

# Электронные медицинские карты как источник данных реальной клинической практики

Гусев А. В.<sup>1,2</sup> , Зингерман Б. В.<sup>4</sup> , Тюфилин Д. С.<sup>2</sup> , Зинченко В. В.<sup>3</sup> 

<sup>1</sup> — ООО «К-Скай», Россия, Петрозаводск

<sup>2</sup> — Федеральное государственное бюджетное учреждение «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, Россия, Москва

<sup>3</sup> — Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», Россия, Москва

<sup>4</sup> — ООО «ТелеПат», Россия, Москва

## Аннотация

В настоящее время в здравоохранении Российской Федерации идёт активное внедрение информационных технологий. Доля государственных и муниципальных медицинских организаций, внедривших различные медицинские информационные системы, увеличилась с 3,9% в 2007 г. до 91% в 2021-м. Одна из ключевых задач информатизации — внедрение электронных медицинских карт (ЭМК), в которых накапливаются большие объёмы данных реальной клинической практики (Real-World Data; RWD). При всей значимости ЭМК как источника RWD, в них имеются ряд недостатков, таких как децентрализованный характер систем ведения баз данных, неструктурированное хранение информации и т. д. В статье представлено описание последовательных процессов по сбору качественных RWD на основе ЭМК, включая применение технологий искусственного интеллекта, для целей научных исследований, создания систем поддержки принятия решений, статистического анализа и т. д. Основу предложенной методики составляет централизованный сбор сведений из ЭМК в так называемом озере данных, где накапливается как можно большее количество «сырых данных» по пациенту (raw data), и последующее извлечение данных из неструктурированных записей посредством моделей natural language processing (NLP). Предложенная технология, при условии постоянного совершенствования, позволит получить правильное и всестороннее решение для умелого понимания любого текста из любой медицинской записи.

**Ключевые слова:** данные реальной клинической практики; электронная медицинская карта; большие данные; искусственный интеллект; машинное обучение; система поддержки принятия решений

### Для цитирования:

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;2 (2):8 — 20. <https://doi.org/10.37489/2782-3784-myrd-13>.

Поступила: 13 мая 2022 г. Одобрена: 15 мая 2022 г. Опубликовано: 7 июня 2022 г.

## Electronic medical records as a source of real-world clinical data

Gusev A. V.<sup>1,2</sup> , Zingerman B. V.<sup>4</sup> , Tyufilin D. S.<sup>2</sup> , Zinchenko V. V.<sup>3</sup> 

<sup>1</sup> — LLC «K-Sky», Russia, Petrozavodsk

<sup>2</sup> — Federal State Budgetary Institution «Central Research Institute for the Organization and Informatization of Healthcare» of the Ministry of Health of Russia, Russia, Moscow

<sup>3</sup> — State Budgetary Institution of Healthcare of the City of Moscow «Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of the City of Moscow», Russia, Moscow

<sup>4</sup> — LLC «TelePat», Russia, Moscow

## Abstract

Currently, information technologies are being actively introduced in the healthcare of the Russian Federation. The share of state and municipal medical organizations that have implemented various medical information systems increased from 3.9% in 2007 to 91% in 2021. One of the key tasks of informatization is the introduction of electronic medical records (EMRs), which accumulate large amounts of Real-World Data (RWD). Despite the importance of EHR as a source of RWD, they have a number of shortcomings, such as the decentralized nature of database management systems, unstructured information storage, etc. The article describes the sequential processes for collecting high-quality RWD based on EHR, including the use of artificial intelligence technologies, for the purposes of scientific research, the creation of decision support systems, statistical analysis, etc. The basis of the proposed methodology is the centralized collection of information from EMR in the so-called data lakes, where as much as possible of raw data on the patient is accumulated and subsequent extraction of data from unstructured records through natural language processing (NLP) models. The proposed technology, subject to continuous improvement, will provide a correct and comprehensive solution for the skilful understanding of any text from any medical record.

**Keywords:** real-world clinical data; electronic medical record; big data; artificial intelligence; machine learning; decision support system

**For citation:**

Gusev AV, Zingerman BV, Tyufilin DS, Zinchenko VV. Electronic medical records as a source of real-world clinical data. *Real-World Data & Evidence*. 2022;2 (2):8 — 20. <https://doi.org/10.37489/2782-3784-myrrwd-13>.

Received: May 13, 2022. Accepted: May 15, 2022. Published: June 7, 2022.

**Введение**

Цифровое здравоохранение является одним из самых больших и быстрорастущих технологических рынков [1]. Повсеместное внедрение электронных медицинских карт (ЭМК), лабораторных и радиологических информационных систем, персональных медицинских устройств и телемедицинских технологий обеспечивает постоянное создание и накопление в здравоохранении так называемых больших данных (*англ.* big data), объём которых удваивается каждый год [2]. Если в 2013 г. произведено 153 экзабайта, то в 2020 г. — уже порядка 2314 экзабайта, что означает общий темп роста не менее 48% ежегодно. Ожидается, что мировой рынок больших данных в здравоохранении уже в 2023 г. достигнет \$9,5 млрд [3].

Развитие искусственного интеллекта (ИИ) и технологий анализа больших данных позволяет создавать новые программные продукты и сервисы, являющиеся основой цифровой трансформации как лечебно-диагностических процессов в медицинских организациях, так и системы здравоохранения в целом. Более того, применение больших данных и ИИ приводит к изменению уже существующих направлений в работе фармацевтической отрасли, включая исследования реальной клинической практики (*англ.* Real-World Data; RWD) [4].

Одним из основных источников данных реальной клинической практики (ДРКП) являются системы ведения ЭМК [5]. Правильная работа с этим источником, включая централизованный сбор сведений из часто децентрализованных медицинских информационных систем (МИС), очистку и подготовку, извлечение информации с помощью ИИ и других технологий, позволяет оценивать распространённость заболеваний и факторов риска [6].

В настоящей статье мы рассмотрим основные процессы извлечения ДРКП из ЭМК и обеспечения качества этих процессов.

**Общие сведения об информатизации здравоохранения РФ**

Исследования и разработки в сфере применения различных информационных технологий в здравоохранении нашей страны начались ещё в середине 60-х гг. прошлого века. Создание первых программных средств в основном проводилось в ведущих научных школах, НИИ и медицинских вузах бывшего СССР. Вначале создаваемое программное обеспечение (ПО) было предназначено для автоматизи-

рованного формирования статистических отчётов и работы бухгалтерии, затем постепенно информатизация стала внедряться в лечебно-диагностический процесс, начиная с учёта поступающих пациентов, ведения врачебных записей и информационной поддержки [7].

Создание коммерческого рынка МИС произошло в РФ на рубеже конца 90-х — начала 2000-х гг., когда появлялись отдельные программы для диагностики и первые МИС, позволявшие вести некое подобие ЭМК. К середине 2000-х гг. в среде практического здравоохранения в целом сформировалось понимание, что информационные технологии действительно могут стать эффективным инструментом для развития медицины, но в то же время уровень их использования в медицинских организациях (МО) был невысок [8].

Ввиду отсутствия финансирования и регулирования со стороны государства проекты информатизации, как правило, запускались по инициативе руководителей, интересовавшихся новыми технологиями. Чаще всего первые компьютеры появлялись в отделениях статистики и бухгалтерии для автоматизации управленческой деятельности, компьютерная техника и ПО приобретались на собственные средства МО. Профессиональная разработка первых отечественных развиваемых МИС велась, в большинстве случаев, небольшими (20-30 человек) частными компаниями, работавшими по заказу ограниченного круга МО. В значительной степени эта работа представляла собой создание плохо тиражируемых заказных систем, ориентированных на специфику работы заказчиков [9]. Тем не менее на рынке появлялись всё новые компании-разработчики, пик числа которых, по данным интернет-каталога «Медицинские информационные технологии» Ассоциации развития медицинских информационных технологий [10], пришёлся на 2012 г. (рис. 1).

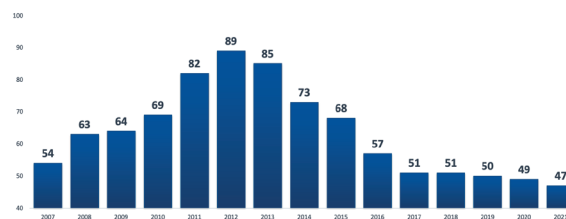


Рис. 1. Динамика числа компаний — разработчиков МИС МО в РФ 2007-2021 гг., по данным Ассоциации развития медицинских информационных технологий

По мнению разработчиков и пользователей МИС в 2003-2008 гг., важнейшим стимулом развития отрасли должно было стать государственное регулирование [11]. Отвечая на этот отраслевой вызов, в 2008 г. Министерство здравоохранения и социального развития РФ начало подготовку к запуску федерального проекта по масштабной информатизации — созданию Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ), фактический старт которого пришёлся на 2011 г.

За 2011-2021 гг. в здравоохранении РФ реализовано несколько этапов и государственных программ, начиная с «Базовой информатизации» в рамках проекта ЕГИСЗ в 2011-2012 гг. и заканчивая запуском в 2019 г. федерального проекта «Создание единого цифрового контура в сфере здравоохранения на основе ЕГИСЗ». Постоянное развитие необходимой инфраструктуры, закупки и внедрение различного ПО обеспечили рост числа МО, внедривших МИС, включая ведение ЭМК. Если в 2007 г. доля таких МО составляла 3,9%, в 2009-м — 10,6%, а в 2011 г. — 15% [12], то уже в 2012 г. она увеличилась до 36,4%. По итогам работы за 2021 г. этот показатель достиг значения в 91% (рис. 2).

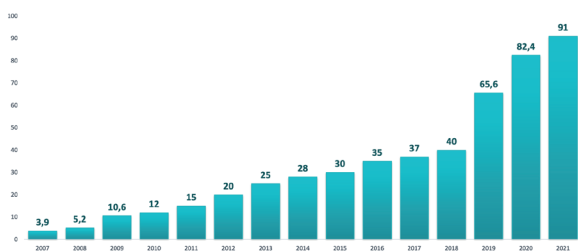


Рис. 2. Динамика доли государственных и муниципальных МО в РФ, внедривших МИС МО

В 2017-2018 гг. при активном участии Министерства здравоохранения РФ в целом определена и утверждена действующая на данный момент система нормативно-правового и технического регулирования. Её основу составляет статья 91 «Информационное обеспечение в сфере здравоохранения» Федерального закона №323-ФЗ, введённая Федеральным законом «О внесении изменений в отдельные законодательные акты РФ по вопросам применения информационных технологий в сфере охраны здоровья» №242-ФЗ от 29.07.2017 г. [13].

В настоящее время всё информационное обеспечение в сфере здравоохранения разделяется на 2 крупных блока: «Информационные системы в сфере здравоохранения» и «Иные информационные системы». К первому блоку относятся программные продукты, созданные по заказу государственных организаций. В соответствии с действующим законодательством они классифицируются по 3 основным уровням (рис. 3):

- **федеральные государственные информационные системы (ГИС)** в сфере здравоохранения, включая Единую государственную информационную систему в сфере здравоохранения (ЕГИСЗ) и Государственную информационную систему обязательного медицинского страхования (ГИС ОМС);
- **государственные информационные системы** в сфере здравоохранения (ГИСЗ) субъектов РФ;
- **учрежденные информационные системы**, представленные медицинскими информационными системами медицинской организации (МИС МО) и информационными системами фармацевтических организаций (ИС ФО).

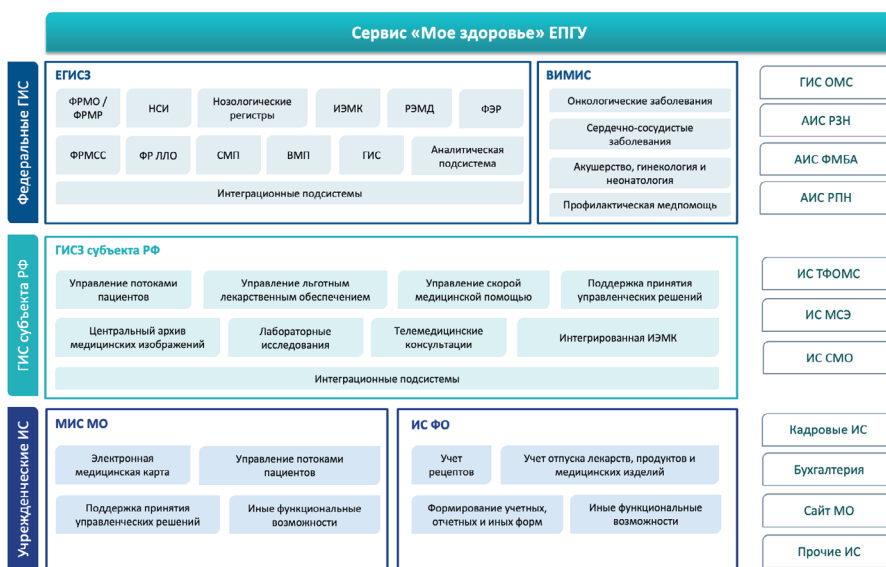


Рис. 3. Функциональная схема информационных систем в сфере здравоохранения в РФ

Все остальные программные продукты, предназначенные для применения в сфере здравоохранения, но разрабатываемые и выводимые на рынок частными компаниями, относятся к так называемым «Иным информационным системам».

Требования к структуре, функциям, порядку и срокам обмена информацией между информационными системами в сфере здравоохранения, включая ЕГИСЗ, ГИС субъектов РФ и МИС МО, определены постановлением Правительства № 140 от 09.02.2022 № 140 «О единой государственной информационной системе в сфере здравоохранения», которое пришло на смену выпущенному в 2018 г. постановлению Правительства № 555. Требования к иным информационным системам, включая требования по защите информации и порядку подключения «Иных ИС» к ЕГИСЗ и другим информационным системам в сфере здравоохранения, определены постановлением Правительства № 447 от 12.04.2018 «О порядке взаимодействия государственных и негосударственных информационных систем в сфере здравоохранения».

В настоящее время нормативное регулирование информатизации здравоохранения включает свыше 30 постановлений и распоряжений Правительства и свыше 40 приказов Министерства здравоохранения, при этом процесс совершенствования нормативного регулирования не останавливается.

#### **Электронные медицинские карты: определение, распространённость, применимое законодательство**

Проблематика определения и утверждения терминологии в сфере ЭМК существует в нашей стране как минимум с начала 2000-х годов, когда были предприняты первые попытки предложить единые определения, в т.ч. используя для этого перевод и адаптацию выпущенных международных стандартов в сфере цифрового здравоохранения. В 2006 г. национальным стандартом ГОСТ Р 52636-2006 был введён термин «Электронная история болезни» (ЭИБ) [14], под которым подразумевалась любая электронная медицинская документация. В настоящее время этот термин вышел из общепринятого применения, поскольку часто «история болезни» ассоциировалась с госпитальным этапом. В 2008 г. был утверждён ГОСТ Р ИСО/ТС 18308-2008 «Информатизация здоровья. Требования к архитектуре электронного учёта здоровья», в котором был предложен термин «электронный учёт здоровья», являющийся некорректным переводом термина *electronic health record* (EHR), хотя именно международный термин EHR ближе всего к термину «ЭМК» [15]. В 2009 г. в работе [16] был представлен обзор различных вариантов терминов и сформулированы предложения по их определению.

В 2013 г. ведущими отраслевыми экспертами был разработан комплекс терминов и определений по электронной медицинской карте, представленный в работе [17]. Данные разработки легли в основу проекта национальных стандартов, который включал ГОСТы «Электронная медицинская карта. Основные принципы, термины и определения», «Электронная медицинская карта, используемая в медицинской организации» и «Интегрированная электронная медицинская карта». Данные документы были одобрены Экспертным советом Министерства здравоохранения России по вопросам использования ИКТ в здравоохранении 10.10.2015 года. Однако затем среди экспертов возник спор о том, каким способом следует регламентировать электронный документооборот (ЭДО), и в частности ЭМК, — с помощью стандартов (добровольного применения) или с помощью приказов Министерства здравоохранения (обязательных к применению). В результате проекты ГОСТ так и не были утверждены.

Таким образом, на сегодняшний день отсутствует нормативное утверждение термина «электронная медицинская карта». ГОСТ Р 52636-2006 по-прежнему остаётся единственным действующим документом, описывающим процессы организации электронного документооборота, связанного с ЭМК.

В этой связи мы пользуемся работой [17], которой предусмотрены следующие понятия:

- **Персональная медицинская запись (ПМЗ)** — любая запись, имеющая отношение к здоровью конкретного человека и выполненная конкретным лицом. ПМЗ — это первичная структурная единица информации о здоровье субъекта, характеризующаяся конкретным автором, отвечающим за содержимое этой записи, конкретным контекстом и моментом выполнения этой записи.

Примечание 1. Это определение несколько расширено по сравнению с ГОСТ Р 52636-2006 за счёт записей о здоровье, которые могут быть сделаны самим пациентом или его доверенными лицами (например, родителями).

Примечание 2. Информация, имеющая отношение к состоянию здоровья, может быть передана по электронным каналам связи непосредственно с медицинского устройства, однако такая запись должна быть подтверждена человеком, отвечающим за организацию измерения, произведённого с помощью данного устройства.

- **Электронная персональная медицинская запись (ЭПМЗ)** — любая персональная медицинская запись, размещённая на электронном

носителе. ЭПМЗ привязана к конкретному электронному хранилищу и характеризуется определённым жизненным циклом в этом хранилище.

- **Электронная медицинская карта (ЭМК)** — совокупность электронных персональных медицинских записей (ЭПМЗ), относящихся к одному человеку, собираемых, хранящихся и используемых в медицинской организации.

Примечание 1. Термин «ЭМК» является аналогом международного термина Electronic Medical Record (EMR).

Примечание 2. Термин «ЭМК» предполагает объединение всей информации (всех ЭПМЗ) о пациенте, имеющейся в данной медицинской организации в электронной форме. При этом ЭПМЗ внутри ЭМК могут быть дополнительно объединены в группы, относящиеся, например, к конкретному законченному случаю заболевания (в амбулаторной практике) или к конкретной госпитализации (при стационарном лечении). Некоторые ЭПМЗ могут не входить ни в одну из групп и не относиться ни к какой конкретной госпитализации или законченному случаю заболевания.

- **Интегрированная электронная медицинская карта (ИЭМК)** — совокупность электронных персональных медицинских записей (ЭПМЗ), относящихся к одному человеку, собираемых, передаваемых и используемых несколькими медицинскими организациями. Входящие в ИЭМК ЭПМЗ могут храниться как централизованно, так и распределённо (в различных МО). При распределённом хранении доступ к отдельным ЭПМЗ, входящим в ИЭМК, осуществляется через централизованный индекс, содержащий информацию о месте хранения и способе доступа к каждой ЭПМЗ. ИЭМК может быть создана группой МО или органом управления здравоохранения.

Примечание: Термин «ИЭМК» является аналогом международного термина Electronic Health Record (EHR).

ИЭМК — это инструмент интеграции медицинских данных, собираемых из различных источников, который может использоваться на различных уровнях. Сегодня этот термин чаще всего используется для региональных (ГИС субъекта РФ) и федеральной (ЕГИСЗ) систем, но он также может использоваться и для сетей клиник или ведомственных сетей, использующих различные МИС [18].

- **Персональная электронная медицинская карта (ПЭМК)** — совокупность электронных

персональных медицинских записей (ЭПМЗ), поступивших из различных источников и относящихся к одному человеку, который и осуществляет их сбор, управление ими, а также определяет права доступа к ним. ПЭМК относится к документам личного хранения и может храниться её субъектом на собственных электронных носителях (личном компьютере, устройствах флэш-памяти и др.) или в специализированных хранилищах, доступных через информационно-телекоммуникационную сеть Интернет.

Примечание 1. Термин «ПЭМК» является аналогом международного термина Personal Health Record (PHR).

ПЭМК предоставляет пациенту и его доверенным лицам возможность вносить информацию о состоянии собственного здоровья, физиологических параметрах своего организма и иную информацию, связанную с собственным здоровьем. Ведение ПЭМК обеспечивает большую приверженность и вовлечённость пациента в лечебный процесс, является эффективным средством поддержания здорового образа жизни, повышает вовлечённость человека в заботу о собственном здоровье и приверженность проводимому лечению. В той или иной степени к ПЭМК сегодня относятся различные классы «личных кабинетов пациентов», создаваемых на различных уровнях от конкретных медицинских организаций до федерального сервиса «Моё здоровье» на едином портале госуслуг.

Приведённая выше терминология стала общепринятой и широко используется в различных, в том числе нормативных, документах [19, 20]. В соответствии с [15] первичными и основными целями ведения ЭМК являются:

- сбор и хранение в электронном виде максимально доступного объёма информации о здоровье конкретного пациента;
- оперативное предоставление доступа к этой информации уполномоченным медицинским работникам, самому пациенту и его доверенным лицам в максимально удобной и доступной для конкретного пользователя форме;
- построение на базе этой информации специализированных электронных сервисов, ориентированных как на медицинский персонал, так и на самого пациента и обеспечивающих увеличение безопасности и качества медицинского обслуживания, а также повышение качества жизни и здоровья пациентов.

Таким образом, понятие ЭМК тесно связано с комплексом задач, охватывающих документирование процессов диагностики и лечения конкретного пациента с помощью информационных технологий, а также процессов диспансеризации, ведения здорового образа жизни и любой другой информации, связанной со здоровьем конкретного индивида. Информация, собираемая в ЭМК, служит, в первую очередь, для обеспечения непрерывности, преемственности и качества лечения.

Согласно [19], ведение ЭМК в МИС МО включает:

- сбор, систематизацию и обработку сведений о лицах, которым оказывается медицинская помощь, а также о лицах, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования;
- назначение диагностических и лабораторных исследований, формирование направления на диагностические и лабораторные исследования;
- получение и выдачу результатов диагностических и лабораторных исследований, медицинских заключений и (или) ссылок на изображения из архивов медицинских изображений;
- учёт временной нетрудоспособности;
- ведение индивидуальных программы реабилитации и реабилитации;
- формирование рецептов на лекарственные препараты и медицинские изделия;
- выдачу отражающих состояние здоровья пациента медицинских документов (их копий), справок и выписок.

Вместе с этим ведение ЭМК, согласно [15], предполагает также ряд вторичных целей, которые могут быть обеспечены в соответствии с требованиями и возможностями конкретных МО, органов управления здравоохранения или провайдеров различных услуг в сфере цифрового здравоохранения. К ним относятся:

- учёт деятельности и автоматизированное построение аналитической и финансовой отчетности МО на базе первичной медицинской информации, получаемой из ЭМК;
- управление МО или здравоохранением региона, а также планирование и выработка политики в отношении медицинских организаций и здравоохранения в целом;
- контроль качества и обоснованности проведенного лечения, юридическое подтверждение проведенного лечения;
- проведение научных и клинических исследований на базе анализа обезличенных данных,

извлечённых из ЭМК, в т. ч. с использованием технологий ИИ;

- использование обезличенных данных из ЭМК для обучения студентов медицинских специальностей, врачей и пациентов, а также для машинного обучения с целью создания новых продуктов в сфере ИИ;
- иные, определённые законодательством, функции, связанные с обеспечением общественного здоровья и безопасности.

Таким образом, хотя стандарт [15] и предусматривает использование ЭМК как источника данных реальной клинической практики, важно подчеркнуть, что это является вторичной целью ведения ЭМК, что обеспечивает особенности и недостатки ЭМК как источника ДРКП, которые мы рассмотрим далее.

Важно подчеркнуть, что за те 15 лет, которые прошли с начала обсуждения темы ЭМК, не только значительно выросли объёмы медицинской информации, собираемой в электронном виде, но и значительно усложнилась структура источников этой информации. Если ранее основным источником данных в ЭМК были медицинские записи, формируемые медработниками внутри единой МИС, то сегодня к этому добавились:

- данные, вносимые самим пациентом с использованием телемедицинских технологий и дистанционного мониторинга (сведения о самочувствии, состоянии, измерениях физиологических параметров, приёме лекарственных препаратов и др.);
- данные от различных медицинских приборов, используемых пациентом в домашних условиях;
- данные, полученные из различных коммерческих медорганизаций (в первую очередь — клинических лабораторий) и предоставляемые самим пациентом;
- данные об образе жизни, которые могут быть получены из различных немедицинских источников (социальных сетей, сотовых операторов, торговых сетей, фитнес-центров и др.).

Эти немедицинские данные также являются ценнейшим и перспективным ресурсом для научного и медицинского анализа, что требует объединения их в единую ЭМК пациента.

### Электронные медицинские карты как источник данных реальной клинической практики

В зарубежной литературе представлено множество примеров использования ЭМК в качестве источника ДРКП. Так, *Hernandez-Boussard et al.* (2019) определили, являются ли данные ЭМК до-

статочными для формирования достоверных клинических утверждений и принятия соответствующих решений в рамках медицинской помощи пациентам с сердечно-сосудистыми заболеваниями. На основании анализа полученных 10840 записей авторы показали, что точность результатов на 98,3% соответствовала данным проведённых ранее рандомизированных клинических исследований (РКИ) [21].

Схожее исследование провели *Kibbelaar et al.* (2017) в рамках нидерландского проекта NemoBase, целью которого являлось обогащение результатов РКИ данными из ЭМК и формирование на основе анализа клинических рекомендаций для пациентов с онкогематологической патологией [22].

Исследовательский коллектив *Moja et al.* (2016) в рамках проекта ONCO-CODES занимался разработкой системы поддержки принятия решений для онкологов, основанной на анализе данных ЭМК, формируемых на этапе оказания первичной медико-санитарной помощи. Авторам удалось доказать эффективность и безопасность разработанной системы [23].

В рамках исследования *Griffith et al.* (2019) изучали возможность использования данных ЭМК пациентов с мелкоклеточным раком лёгкого для формирования прогноза излечения пациентов, используя критерии существующих клинических рекомендаций. Авторы установили, что такой подход может быть оправдан при использовании комплексно как клинических данных, так и результатов объективных инструментальных исследований [24].

Использование ЭМК в качестве источника ДРКП, в т. ч. с точки зрения машинного обучения и научных исследований, подвергается критике в силу следующих недостатков:

1. Низкое качество и удобство интерфейса ЭМК.
2. Применение разработчиками функций повторного использования (копирования) однажды внесённых данных в новые записи [25].
3. Децентрализованный характер систем ведения ЭМК, многие из которых созданы на устаревших технологиях и используют локально устанавливаемые в МО сервера без возможности ведения единой общей БД.
4. Отсутствие единой нормативно-справочной информации для кодирования записей в электронных медицинских документах [26].

5. Пропуски данных и некачественное заполнение экранных форм пользователями [27].
6. Около 80% информации ЭМК представлено в виде неструктурированных записей, в том числе хранимых в виде обычных текстовых документов [28, 29, 30, 31].

Причинами отмеченных недостатков являются:

1. Низкий уровень заинтересованности компаний — разработчиков ЭМК в повышении качества данных и удобства интерфейса, т. к. их выручка в основном зависит от размера клиентской базы и реального уровня использования ЭМК в медицинской организации, а не качества собираемой в ЭМК информации.
2. Отсутствие законодательных и иных программ мотивации повышать качество и полноту сведений, формируемых в ЭМК.
3. Высокие затраты времени на заполнение подробных экранных форм электронных медицинских документов, что делает данный способ менее популярным среди разработчиков по сравнению с использованием шаблонов и свободно заполняемых полей.

Тем не менее, несмотря на отмеченные проблемы, ЭМК являются одним из самых важных источников ДРКП. При правильном использовании ЭМК можно получить огромное количество информации, направленное на решение различных задач в системе здравоохранения. С целью уменьшения рисков, вызываемых существующими на сегодня проблемами в использовании ЭМК, важно обеспечивать корректное извлечение данных.

#### Извлечение данных из электронных медицинских карт: общая схема

Внедрение продуманной стратегии, включающей ряд взаимосвязанных последовательных этапов обработки данных, и её последующая качественная реализация — ключевой фактор успеха при получении ДРКП из ЭМК (рис. 4). Без подобной специальной подготовки информация из ЭМК будет непригодной для задач машинной обработки, включая формирование наборов данных для целей научных исследований, создания систем искусственного интеллекта (ИИ), статистического анализа и т. д.



Рис. 4. Схема формирования наборов данных реальной клинической практики из систем ведения электронных медицинских карт

Рассмотрим подробнее каждый из процессов.

### Накопление медицинских записей в системах ведения ЭМК

Важнейшей проблемой накопления данных в ЭМК является значительное нежелание врачей работать с МИС. Эта проблема существует во всём мире. Более того, неудобство интерфейсов ЭМК и повышенная нагрузка на врачей в связи с необходимостью использования ЭМК являются одной из основных причин их эмоционального «выгорания» [32].

Не секрет, что даже в медорганизациях с высоким уровнем внедрения МИС все карты пациентов ведутся в бумажном формате, хотя и с компьютерными распечатками. Конечно, у врача появляется повод для недовольства: нужно внести запись и в компьютер, и в бумажную карту, искать информацию либо в компьютере, либо в бумажной карте. Такие действия порождают неудобство и усиливают нагрузку. Основной причиной такого дублирования до последнего времени было отсутствие легитимного статуса электронного документооборота (ЭДО). Однако с февраля 2021 года эта проблема была решена с выходом приказа Министерства здравоохранения № 947н [20], разрешившего использовать ЭДО без дублирования на бумажном носителе и прояснившего основные проблемы такого безбумажного использования. Однако год, прошедший с момента вступления в силу приказа № 947н, показал, что сама по себе юридическая возможность ЭДО ещё не достаточна. Такой документооборот необходимо активно стимулировать.

Сегодня единственным стимулом для ЭДО являются нормативные требования, обязывающие передавать определённые медицинские документы в ЕГИСЗ. Это даже включено в лицензионные требования для медицинских организаций. Такое стимулирование приводит к тому, что ЭМК состоит в основном из формально требуемых документов — статталонов, выписок, реестров счетов и других документов, включающих не очень большой объём именно медицинских данных о пациен-

те. Такое насыщение ЭМК формально требуемыми документами значительно снижает ценность ЭМК для аналитической и научной работы.

Необходимо вырабатывать позитивные стимулы и программы мотивации для насыщения ЭМК именно медицинскими документами, содержащими клинически ценную информацию. Для этого ЭМК должна стать полезной врачам в их повседневной деятельности. В частности, работой [32] отмечены следующие ценности:

- способность передавать данные коллегам в электронном виде (это требуется 70% врачей);
- обеспечивать доступ к ЭМК из дома (76%);
- делиться результатами исследований с пациентами (48%).

С учётом требований законодательства о защите персональных данных, система ведения ЭМК должна обезличить накопленные персональные данные пациента и передать их в централизованную систему озера данных (*англ.* data lake). Процесс обезличивания должен осуществляться строго в защищённой инфраструктуре оператора (медицинской организации или ведомственного ЦОДа в случае централизованной МИС). Обезличивание необходимо реализовывать по единым техническим принципам во всех системах ведения ЭМК. Подобный подход обеспечит в последующем соединение различных эпизодов обращения пациента за медицинской помощью, полученных из разных МО и систем ведения ЭМК, в единую интегрированную ЭМК пациента (ИЭМК).

### Централизованный сбор «сырых данных» из ЭМК в озеро данных

Главной задачей озера данных является централизованное накопление как можно большего количества любой необработанной информации по пациенту — так называемых «сырых данных» (*англ.* raw data). Чем больше «сырых данных» будет накоплено, тем лучше. Очень важно, чтобы в озеро данных были загружены все эпизоды обращения па-



циента за медицинской помощью, включая случаи амбулаторного и стационарного лечения по всем поводам, данные об обращении в станцию скорой медицинской помощи, данные диспансеризации, реабилитации и т. д.

Ценность озера данных будет намного больше, если, кроме данных из обезличенных ЭМК, в него можно будет загрузить данные самих пациентов,

включая сведения из аккаунтов социальных сетей, данные носимых устройств, справочную информацию об условиях жизни пациента, включая характеристики места жительства, окружающей среды, данные об условиях, вредных и опасных факторах места труда, характеристики системы здравоохранения в районе постоянного проживания пациента и т. д. (рис. 5).



Рис. 5. Состав данных, которые необходимо собирать в озере данных

С технической точки зрения существует несколько ключевых требований к озеру данных:

Исключение любой возможности воссоздания из принятых данных идентификаторов пациента, включая фамилию, имя и отчество, номера документов, удостоверяющих личность, полисов страхования и СНИЛС, номера телефонов, адреса электронной почты и т. д.

Приём информации из систем ведения ЭМК в виде структурированных электронных медицинских документов (СЭМД), основанных на стандарте HL7 CDA 3.0.

Автоматическое объединение различных эпизодов течения заболеваний пациента, полученных из несовместимых друг с другом и разнородных источников информации, вокруг пациента.

### Извлечение данных из неструктурированных записей

Для извлечения из неструктурированных медицинских записей возможно применение различных технологий, в т. ч. natural language processing (NLP). С технической точки зрения задачей данного этапа является анализ сохранённых, в т. ч. неструктурированных, записей с целью выделения из них отдель-

ных структурированных признаков (*англ.* features). Различные типы признаков представлены в табл. 1.

Таблица 1. Виды признаков, извлекаемых из «сырых данных»	
Признак	Пример
Бинарный	Курение, приём антигипертензивных препаратов и т. д.
Числовой	Температура, ЧСС, АД, рост, масса тела, значения лабораторных тестов и т. д.
Дата	Дата рождения, дата события, дата смерти и т. д.
Текстовый	Симптом, место работы и т. д.
Код из справочника	Код МКБ, пол и т. д.

Для этого разрабатываются модели машинного обучения, которые могут находить заранее определённые признаки в получаемых на вход текстовых блоках и возвращать структурированную информацию, которая уже затем будет записана в базу данных и будет пригодна для дальнейшей обработки (рис. 6).

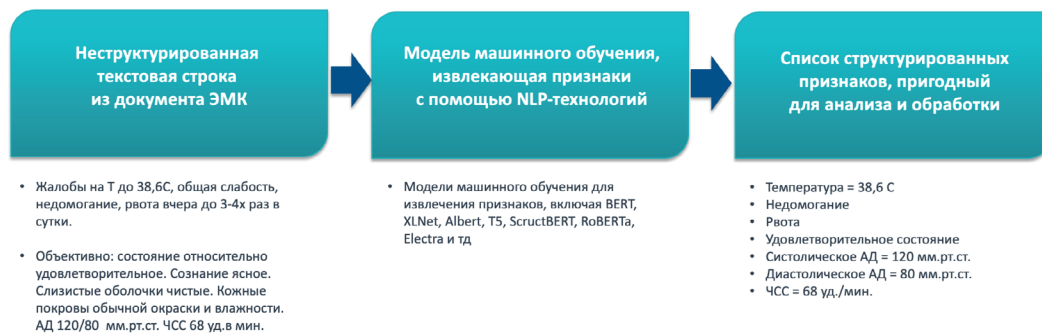


Рис. 6. Пример извлечения структурированных признаков из текстовой записи с помощью моделей NLP

Конечно, получить 100%-но правильное и всестороннее решение, которое бы умело понимать любой текст и извлекать любые признаки, будет очень сложно. Однако работая постоянно над совершенствованием NLP-моделей, разработчики могут получить достаточно мощный инструмент для обработки и извлечения данных практически из любой медицинской записи.

### Формирование цифрового профиля

После извлечения признаков из всех накопленных в озере данных документов возможно формирование так называемого цифрового профиля пациента. Иногда в литературе комплексно собранные по пациенту данные называют цифровым двойником пациента (*англ.* digital twin) [33], что, на наш взгляд, является не совсем корректным, поскольку цифровые двойники должны позволять проводить полноценное моделирование изменения объекта в различных условиях, что без комплекса математических моделей здоровья пациента и его заболеваний невозможно.

Чем больше разнообразных данных о пациенте накоплено в его цифровом профиле, тем больше его

ценность с точки зрения проведения исследований ДРКП, а также выполнения задач исследований и разработок в сфере ИИ [34]. Состав таких данных представлен на рис. 7.

Обязательной задачей данного процесса является форматно-логический контроль каждого извлечённого признака. Для этого в справочнике соответствующей информационной системы должны храниться границы допустимых значений для соответствующей единицы измерения. В случае если NLP-модели извлекли какое-то значение признака, которое не укладывается в допустимые границы, информационная система должна пометить это значение как некорректное, чтобы исключить дальнейшую обработку. Не рекомендуется удалять ошибочные записи из базы данных системы. Они необходимы для последующего анализа причин появления некачественных данных, определения распространённости их в разрезе различных МО или систем ведения ЭМК. Такой анализ может представлять весомую ценность для последующих мер повышения качества ведения записей в ЭМК.

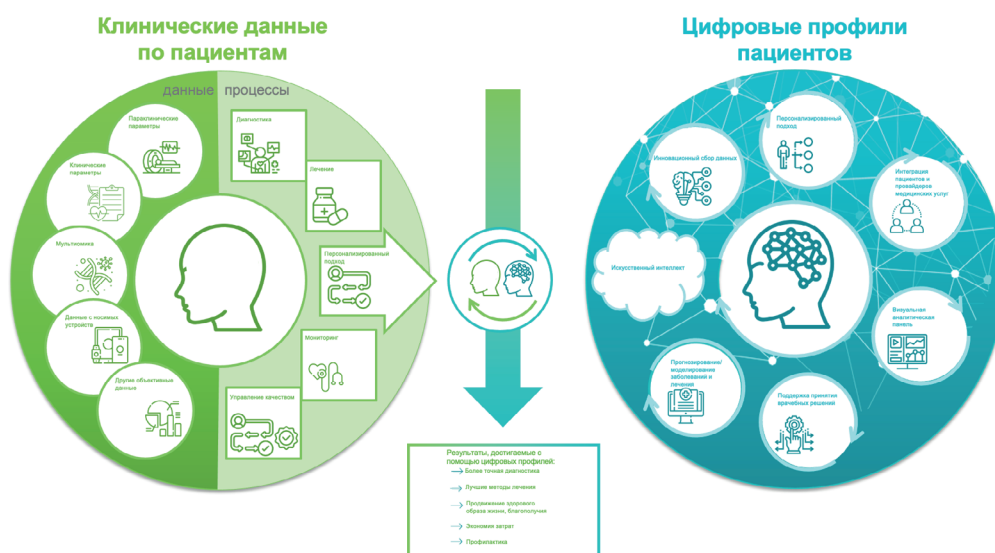


Рис. 7. Концепция цифрового профиля пациента по [34]

Также на данном этапе осуществляется комплексная интерпретация всех подготовленных данных: проверка на дубли и противоречивость данных, расчёт вторичных признаков (например, ИМТ из извлечённых данных о массе тела и росте), выявление и фиксация финальных сведений о факторах риска, зарегистрированных заболеваниях, прогнозах и т. д.

### Формирование наборов данных по запросу

Сформированные цифровые профили пациентов полностью хранятся в базе данных в структурированном виде, пригодном для выполнения запросов, аналитической обработки и формирования наборов данных на основании определённых критериев. Сформированные наборы могут быть выгружены в виде машиночитаемых файлов (например, в формате CSV) и использоваться для дальнейшего анализа и обработки, в том числе в научных исследованиях, машинном обучении и т. д.

### Заключение

В российском здравоохранении вот уже более 10 лет реализуется ряд крупных государственных проектов по цифровой трансформации, которые обеспечили накопление в медицинских организациях архивов ЭМК. Развитие технологий обработки больших данных, таких как ИИ, позволяет извлекать из накапливаемых ЭМК ценную клиническую информацию и использовать её как для создания инновационных продуктов, таких как системы поддержки принятия врачебных решений, так и для проведения исследований реальной клинической практики.

В настоящее время ЭМК являются одним из самых важных источников ДРКП. Внедрение продуманной стратегии, включающей ряд взаимосвязанных последовательных этапов обработки данных, и её последующая качественная реализация, в свою очередь, является ключевым фактором успеха при получении ДРКП из ЭМК. Мы выделили 5 ключевых этапов, позволяющих получить чёткие наборы данных для достижения конкретных задач:

- 1) накопление медицинских записей в системах ведения ЭМК;
- 2) централизованный сбор обезличенных медицинских записей из ЭМК в т. н. озере данных;
- 3) извлечение признаков из неструктурированных медицинских документов;
- 4) формирование цифрового профиля пациента;
- 5) формирование наборов данных по запросу.

Каждый из представленных этапов содержит ряд требований и последовательные процессы их проведения.

Для того чтобы обеспечить доверие к разработкам и выводам, формируемым на основе анализа

ДРКП, полученных из ЭМК, необходимо обеспечивать качество выполнения всех этапов и процессов формирования наборов ДРКП.

### ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

**Благодарность.** Авторы выражают благодарность Т. А. Гольдиной, руководителю направления по данным рутинной практики и научной коммуникации, АО «Санофи Россия», за оказанную помощь при написании настоящей статьи.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Финансирование:** статья опубликована без финансовой поддержки.

### СВЕДЕНИЯ ОБ АВТОРАХ

**Гусев Александр Владимирович** — кандидат технических наук, директор по развитию бизнеса ООО «К-Скай», Петрозаводск, Россия

*Автор, ответственный за переписку*

email: [agusev@webiomed.ai](mailto:agusev@webiomed.ai)

<https://orcid.org/0000-0002-7380-8460>

**Зингерман Борис Валентинович** — директор ООО «ТелеПат», Москва, Россия

email: [boriszing@gmail.com](mailto:boriszing@gmail.com)

<https://orcid.org/0000-0002-1855-1834>

**Тюфилин Денис Сергеевич** — начальник управления стратегического развития здравоохранения ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, Москва, Россия

email: [tyufilinds@mednet.ru](mailto:tyufilinds@mednet.ru)

<https://orcid.org/0000-0002-9174-6419>

**Зинченко Виктория Валерьевна** — начальник сектора клинических и технических испытаний Государственного бюджетного учреждения здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы», Москва, Россия

email: [v.zinchenko@npcmr.ru](mailto:v.zinchenko@npcmr.ru)

<https://orcid.org/0000-0002-2307-725X>

### Литература

1. Digital Health Market Size By Technology, Telehealth, mHealth, Apps, Health Analytics, Digital Health System (EHR), By Component, Industry Analysis Report, Regional Outlook, Application Potential, Price Trends, Competitive Market Share & Forecast, 2020-2026. <https://www.gminsights.com/industry-analysis/digital-health-market>.

2. Harnessing the Power of Data in Health: Stanford Medicine 2017 Health Trends Report. <https://med.stanford.edu/content/dam/sm/sm-news/documents/StanfordMedicineHealthTrendsWhitePaper2017.pdf>.
3. 2020 Global Health Care Outlook. <https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html>.
4. Гольдина Т.А., Колбин А.С., Белоусов Д.Ю., Боровская В.Г. Обзор исследований реальной клинической практики. *Качественная клиническая практика*. 2021; (1):56-63. <https://doi.org/10.37489/2588-0519-2021-1-56-63>.
5. Kim HS, Lee S, & Kim JH. Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. *Journal of Korean medical science*. 2018;33 (34):e213. <https://doi.org/10.3346/jkms.2018.33.e213>.
6. Ka-Shing Cheung. Application of Big Data analysis in gastrointestinal research/Ka-Shing Cheung, Wai K Leung, Wai-Kay Seto. 2019. <https://pubmed.ncbi.nlm.nih.gov/31293336/>.
7. Гаспарян С.А., Пашкина Е.С. Страницы истории информатизации здравоохранения России. — М., 2002. — 304 с.
8. Фролов С.В., Маковеев С.Н., Семёнова С.В., Фареа С.Г. Современные тенденции развития рынка медицинских информационных систем. *Вестник Тамбовского государственного технического университета*. 2010;16 (2):266-72.
9. Гусев А.В., Романов Ф.А., Дуданов И.П. Обзор медицинских информационных систем на отечественном рынке в 2005 году. *Медицинский академический журнал*. 2005;5 (3):Приложение 7. 72-84.
10. Ассоциация развития медицинских информационных технологий. <https://www.armit.ru/>.
11. Гусев А.В. Обзор рынка комплексных медицинских информационных систем. *Врач и информационные технологии*. 2009; (6):4-17.
12. Гусев А.В. Медицинские информационные системы: состояние, уровень использования и тенденции. *Врач и информационные технологии*. 2011; (3):6-14.
13. Бойко Е.Л. Цифровое здравоохранение. *Вестник Росздрава*. 2018; (3):5-8.
14. ГОСТ Р 52636-2006 «Электронная история болезни. Общие положения». <https://docs.cntd.ru/document/1200048924>.
15. ГОСТ Р ИСО/ТС 18308-2008 «Информатизация здоровья. Требования к архитектуре электронного учёта здоровья». <https://docs.cntd.ru/document/1200067414>.
16. Емелин И.В., Зингерман Б.В., Лебедев Г.С. Проблемы определения ключевых терминов медицинской информатики. Информационно-измерительные и управляющие системы. 2009; (12):15-23.
17. Зингерман Б.В., Шкловский-Корди Н.Е. Электронная медицинская карта и принципы её организации. *Врач и информационные технологии*. 2013; (2):37-58.
18. Зингерман Б.В., Шкловский-Корди Н.Е. Интегрированная электронная медицинская карта: задачи и проблемы. *Врач и информационные технологии*. 2015; (1):24-34.
19. Приказ Министерства здравоохранения РФ от 24.12.2018 № 911н «Об утверждении требований к государственным информационным системам в сфере здравоохранения субъектов Российской Федерации, медицинским информационным системам медицинских организаций и информационным системам фармацевтических организаций».
20. Приказ Министерства здравоохранения РФ от 07.09.2020 № 947н «Об утверждении Порядка организации системы документооборота в сфере охраны здоровья в части ведения медицинской документации в форме электронных документов».
21. Hernandez-Boussard T, Monda KL, Crespo BC, & Riskin D. Real world evidence in cardiovascular medicine: ensuring data validity in electronic health record-based studies. *Journal of the American Medical Informatics Association: JAMIA*. 2019;26 (11):1189-94. <https://doi.org/10.1093/jamia/ocz119>.
22. Kibbelaar RE, Oortgiesen BE, van der Wal-Oost AM, Boslooper K, Coebergh JW, Veeger NJGM, Joosten P, Storm H, van Roon EN, Hoogendoorn M. Bridging the gap between the randomised clinical trial world and the real world by combination of population-based registry and electronic health record data: A case study in haematology. *Eur J Cancer*. 2017 Nov;86:178-85. doi: 10.1016/j.ejca.2017.09.007. Epub 2017 Oct 6. PMID: 28992561. <https://doi.org/10.1016/j.ejca.2017.09.007>.
23. Moja L, Passardi A, Capobussi M, Banzi R, Ruggiero F, Kwag K, Liberati EG, Mangia M, Kunnamo I, Cinquini M, Vespignani R, Colamartini A, Di Iorio V, Massa I, González-Lorenzo M, Bertizzolo L, Nyberg P, Grimshaw J, Bonovas S, Nanni O. Implementing an evidence-based computerized decision support system linked to electronic health records to improve care for cancer patients: the ONCO-CODES study protocol for a randomized controlled trial. *Implement Sci*. 2016 Nov 25;11 (1):153. doi: 10.1186/s13012-016-0514-3. PMID: 27884165; PMCID: PMC5123241. <https://doi.org/10.1186/s13012-016-0514-3>.

24. Griffith SD, Tucker M, Bowser B, Calkins G, Chang CJ, Guardino E, Khozin S, Kraut J, You P, Schrag D, Miksad RA. Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer. *Adv Ther*. 2019 Aug;36 (8):2122-36. doi: 10.1007/s12325-019-00970-1. Epub 2019 May 28. PMID: 31140124; PMCID: PMC6822856. <https://doi.org/10.1007/s12325-019-00970-1>.
25. Wang MD, Khanna R, and Najafi N. Characterizing the Source of Text in Electronic Health Record Progress Notes. *JAMA Internal Medicine*. 2017;177 (8):1212-3.
26. Topol EJ. Editor. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books, 2019.
27. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, Spreafico R, Hafler DA and McKinney EF. From Big Data to Precision Medicine. *Front. Med*. 2019;6:34. doi: 10.3389/fmed.2019.00034
28. Gilmore-Bykovskyi AL, Block LM, Walljasper L, Hill N, Gleason C, Shah MN. Unstructured clinical documentation reflecting cognitive and behavioral dysfunction: toward an EHR-based phenotype for cognitive impairment. *J Am Med Inform Assoc*. 2018 Sep 1;25 (9):1206-12. doi: 10.1093/jamia/ocy070. PMID: 29947805; PMCID: PMC6118865.
29. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J Biomed Health Inform*. 2018 Sep;22 (5):1589-604. doi: 10.1109/JBHI.2017.2767063. Epub 2017 Oct 27. PMID: 29989977; PMCID: PMC6043423.
30. Kong HJ. Managing Unstructured Big Data in Healthcare System. *Healthcare informatics research*. 2019;25 (1):1-2. <https://doi.org/10.4258/hir.2019.25.1.1>.
31. Исследование полноты и структурированности данных медицинских информационных систем Санкт-Петербурга. <https://actcognitive.org/proekty/city-healthcare?>
32. Kroth PJ, Morioka-Douglas N, Veres S, et al. Association of Electronic Health Record Design and Use Factors With Clinician Stress and Burnout. *JAMA Netw Open*. Published online August 16, 2019;2 (8):e199609. doi: 10.1001/jamanetworkopen.2019.9609
33. Björnsson B, Borrebaeck C, Elander N et al. Digital twins to personalize medicine. *Genome Med*. 2020;12:4. <https://doi.org/10.1186/s13073-019-0701-3>.
34. Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K, Ziemssen T. Digital Twins for Multiple Sclerosis. *Front Immunol*. 2021 May 3; 12:669811. doi: 10.3389/fimmu.2021.669811. PMID:34012452;PMCID:PMC8128142.<https://doi.org/10.3389/fimmu.2021.669811>. 