацетилсалициловая Кислота+Кофеин»,

Таблица 7 — Распределение извлеченных признаков по видам и их количеству

Вид признака	Число видов признаков	Кол-во признаков	Доля	В т.ч. ошибочных	Доля
Демографические данные (пол, дата рождения, регион проживания, данные о месте работы, социальный статус и т.д.)	39	394 379 146	8,65%	0	0%
Данные об обращении в МО (даты и типы случаев обращения, исходы и результаты лечения, коды диагнозов и т.д.)	14	1 404 274 342	30,81%	0	0%
Данные о зарегистрированных заболеваниях (названия, коды МКБ-10, даты первой регистрации заболевания и последнего обращения, тип заболевания (острое или хроническое) и т.д.)	6	1 790 155 884	39,27%	0	0%
Клинические данные – физиологические данные и данные объективного осмотра пациента (например, АД, рост, вес, ЧСС и т.д.), а также симптомы и синдромы, извлеченные из врачебных осмотров (например, курение, головная боль, отягощенный анамнез и т.д.)	394	622 205 993	13,65%	28 690 750	4,61%
Лабораторные данные (общий холестерин, глюкоза крови, креатинин и т.д.)	326	202 068 195	4,43%	12 152 291	6,01%
Данные инструментальных исследований (извлекаются из протоколов обследований, включая УЗИ, ЭКГ, КТ, рентгенографию, маммографию и т.д.)	134	14 203 332	0,31%	40 080	0,28%
Данные о лекарственном лечении (наименование, кратность приема, доза и т.д.)	2148	131 085 899	2,88%	0	0%
Итого	3061	4 558 372 791	100,00	40 883 121	0.9%

«Дротаверина Гидрохлорид+Кодеина Фосфата Полугидрат+Парацетамол».

Извлеченные признаки можно объединить в 7 групп (таблица 7).

Таким образом, на момент проведения выгрузки данных (16.10.2023) в платформе Webiomed:

- зарегистрировано 856 МО в 28 субъектах РФ,
- включено 229 017 146 документов, представляющих данные 29 230 375 пациентов,
- 8 393 403 пациента имеют в ЭМК медицинскую информацию, из них 3 448 797 пациентов имеют более 10 медицинских документов,
- у 4 456 263 пациентов отмечена глубина сбора данных свыше 3х лет,
- извлечено 4 558 372 791 признаков.

ОБСУЖДЕНИЕ

ЭМК как источник данных

Внедрение ЭМК является одной из ключевых задач информатизации в России. В соответствии с федеральным проектом «Создание единого цифрового контура в сфере здравоохранения»,

100% государственных МО должны перейти на ведение ЭМК к концу 2024 г. Таким образом, стратегия государства в области информатизации здравоохранения предусматривает накопление цифровых данных об оказанной медицинской помощи, что в свою очередь, приводит к постоянному росту интереса к ЭМК как одному из важнейших источников ДРКП.

Сбор и анализ ЭМК как источника ДРКП является основной технологической задачей платформы Webiomed, позволяя ей в полностью автоматическом режиме формировать цифровые профили пациентов с подробной клинической информацией, которая далее выступает источником для различных задач, включая работу системы поддержки принятия врачебных решений (СППВР) и управленческую аналитику.

Возможности и ограничения ЭМК

В настоящее время можно выделить 2 ключевых ограничения использования ЭМК как источника данных:

1. «Пустые» ЭМК (которые не содержали ни одной медицинской записи, 71,29% от общего числа ЭМК);
2. Частично-заполненные ЭМК, в которых медицинская информация представлена не полностью (включены не все данные, есть не все медицинские документы и т.п.).

Подобные ЭМК не могут быть использованы (либо их использование сильно ограничено) для решения многих задач, которые стоят перед системой здравоохранения [3–5].

Улучшение качества заполнения ЭМК – одна из ключевых задач, которая на сегодня должна стоять перед разработчиками МИС и руководителями в сфере здравоохранения. ДРКП, которые извлекаются из ЭМК, лежат не только в основе многих задач/проектов/инициатив, но и формируют основы ценностно-ориентированных подходов системы здравоохранения. Таким образом, улучшение качества в начале этой цепочки – на этапе включения этих данных в ЭМК, значительно улучшит качество реализуемых на основе ДРКП проектов [1].

В представленном анализе мы не проводили определения причин «пустых регистраций» и недостающих данных, но потенциально могут быть определены следующие направления для исследований и последующего поиска решений:

- **структура ЭМК:** детализация разделов и полей, удобство заполнения для пользователей, обязательность полей и т.д., использование листингов и иных инструментов работы с данными, упрощающие ввод данных;
- **обеспечение и контроль качества на уровне ЭМК:** валидационные правила для полей, автоматические подсчеты показателей, определяемых по формулам, синхронизация ввода лабораторных данных в выписках и заключениях и т.д.;
- **ввод данных:** обязательность ввода ключевых полей, формирование культуры работы с данными на уровне МО, упрощение работы по вводу (например, за счет структуры ЭМК).

Важно подчеркнуть, что обязательность ввода данных с помощью структурированных полей без изменений на уровне ЭМК (структура, контроль качества), а также без выделения у специалистов здравоохранения необходимых временных ресурсов, может привести к ухудшению качества ведения ЭМК, что, в свою очередь,

приведет к ошибкам (порой значительным) на уровне результатов задач/проектов/инициатив, построенных на этих данных.

Тем не менее, несмотря на существующие ограничения, ЭМК уже на текущем этапе своего развития обладают не только высоким потенциалом, но и значимыми существующими (внедренными и продолжающимися внедряться) возможностями их использования. Доля ЭМК, классифицированных нами как имеющие достаточное наполнение (более чем 10 медицинских документов), составила 11,8%. На первый взгляд, относительные цифры (%) ЭМК, в которых есть медицинские документы, кажутся незначительными, недостаточными для решения вопросов, которые можно ставить перед платформой. Однако абсолютные цифры обеспечивают возможность достижения необходимых целей и решения любых задач: 8 393 403 пациента имеют достаточную для исследований и повторного использования медицинскую информацию, в том числе 3 448 797 пациентов имеют более 10 медицинских документов (таблица 4). Более того, динамика регистрации медицинских документов в ЭМК на платформе Webiomed ежедневно продолжается (рисунки 3 и 4). И это означает, что абсолютное число пациентов, имеющих медицинскую информацию и данные которых представлены на платформе Webiomed, постоянно растет с течением времени.

Таким образом, несмотря на существующие ограничения, формирование наборов ДРКП из «больших данных» ЭМК является ценным и перспективным источником информации, который со временем будет только расширяться как в объемах данных, так и в масштабах его использования. Обеспечение качества процессов на уровне извлечения данных из ЭМК является одной из приоритетных задач платформы Webiomed.

Извлечение данных из ЭМК

Важно понимать, что, несмотря на гигантские объемы накопленных в России ЭМК, с точки зрения анализа они являются так называемыми «сырыми» (от англ. raw) данными, не пригодными вследствие текстового и неструктурированного формата к интерпретации и построению аналитики.

Для формирования наборов ДРКП из ЭМК и последующего их использования, необходимо

создание и постоянное совершенствование довольно сложного программного комплекса. Применение технологий ИИ, в частности NLP, позволяет обеспечить работу такого комплекса в полностью автоматическом режиме. Точность извлечения данных платформы Webiomed достигает 95–99%.

Анализ опыта функционирования платформы Webiomed показывает, что главными сложностями обеспечения качества работы на этапе извлечения ДРКП из ЭМК в России являются:

- большое влияние на точность извлечения особенностей русского языка;
- локальные традиции ведения врачебных записей (на уровне субъектов РФ, отдельных МО или научных школ);
- технические особенности заполнения ЭМК в интерфейсе МИС;
- отсутствие единой нормативно-справочной информации для кодирования клинических записей в электронных медицинских документах;
- преобладание неструктурированных записей, в том числе хранимых в виде обычных текстовых документов (таблица 5).

В этой связи при создании и развитии программных комплексов для сбора и извлечения ДРКП из ЭМК необходимо уделять большое внимание постоянному контролю качества и точности извлечения данных, переобучению моделей машинного обучения, использующих технологии NLP для интерпретации медицинских записей. Такая работа может быть связана с существенными инвестициями в научные исследования и опытно-конструкторские работы, включая создание новых способов и методов работы с данными, а также патентованием полученных результатов интеллектуальной деятельности [13].

Накопление данных

Формируемые из ЭМК с помощью технологий ИИ цифровые профили пациентов позволяют создавать наборы ДРКП, на основании анализа которых, в свою очередь, формируется новая клиническая и управленческая информация. Основу процесса получения новых знаний составляет понятие «Извлеченный признак», который может быть охарактеризован по 3 классификациям (таблицы 6 и 7):

- 1) метод извлечения (NLP, маппинг, программное извлечение);
- 2) вид признака (физиологические, клинические и др.);
- 3) доверие к признаку (ошибочный или нет).

Использование правил форматно-логического контроля позволило оценивать извлеченные данные на предмет ошибок [4]. Благодаря такому контролю было выявлено, что 4,20% признаков от общего числа (4 558 372 79) являются ошибочными. Самая большая доля ошибок содержится в данных о зарегистрированных у пациента окончательных клинических диагнозах (таблица 6). Основной причиной ошибок является дублирование хронических заболеваний – это связано с тем, что в объединенной ЭМК пациента находится несколько записей о регистрации одного и того же хронического заболевания.

Кроме того, наличие ошибочных признаков может быть обусловлено децентрализованной архитектурой внедрения МИС в некоторых субъектах РФ, в результате чего одно и то же хроническое заболевание регистрируется в разных инсталляциях МИС, что приводит потом к дублированию записей о таких заболеваниях при их централизованном сборе, включая формируемые формы государственной статистической отчетности. Еще одной причиной является отсутствие в некоторых МИС контроля правильности кодирования сведений о заболеваниях.

Выявленные ошибочные записи из базы данных не удаляются. Это крайне важно для последующего научного анализа частоты ошибок и совершенствования методов контроля качества, определения распространенности их в разрезе различных МО или систем ведения ЭМК. При этом важно подчеркнуть, что ошибочные признаки никогда не используются при формировании и анализе цифрового профиля пациента. Другими словами, ошибочные признаки не участвуют в формировании новых знаний, используемых целевой аудиторией платформы в принятии решений.

Чем больше разнообразных данных о пациенте накоплено в его цифровом профиле, тем больше его ценность с точки зрения применения ДРКП для решения различных задач в системе здравоохранения, в фармацевтической индустрии [4]. Анализ популяции пациентов (рисунок 2, таблица 3) и анализ распределения

извлеченных признаков по видам и их количеству (таблица 7) показывают пользователям платформы возможности ее использования для прикладных задач.

Так, например, разработка концепции исследования реальной клинической практики (Real World Evidence, RWE), в отличие от клинических (интервенционных) исследований, включает обязательный этап – оценку ДРКП [7]. В рамках этого процесса необходимо оценить как источник данных, так и получаемые в итоге данные. Знание видов признаков и их числа показывает направления возможностей для формулировки целей и конечных точек разрабатываемого научного исследования (таблица 7). На следующем

этапе разработки исследования необходимо оценить, какие конкретно признаки имеются в базе для описанной популяции исследования (критерии включения/невключения) и сформулированных целей и конечных точек. Сформулированные на основании анализа и отбора наборы могут быть выгружены в виде машинно-читаемых файлов и использоваться в статистическом анализе научного исследования [4].

Целевая аудитория и типы проектов на основе данных из ЭМК

Анализ накопленного опыта сбора и обработки ДРКП с точки зрения их применения в различных проектах представлен в таблице 8.

Таблица 8 — Целевая аудитория проектов, построенных на ДРКП, типы проектов, их цели и задачи

Целевая аудитория (ЦА)	Типы проектов	Цели/задачи
Государственные органы управления здравоохранением, государственные МО	Внедрение СППВР для врачей	Сокращение нагрузки на врачей в части оценки данных ЭМК, сокращение врачебных ошибок, улучшение профилактики и лечения заболеваний [14–16]
	Внедрение управленческой аналитики для руководителей	Улучшение контроля заболеваемости и тем самым положительное влияние на сокращение заболеваемости и смертности, повышение оперативности и эффективности принятия решений [15, 16]
Коммерческие МО	Внедрение СППВР для врачей	Обращение внимания на пациентов высокого риска, поддержка в принятии решений по дополнительной диагностике и рекомендации по лечению [14]
	Внедрение управленческой аналитики для руководителей	Улучшение контроля за отработкой выявленных подозрений, контроль выполнения критериев качества оказания медицинской помощи. Поиск клиентов / предложения услуг клиники, повышение продажи услуг клиники, в т.ч. повторных продаж в существующей клиентской базе
Фармацевтические компании	Проекты, построенные на ДРКП	Выявление подозрений на заболевания, пациентов высокого риска с помощью СППВР [17, 18]
	Исследования ДРКП	Оценка распространенности факторов риска и заболеваний, оценка эффективности и безопасности, паттерны лечения и т.д. [19–21]
	Фармако-экономический анализ	Повышение эффективности исследований на основе результатов ДРКП
Контрактно-исследовательские компании (КИО)	ДРКП	Повышение продаж услуг КИО, в некоторых случаях возможность сохранить бизнес/коллектив компании из-за остановки КИ новых лекарств [22]
Научно-исследовательские организации	Проведение научных исследований и разработок на основе ДРКП	Исследования ДРКП, разработка и валидация моделей машинного обучения. Возможность привлечь грантовое финансирование и гарантированно выполнить условия гранта, повышение наукометрических показателей ученых, включая публикации в журналах WoS и Scopus, выступления на ведущих научных конференциях [23–28]
Промышленные предприятия	Внедрение комплексных систем управления здоровьем	Управление риском потери трудоспособности, повышение эффективности профилактических медицинских осмотров – сокращение издержек предприятия из-за утраты здоровья, повышение эффективности корпоративных программ сохранения здоровья

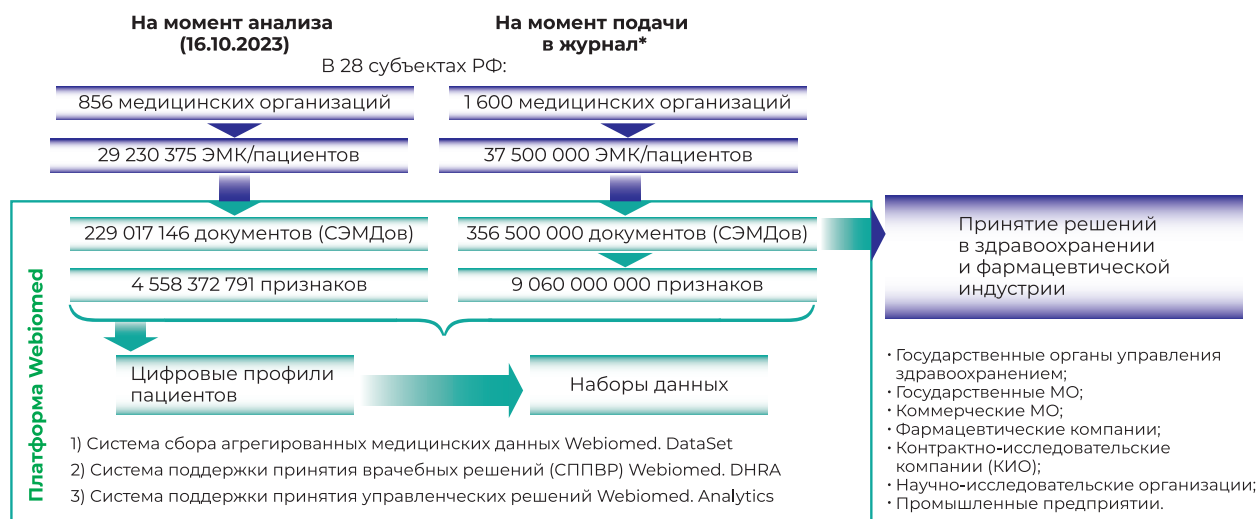


Рисунок 4 — Общая структура, состав и ключевые возможности платформы Webiomed. * – т.к. на момент подачи статьи в журнал мы получили обновление по ключевым показателям, то мы решили целесообразным отобразить их в статье.

ЗАКЛЮЧЕНИЕ

Использование ДРКП для решения различных задач здравоохранения и фармацевтической индустрии приобретает все большее значение как в мире, так и в России. В настоящей публикации на примере платформы Webiomed мы проследили ключевые процессы, этапы и аспекты работы с ЭМК как источником ДРКП, начиная непосредственно от источника данных до возможностей применения полученных результатов в системе здравоохранения и фармацевтической индустрии. Основные показатели представлены на рисунке 4.

Полученные результаты свидетельствуют, что анализ и извлечение ДРКП из обезличенных ЭМК действительно позволяет создавать большие наборы структурированных данных. Однако, наряду с существенными перспективами, на данном этапе развития ЭМК имеются и значительные ограничения, влияющие на качество извлекаемых данных. В этой связи важно и необходимо применять продуманную стратегию повышения качества данных как на уровне ЭМК, так и в процессе

извлечения признаков и создания цифровых двойников.

По нашим данным, представленный в данной статье анализ является первым в России исследованием потенциала извлечения ДРКП из ЭМК с использованием технологий ИИ. Постоянный и усиливающийся рост и развитие технологии обеспечиваются за счет следующих ключевых факторов: интеграция с лидирующими МИС МО России, подключение как можно большего числа МО, постоянное развитие возможностей в анализе накопленных данных, а также создание ценной для системы здравоохранения и фармацевтической отрасли клинической и управленческой аналитики.

Обеспечение качества работы с ДРКП на всех этапах, начиная от разработки структуры ЭМК и ввода данных до формирования цифровых двойников, является важнейшим условием их применения для решения различных задач в системе здравоохранения и фармацевтической индустрии.

Конфликт интересов. Авторы статьи являются сотрудниками компании «К-Скай» и фармацевтической компании АО «Санофи Россия».

ЛИТЕРАТУРА/REFERENCES

1. Гольдина Т.А., Бурмистров В.А., Ефименко И.В., Хорошевский В.Ф. Искусственный интеллект в здравоохранении: Real World Data и Patient Voice – готовы ли мы к новым реалиям? // Медицинские технологии. Оценка и выбор. – 2021. – №2(43). – С.22-31. [Gol'dina TA, Burmistrov VA, Efimenko IV, Horoshevskij VF. Iskusstvennyj intellekt v zdravoohranenii: Real World Data i Patient Voice – gotovy li my k novym realiym? Medicinskie tekhnologii. Ocenka i vybor. 2021; 2(43): 22-31. (In Russ.)] doi: 10.17116/medtech20214302122.
2. Гусев А.В., Владимирский А.В., Голубев Н.А., Зарубина Т.В. Информатизация здравоохранения Российской Федерации: история и результаты развития // Национальное здравоохранение. – 2021. – Т.2/ – №3. – С.5-17. [Gusev AV, Vladimirskij AV, Golubev NA, Zarubina TV. Informatizaciya zdravoohraneniya Rossijskoj Federacii: istoriya i rezul'taty razvitiya. Nacional'noe zdravoohranenie. 2021; 2(3): 5-17. (In Russ.)] doi: 10.47093/2713-069X.2021.2.3.5-17.
3. Колбин А.С., Белоусов Д.Ю., Зырянов С.К. и др. Исследования реальной клинической практики. – М.: Буки Веди, 2020. – 208 с.: ил. [Kolbin AS, Belousov DYU, Zyryanov SK, et al. Issledovaniya real'noj klinicheskoy praktiki. M.: Buki Vedi, 2020. 208 p.: il. (In Russ.)]
4. Гусев А.В., Зингерман Б.В., Тюфилин Д.С., Зинченко В.В. Электронные медицинские карты как источник данных реальной клинической практики // Реальная клиническая практика: данные и доказательства. – 2022. – Т.2. – №2. – С.8-20. [Gusev AV, Zingerman BV, Tyufilin DS, Zinchenko VV. Elektronnye medicinskie karty kak istochnik dannyh real'noj klinicheskoy praktiki. Real'naya klinicheskaya praktika: dannye i dokazatel'stva. 2022; 2(2): 8-20. (In Russ.)] doi: 10.37489/2782-3784-myrwd-13.
5. Гольдина Т.А. Исследования рутинной клинической практики: от получения данных к оценке медицинских технологий и принятию решений в здравоохранении // Медицинские технологии. Оценка и выбор. – 2018. – №1(31). – С.21-29. [Gol'dina TA. Issledovaniya rutinnoj klinicheskoy praktiki: ot polucheniya dannyh k ocenke medicinskih tekhnologij i prinyatiyu reshenij v zdravoohranenii. Medicinskie tekhnologii. Ocenka i vybor. 2018; 1(31): 21-29. (In Russ.)]
6. Yamamoto K, Sumi E, Yamazaki T, et al. A pragmatic method for electronic medical record-based observational studies: developing an electronic medical records retrieval system for clinical research. BMJ Open. 2012; 2: e001622. doi: 10.1136/bmjopen-2012-001622.
7. Хомицкая Ю.В., Осина Г.С., Новодережкина Е.А., Гольдина Т.А. Исследования рутинной практики: как разработать и внедрить новое направление. Системный подход в обеспечении качества результатов // Медицинские технологии. Оценка и выбор. – 2021. – №4(43). – С.16-27. [Homickaya YUV, Osina GS, Novoderezhkina EA, Gol'dina TA. Issledovaniya rutinnoj praktiki: kak razrabotat' i vnedrit' novoe napravlenie. Sistemnyj podhod v obespechenii kachestva rezul'tatov. Medicinskie tekhnologii. Ocenka i vybor. 2021; 4(43): 16-27. (In Russ.)] doi: 10.17116/medtech20214304116.
8. Гольдина Т.А., Колбин А.С., Белоусов Д.Ю., Боровская В.Г. Обзор исследований реальной клинической практики // Качественная клиническая практика. – 2021. – №1. – С.56-63. [Gol'dina TA, Kolbin AS, Belousov DYU, Borovskaya VG. Obzor issledovanij real'noj klinicheskoy praktiki. Kachestvennaya klinicheskaya praktika. 2021; 1: 56-63. (In Russ.)] doi: 10.37489/2588-0519-2021-1-56-63.
9. Fu S, Chen D, He H, et al. Clinical concept extraction: A methodology review. J Biomed Inform. 2020; 109: 103526. doi: 10.1016/j.jbi.2020.103526.
10. Zhao ., Weroh SJ, Goode EL, et al. Generating real-world evidence from unstructured clinical notes to examine clinical utility of genetic tests: use case in BRCAness. BMC Med Inform Decis Mak. 2021; 21: 3. doi: 10.1186/s12911-020-01364-y.
11. van Laar S, Gombert-Handoko K, Groenwold R, et al. Real-World Metastatic Renal Cell Carcinoma Treatment Patterns and Clinical Outcomes in The Netherlands. Front. Pharmacol. 2022; 13: 803935. doi: 10.3389/fphar.2022.803935.
12. Webiomed.Connect. <https://webiomed.ru/products/integratsionnyi-shliuz-webiomedconnect>.
13. Кобякова О.С. Технологии искусственного интеллекта в здравоохранении в глобальной патентной экосистеме: 2000-2023 гг. // Врач и информационные технологии. – 2024. – №1. – С.44-59. [Kobyakova OS. Tekhnologii iskusstvennogo intellekta v zdravoohranenii v global'noj

- patentnoj ekosisteme: 2000-2023 gg. Vrach i informacionnye tekhnologii. 2024; 1: 44-59. (In Russ.)) doi: 10.25881/18110193_2024_1_44.
14. Гусев А.В., Токарев С.А., Гаврилов Д.В., Кузнецова Т.Ю. Применение автоматизированной системы поддержки принятия врачебных решений при диспансеризации взрослого населения для контроля правильности оценки уровня сердечно-сосудистого риска // Менеджмент качества в медицине. – 2022. – №1. – С.72-79. [Gusev AV, Tokarev SA, Gavrilo DV, Kuznecova TYU. Primenenie avtomatizirovannoj sistemy podderzhki prinyatiya vrachebnyh reshenij pri dispanserizacii vzroslogo naseleniya dlya kontrolya pravil'nosti ocenki urovnya serdechno-sosudistogo riska. Menedzhment kachestva v medicine. 2022; 1: 72-79. (In Russ.)]
 15. Курдюмов Д.А., Кашин А.В., Рябов Н.Ю. и др. Опыт применения технологий искусственного интеллекта для развития профилактического здравоохранения на примере Кировской области // Менеджер здравоохранения. – 2023. – №6. – С.62-69. [Kurdyumov DA, Kashin AV, Ryabov NYU, et al. Opyt primeneniya tekhnologij iskusstvennogo intellekta dlya razvitiya profilakticheskogo zdravoohraneniya na primere Kirovskoj oblasti. Menedzher zdravoohraneniya. 2023; 6: 62-69. (In Russ.)]. doi: 10.21045/1811-0185-2023-6-62-69.
 16. Белорус О.В., Токарев С.А., Захарова М.Г., Гусев А.В. Результаты пилотного внедрения технологий искусственного интеллекта в систему здравоохранения Ямало-Ненецкого автономного округа // Менеджмент качества в медицине. – 2023. – №3. – С.39-47. [Belorus OV, Tokarev SA, Zaharova MG, Gusev AV. Rezul'taty pilotnogo vnedreniya tekhnologij iskusstvennogo intellekta v sistemu zdravoohraneniya YAmalo-Neneckogo avtonomnogo okruga. Menedzhment kachestva v medicine. 2023; 3: 39-47. (In Russ.)]
 17. Гиляревский С.Р., Гаврилов Д.В., Гусев А.В. Результаты ретроспективного анализа записей электронных амбулаторных медицинских карт пациентов с хронической сердечной недостаточностью: первый российский опыт // Российский кардиологический журнал. – 2021. – Т.26. – №5. – С.147-155. [Gilyarevskij SR, Gavrilo DV, Gusev AV. Rezul'taty retrospektivnogo analiza zapisej elektronnyh ambulatornyh medicinskih kart pacientov s hronicheskoj serdechnoj nedostatochnost'yu: pervyj rossijskij opyt. Rossijskij kardiologicheskij zhurnal. 2021; 26(5): 147-155. (In Russ.)] doi: 10.15829/1560-4071-2021-4502.
 18. Батюшин М.М., Касимова И.С., Гаврилов Д.В. и др. Распространенность хронической болезни почек по данным ретроспективного когортного исследования «эпидемиология ХБП» (город Киров) // Нефрология и диализ. – 2021. – Т.23. – №2. – С.192-202. [Batyushin MM, Kasimova IS, Gavrilo DV, et al. Rasprostranennost' hronicheskoj bolezni pochek po dannym retrospektivnogo kogortnogo issledovaniya «epidemiologiya HBP» (gorod Kirov). Nefrologiya i dializ. 2021; 23(2): 192-202. (In Russ.)] doi: 10.28996/2618-9801-2021-2-192-202.
 19. Дружилов М.А., Кузнецова Т.Ю., Дружилова О.Ю. и др. Фибрилляция предсердий и хроническая болезнь почек: основные клинические характеристики пациентов в отдельных субъектах Российской Федерации // Кардиоваскулярная терапия и профилактика. – 2023. – Т.22. – №4. – С.57-67. [Druzhilov MA, Kuznecova TYU, Druzhilova OYU, et al. Fibrillyaciya predserdij i hronicheskaya bolezni' pochek: osnovnye klinicheskie harakteristiki pacientov v otdel'nyh sub"ektah Rossijskoj Federacii. Kardiovaskulyarnaya terapiya i profilaktika. 2023; 22(4):57-67. (In Russ.)] doi: 10.15829/1728-8800-2023-3544.

20. Дружилов М.А., Кузнецова Т.Ю., Дружилова О.Ю. и др. Результаты ретроспективного анализа частоты ишемического инсульта и назначения антикоагулянтной терапии пациентам с фибрилляцией предсердий в зависимости от индекса массы тела // Российский кардиологический журнал. – 2023. – Т.28. – №5. – С.46-54. [Druzhilov MA, Kuznecova TYU, Druzhilova OYU, et al. Rezul'taty retrospektivnogo analiza chastoty ishemicheskogo insul'ta i naznacheniya antikoagulyantnoj terapii pacientam s fibrillyaciej predserdij v zavisimosti ot indeksa massy tela. Rossijskij kardiologicheskij zhurnal. 2023; 28(5): 46-54. (In Russ.)]. doi: 10.15829/1560-4071-2023-5359.
21. Korneva V, Druzhilov M, Kuznetsova T, Gavrilov D, Gusev A. Obesity and the risk of ischemic stroke in patients with atrial fibrillation: Findings from a big data analysis. *Atherosclerosis*. 2023; 379: 88. doi: 10.1016/j.atherosclerosis.2023.06.327.
22. Дружилов М.А., Кузнецова Т.Ю., Гаврилов Д.В., Гусев А.В. Верификация субклинического каротидного атеросклероза в рамках риск-стратификации при избыточном весе и ожирении: роль методов машинного обучения в формировании диагностического алгоритма // Кардиоваскулярная терапия и профилактика. – 2022. – Т.21. – №7. – С.79-87. [Druzhilov MA, Kuznecova TYU, Gavrilov DV, Gusev AV. Verifikaciya subklinicheskogo karotidnogo ateroskleroza v ramkah risk-stratifikacii pri izbytochnom vese i ozhireнии: rol' metodov mashinnogo obucheniya v formirovanii diagnosticheskogo algoritma. Kardiovaskulyarnaya terapiya i profilaktika. 2022; 21(7): 79-87. (In Russ.)] doi: 10.15829/1728-8800-2022-3222.
23. Гаврилов Д.В., Кузнецова Т.Ю., Дружилов М.А. и др. Прогнозирование наличия субклинического каротидного атеросклероза у пациентов с избыточным весом и ожирением при помощи модели машинного обучения // Российский кардиологический журнал. – 2022. – Т.27. – №4. – С.40-47. [Gavrilov DV, Kuznecova TYU, Druzhilov MA, et al. Prognozirovanie nalichiya subklinicheskogo karotidnogo ateroskleroza u pacientov s izbytochnym vesom i ozhireniem pri pomoshchi modeli mashinnogo obucheniya. Rossijskij kardiologicheskij zhurnal. 2022; 27(4): 40-47. (In Russ.)] doi: 10.15829/1560-4071-2022-4871.
24. Gavrilov D, Kuznetsova T, Gusev A, Korsakov I, Novitskiy R. Application of a clinical decision support system to assess the severity of the new coronavirus infection COVID-19, *European Heart Journal*. 2021; 42(1): ehab724.3054. doi: 10.1093/eurheartj/ehab724.3054.
25. Андрейченко А.Е., Лучинин А.С., Ившин А.А. и др. Разработка и валидация моделей прогнозирования общего риска преэклампсии и риска ранней преэклампсии с использованием алгоритмов машинного обучения в первом триместре беременности // Акушерство и гинекология. – 2023. – №10. – С.94-107. [Andrejchenko AE, Luchinin AS, Ivshin AA, et al. Razrabotka i validaciya modelej prognozirovaniya obshchego riska preeklampsii i riska rannej preeklampsii s ispol'zovaniem algoritmov mashinnogo obucheniya v pervom trimestre beremennosti. Akusherstvo i ginekologiya. 2023; 10: 94-107. (In Russ.)] doi: 10.18565/aig.2023.101.
26. Ivshin AA, Boldina JS, Gusev AV, et al. Artificial Intelligence on Guard of Reproductive Health. *Ad Alta, Journal of Interdisciplinary Research*. 2022; 12(02): 325-329. doi: 10.33543/1202.
27. Gusev AV, Novitskiy RE, Ivshin AA, et al. Extraction of medical data from electronic medical records using NLP algorithms. *Ad Alta, Journal of Interdisciplinary Research*. 2022; 12(02): 314-319. doi: 10.33543/1202.