

Возможности больших данных в фармакоэпидемиологии, проблемы использования, юридическое регулирование

Н.В. Орлова^{1,*}, Г.Н. Суворов^{2,3}, Н.Д. Карселадзе⁴, Ю.Ю. Якушев¹, Т.В. Гололобова^{2,5}

¹ ФГАОУ ВО «Российский национальный исследовательский медицинский университет им. Н.И. Пирогова» Министерства здравоохранения Российской Федерации, ул. Островитянова, д. 1, г. Москва, 117513, Россия

² ФГБОУ ДПО «Российская медицинская академия непрерывного профессионального образования» Министерства здравоохранения Российской Федерации, ул. Баррикадная, д. 2/1, стр. 1, г. Москва, 125993, Россия

³ ФГАНУ «Центр социологических исследований», Тверской бульвар, д. 13, стр. 1, г. Москва, 123104, Россия

⁴ Медицинский центр DMEMED аэропорта Домодедово имени М.В. Ломоносова, ст. 1а, Московская область, 142014, Россия

⁵ ФГУП «Государственный научно-исследовательский институт биологического приборостроения» ФМБА России, Волоколамское шоссе, д. 75, к. 1, г. Москва, 125424, Россия

Аннотация

Рост цифровизации в медицине значительно расширил возможности использования больших данных в фармакоэпидемиологии. Использование больших данных позволяет сократить расходы на клинические исследования, увеличить скорость набора и расширить выборку, дает возможность оценить действие лекарственных препаратов у беременных и эффективность при редких заболеваниях. Проанализированы базы данных PubMed, Scopus, Web of Science и Google Scholar за 12 лет, с 2012 по 2023 год, документы Организации Объединенных Наций, Всемирной организации здравоохранения, федеральные законы Российской Федерации в области использования интеллектуального интеллекта, защиты конфиденциальной информации, проведения клинических исследований. Поиск осуществлялся по ключевым словам: «большие данные», «регистры», «фармакоэпидемиология», «персональные данные», «правовое регулирование», «методы защиты». В статье приведены примеры использования больших данных в здравоохранении, включающих от 25 до 50 млн человек. Анализ литературных данных выявил однотипные проблемы: отсутствие единообразия внесения информации, неполная информация, ограничение доступности. В обзоре определены проблемы защиты конфиденциальности информации. Рассмотрены механизмы стандартизации информации, хранения, обработки данных. Представлена международная и российская законодательные базы, регламентирующие проведение клинических исследований с использованием больших данных.

Ключевые слова: большие данные; фармакоэпидемиология; персональные данные; правовое регулирование

Для цитирования: Орлова Н.В., Суворов Г.Н., Карселадзе Н.Д., Якушев Ю.Ю., Гололобова Т.В. Возможности больших данных в фармакоэпидемиологии, проблемы использования, юридическое регулирование. Национальное здравоохранение. 2024; 5 (2): 25–35. <https://doi.org/10.47093/2713-069X.2024.5.2.25-35>

Контактная информация:

* Автор, ответственный за переписку: Орлова Наталья Васильевна. E-mail: vrach315@yandex.ru

Статья поступила в редакцию: 12.04.24

Статья принята к печати: 17.05.24

Дата публикации: 29.07.24

The possibilities of big data in pharmacoepidemiology, problems of use, legal regulation

Natalya V. Orlova^{1,*}, Georgy N. Suvorov^{2,3}, Natalia D. Karseladze⁴, Yuri Yu. Yakushev¹, Tatyana V. Gololobova^{2,5}

¹ N.I. Pirogov Russian National Research Medical University, Ostrovityanova str., 1, Moscow, 117513, Russia

² Russian Medical Academy of Continuing Professional Education, Barricadnaya str., 2/1, building 1, Moscow, 125993, Russia

© Н.В. Орлова, Г.Н. Суворов, Н.Д. Карселадзе, Ю.Ю. Якушев, Т.В. Гололобова, 2024

³ Center for Sociological Research, Tverskoy Boulevard, 13/1, Moscow, 123104, Russia

⁴ Medical Center DMEMED, M.V. Lomonosov Airport Domodedovo, 1a, Moscow region, 142014, Russia

⁵ State Scientific Research Institute of Biological Instrumentation FMBA of Russia, Volokolamsk highway, 75/1, Moscow, 125424, Russia

Abstract

The growth of digitalization in medicine has significantly expanded the possibilities of using big data in pharmacoepidemiology. The use of big data makes it possible to reduce the cost of clinical research, increase the speed of recruitment and expand the sample, makes it possible to evaluate the effect of drugs in pregnant women and effectiveness in rare diseases. The databases PubMed, Scopus, Web of Science and Google Scholar for 12 years from 2012 to 2023, United Nations documents, World Health Organization, Federal laws of the Russian Federation in the field of artificial intelligence, protection of confidential information, and clinical research were analyzed. The search was carried out by keywords: «big data», «registers», «pharmacoepidemiology», «personal data», «legal regulation», «protection methods». The article provides examples of the use of big data in healthcare, including from 25 to 50 million people. The analysis of the literature data revealed the same type of problems – the lack of uniformity in the introduction of information, incomplete information, limited availability. The review identifies the problems of protecting the confidentiality of information. The mechanisms of information standardization, storage, and data processing are considered. The international and Russian legislative framework regulating the conduct of clinical trials using big data is presented.

Keywords: big data; pharmacoepidemiology; personal data; legal regulation

For citation: Orlova N.V., Suvorov G.N., Karseladze N.D., Yakushev Yu.Yu., Gololobova T.V. The possibilities of big data in pharmacoepidemiology, problems of use, legal regulation. National Health Care (Russia). 2024; 5 (2): 25–35. <https://doi.org/10.47093/2713-069X.2024.5.2.25-35>

Contacts:

* Corresponding author: Natalya V. Orlova. E-mail: vrach315@yandex.ru

The article received: 12.04.24

The article approved for publication: 17.05.24

Date of publication: 29.07.24

Список сокращений:

БД – большие данные

ДИ – доверительный интервал

ЗНО – злокачественные новообразования

КИ – клинические исследования

ЛС – лекарственные средства

ОШ – отношение шансов

CNODES – Canadian Network for Observational Drug Effect Studies, Канадская сеть наблюдательных исследований воздействия лекарств

EUROCAT – European Concerted Action on Congenital Anomalies and Twins, Европейская организация согласованных действий по врожденным аномалиям
SEER – The Surveillance, Epidemiology, and End Results, Программа эпиднадзора, эпидемиологии и конечных результатов (США)

TARGET – Therapeutically Applicable Research to Generate Effective Treatments, терапевтически применимые исследования для разработки эффективных методов лечения

Фармакоэпидемиология – это изучение взаимодействия между лекарственными средствами (ЛС) и популяцией человека, изучение в реальных условиях жизни преимуществ и рисков использования ЛС [1]. Регистрации ЛС предшествуют клинические исследования (КИ), основанные на экспериментальном дизайне с рандомизацией участников. К сожалению, выводы КИ ограничиваются пациентами, включенными в них. При проведении КИ не всегда удается достичь целевого размера выборки, например при редких заболеваниях. Как правило, КИ предполагают высокие экономические затраты, например связанные с исследовательским персоналом [2]. Результаты могут иметь ограниченную возможность обобщения из-за строгих критериев отбора, различных условий проведения исследований, нерепрезентативности выборки участников или других причин.

Для расширения возможностей фармакоэпидемиологии требуются новые форматы исследований: 1) прогнозные моделирование и создание более экономичного, быстрого и целенаправленного процесса исследований; 2) применение статистических инструментов и алгоритмов для улучшения дизайна КИ и набора пациентов, в т.ч. их отдельных категорий, тем самым уменьшая количество неудачных испытаний и ускоряя вывод новых ЛС на рынок; 3) анализ КИ и историй болезни пациентов для выявления последующего расширения показаний и выявления побочных эффектов ЛС [3].

Цель исследования – проанализировать возможности использования больших данных в фармакоэпидемиологии. Были проанализированы статьи из наукометрических баз данных PubMed, Scopus, Web of Science и Google Scholar за период с 2012 по 2023

© Natalya V. Orlova, Georgy N. Suvorov, Natalia D. Karseladze, Yuri Yu. Yakushev, Tatyana V. Gololobova, 2024

годы. Поиск осуществлялся по ключевым словам: «большие данные», «регистры», «фармакоэпидемиология», «персональные данные», «правовое регулирование», «методы защиты». Всего просмотрено более 360 источников. Отобраны 38 статей (36 англоязычных и 2 русскоязычные), 84 % были изданы в течение 2018–2023 гг. Проанализированы документы Организации Объединенных Наций, Всемирной организации здравоохранения, федеральные законы Российской Федерации в области использования искусственного интеллекта, защиты конфиденциальной информации, проведения клинических исследований.

Большие данные в здравоохранении

Большие данные (БД) – большие объемы высокоскоростных, сложных и переменных данных, которые требуют передовых методов и технологий для сбора, хранения, распространения, управления и анализа информации.

Большие данные включают в себя такие характеристики, как разнообразие, скорость и, применительно к здравоохранению, достоверность [1]. Рост цифровизации в медицине значительно расширил возможности использования БД. Их анализ в здравоохранении позволяет получать ценную информацию на основе очень больших наборов данных и улучшать результаты при одновременном сокращении затрат. Компьютеризированные медицинские БД все чаще используются для реальных постмаркетинговых (пострегистрационных) наблюдательных исследований [4].

Использование БД обладает рядом преимуществ: позволяет изучать редкие события или воздействия; дает возможность оценить действие ЛС у беременных, детей, больных с психическими заболеваниями, которые чаще всего ограниченно представлены в КИ; а также получить информацию об эффектах использования ЛС в ситуациях, не указанных в инструкции к их использованию. КИ на основе БД дополняют рандомизированные исследования, предоставляя доступ к большой выборке данных. Преимущества использования БД в здравоохранении связаны не только с их объемом, но также с разнообразием типов данных и скорости, с которой ими можно управлять. Применение аналитики, машинного обучения и искусственного интеллекта к БД позволяет выявлять закономерности и корреляции и, следовательно, дает полезную информацию для улучшения оказания медицинской помощи [5].

В сфере здравоохранения к источникам БД относятся электронные медицинские записи, медицинские карты пациентов, результаты медицинских осмотров и обследований, а также другие информационные ресурсы (лаборатории, аптеки, страховые компании, сайты различных поставщиков медицинских услуг и др.). Данные здравоохранения, как правило, неоднородны, недостаточно структурированы, формализованы

и стандартизованы, часто фрагментируются или генерируются в устаревших ИТ-системах в несовместимых форматах.

Для возможности проведения анализа БД необходимо объединение различных ресурсов, для чего информация должна быть обработана и преобразована. Посредством этапов извлечения, преобразования и загрузки данные из различных источников очищаются и подготавливаются. В зависимости от того, являются ли данные структурированными или неструктурированными, в платформу анализа БД можно вводить несколько форматов данных.

Аналитика БД в режиме реального времени является ключевым требованием в здравоохранении, что требует устранения задержки между сбором и обработкой данных [6].

БД в фармакоэпидемиологии ЛС у беременных

Одним из преимуществ использования БД в здравоохранении является расширение информации о безопасности и эффективности применения ЛС у беременных. Использование лекарств во время беременности часто ограничено в связи с невозможностью проведения рандомизированных контролируемых КИ среди этой категории женщин [7]. Изучение тератогенеза, как правило, проводится в исследованиях на животных. Безопасность применения ЛС во время беременности чаще всего можно оценить только после его выхода на рынок [8]. Врожденные аномалии встречаются редко и возникают из разных эмбриональных тканей и на разных сроках беременности [9].

Одной из сильных сторон электронных баз данных здравоохранения является большое количество беременностей, которые они фиксируют, и отчет о препаратах, которые им выписывались.

Для сбора информации широко используются регистры. Например, European Concerted Action on Congenital Anomalies and Twins (EUROCAT), Европейская организация согласованных действий по врожденным аномалиям – сеть популяционных реестров врожденных аномалий, предоставляющая стандартизированную эпидемиологическую информацию о врожденных аномалиях в Европе. В эту базу входят 43 регистра врожденных аномалий в 23 странах, а также данные о воздействии лекарств. В настоящее время европейский исследовательский консорциум EUROmediCAT на основе использования электронных баз данных здравоохранения в сочетании с данными EUROCAT о врожденных аномалиях разрабатывает европейскую систему оценки использования лекарств во время беременности в связи с риском врожденных аномалий. Общее число включенных в анализ EUROmediCAT составило 25 миллионов человек. Данные о воздействии препаратов оценивались на основе выдачи рецептов и отпуска лекарств в аптеках.

Среди проблем использования БД исследователи отметили отсутствие информации о препаратах, выписанных в стационаре и частными врачами. Контроль за выпиской лекарств различался в разных странах, также различия касались информации о пациентках (например, в электронных историях болезни не везде отмечалась дата последней менструации). В странах также различались организации и структуры здравоохранения, ответственные за ведение учета аномалий плода, задержки внутриутробного развития, преждевременных родов, развития в детском возрасте, информации о хронических заболеваниях беременных [10].

Примером эффективности использования БД для изучения применения ЛС у беременных являются исследования, проведенные в Японии. В исследование по оценке взаимосвязи между употреблением антидепрессантов во время беременности и аутизмом у детей были включены 26 925 пар «мать – ребенок». Анализ показал, что распространенность аутизма у детей была значительно выше при использовании любого антидепрессанта, чем при его неиспользовании (отношение шансов (ОШ) 2,32; 95 % доверительный интервал (ДИ) 1,08–4,95; $p = 0,03$). Однако при дальнейшем статистическом анализе, с поправками на факторы (например, наличие депрессии), статистическая значимость была снижена (ОШ 0,76; ДИ 0,27–2,18; $p = 0,61$).

Изучение в исследовании влияния пренатального воздействия антибиотиков на развитие астмы у детей показало, что воздействие антибиотиков в период внутриутробного развития было связано с ранним развитием астмы – до 3 лет (относительный риск 1,18; 95 % ДИ 1,08–1,30) [11].

БД в фармакоэпидемиологии ЛС при редких заболеваниях

Еще одной возможностью КИ с использованием БД является решение проблемы оценки распространенности и эффективности ЛС для лечения редких заболеваний, поскольку эпидемиологические данные о них зачастую трудно оценить, если они не собираются в большой популяции [12]. Примером является исследование, включившее 12 713 человек в возрасте старше 20 лет, у которых впервые была диагностирована акромегалия. Исследование позволило оценить распространенность этого заболевания и применяемые методы лечения. Распространенность акромегалии составляла 9,2 случая на 100 000. Годовая заболеваемость в Японии в 2017 году составила 0,49 случаев на 100 000. Что касается схемы лечения, то 54 % пациентов получали хирургическое лечение и 45 % – медикаментозное. Из 1569 пациентов, перенесших операцию, 29 % получали медицинскую помощь до операции. Было установлено, что распространенность и заболеваемость акромегалией в Японии были аналогичны таковым в других странах [13].

БД в фармакоэпидемиологии противоопухолевой терапии

Несмотря на внедряемые меры профилактики злокачественных новообразований (ЗНО), ранний скрининг и индивидуальное лечение, масштабы их продолжают увеличиваться: в 2025 году прогнозируется более 20 миллионов новых случаев рака по сравнению с 14,1 миллиона в 2012 году.

Использование БД широко востребовано для оценки распространенности ЗНО, факторов риска, эффективности и безопасности противоопухолевого лечения, в том числе при редких ЗНО [14]. Одна из крупнейших баз данных SEER была создана в 1973 году Национальным институтом рака США с охватом около 28 % населения США. Эта база имеет большой размер выборки, высокое качество и высокую статистическую мощность, что может предоставить исследователям, связанным с опухолями, данные высокой клинической ценности. База данных SEER является одной из наиболее репрезентативных по опухолям в Северной Америке. Хотя эта база имеет некоторые недостатки, такие как отсутствие семейного анамнеза, результатов генетических анализов, данных о рецидиве заболевания и адъювантной химиотерапии, она по-прежнему является хорошим источником данных, обеспечивающим высокое качество для клинических исследований [15]. На основе анализа 9016 больных раком яичников из базы SEER были получены данные, позволившие провести многофакторный анализ, который выявил, что представители негроидной расы (ОШ 0,58; 95 % ДИ 0,46–0,73; $p < 0,001$) и латиноамериканские женщины (ОШ 0,69; 95 % ДИ 0,54–0,88; $p < 0,001$) с меньшей вероятностью получали как химиотерапию, так и хирургическое вмешательство. Принадлежность к негроидной расе достоверно коррелировала с худшей общей пятилетней выживаемостью (относительный риск 1,13; 95 % ДИ 1,03–1,23; $p = 0,02$) [16].

Использование базы SEER позволило оценить выживаемость детей с острым лимфобластным лейкозом на основе данных наблюдения, эпидемиологии и конечных результатов. Анализ показал, что в течение трех десятилетий, с 1981 по 2010 год, общий показатель выживаемости у детей в возрасте 0–14 лет улучшался последовательно каждое десятилетие с 69,3 до 80,9 и 85,5 % [17]. В последние годы, благодаря постоянному совершенствованию методов лечения, общий прогноз детского рака значительно улучшился, но детские злокачественные опухоли по-прежнему остаются основной причиной смертности детей.

Программа «Терапевтически применимые исследования для разработки эффективных методов лечения» (TARGET) использует комплексный геномный подход для определения молекулярных изменений, которые вызывают рак у детей. Набор данных на данный момент включает 33 402 файла, представляющих

6197 случаев. Целью программы является использование данных для разработки более эффективных и менее токсичных методов лечения детей. База данных TARGET нацелена на детские опухоли, а основные заболевания включают острый лимфобластный лейкоз, острый миелоидный лейкоз, опухоли почек, нейробластому и остеосаркому. TARGET выявляет геном, связанный с конкретным видом рака у детей, и обеспечивает доступ к обнаружению терапевтических мишеней для лечения, что помогает создавать более точные варианты лечения [18].

В связи со значительным улучшением выживаемости больных ЗНО на фоне современных методов лечения становится актуальной оценка отдаленных последствий проводимого лечения.

Исследование, проведенное Suh E. и соавт. [19], оценивало наличие хронических заболеваний и позднюю смертность выживших с ЗНО. Исследование представляло собой ретроспективную когорту с продольным наблюдением за 24 363 выжившими в течение пяти лет с диагнозом рака людьми моложе 21 года в 27 академических институтах Северной Америки в период с 1970 по 1999 год. Из них в возрасте до 15 лет ЗНО было диагностировано у 5804 детей. Контрольную группу для детей составили 5059 братьев и сестер. Результаты показали, что лица, перенесшие рак в раннем подростковом и молодом возрасте, имели более высокий риск смертности и развития тяжелых и опасных для жизни хронических заболеваний, чем их братья и сестры [19].

ЗНО у детей является редкой патологией, но ведущей причиной смертности. Использование БД значительно повышает эффективность изучения причин, последствий и эффективности лечения ЗНО у детей.

Международный консорциум когорт детского рака (I4C) является одной из наиболее обширных исследовательских платформ перспективных когорт для изучения этиологии детского рака.

В настоящее время в него входят шесть когорт, набранных за шесть десятилетий (1950–2000-е годы), которые предоставили данные о 388 120 парах «мать – ребенок». Ожидается, что в течение 5 лет еще девять новых когорт из семи стран предоставят данные о 627 500 дополнительных парах «мать – ребенок». Такой размер выборки позволяет расширить возможности этиологических исследований, выяснения основных молекулярных механизмов ЗНО, оценки эффективности лечения и прогноза заболевания [20].

Для изучения последствий перенесенного ЗНО в детском возрасте планируется проведение широкомасштабного международного исследования по оценке психосоциального и нейрокогнитивного статуса, состояния физического здоровья, качества жизни, наличия хронических заболеваний у пациентов, переживших 17 видов ЗНО в возрасте до 18 лет. Планируется использование БД здравоохранения,

медицинских записей и онлайн-опроса. Целью проекта является улучшение состояния здоровья и качества жизни пациентов, переживших ЗНО в детском возрасте [21].

В отдаленном периоде возникают проблемы достоверности в оценке нежелательных явлений в случаях, когда предполагаемый негативный эффект имеет кумулятивный характер с отсроченным началом, когда существует небольшая вероятность взаимосвязи эффекта с приемом препарата или сроки исследования ограничены в пределах окна наблюдения. При длительном воздействии или при развитии отдаленных последствий снижается вероятность регистрации нежелательных явлений из-за отсутствия настороженности, что эти события взаимосвязаны. Примером могут служить такие нежелательные явления, как депрессия, онкологические заболевания, низкая информативность о которых связана с ошибками в классификации, регистрации и сроках начала их развития.

Продолжительность наблюдения за пациентами в базах данных больниц и аптек следует интерпретировать с осторожностью, поскольку за ними можно следить только в пределах одного и того же учреждения. Любое обращение или обращение за пределами учреждения не будет фиксироваться, поэтому данные долгосрочного наблюдения за вновь возникшими клиническими событиями в этих базах данных ненадежны [22].

БД в оценке эффективности и безопасности вакцинопрофилактики

В связи с пандемией COVID-19 остро встала проблема вакцинопрофилактики инфекционных заболеваний. Разработка новых вакцин требует мониторинга их безопасности [23].

База данных Всемирной организации здравоохранения ограничена спонтанными сообщениями о побочных эффектах иммунизации и не включает информацию из БД (электронных медицинских карт, зарегистрированных претензий пациентов или данных реестра иммунизации). В связи с этим активно проводятся исследования на региональных уровнях [24].

Исследование, инициированное Duszynski K.M. и соавт. [25], включало информацию из электронных баз данных 8 стран Азиатско-Тихоокеанского региона (Австралия, Китай, Гонконг, Республика Корея, Малайзия, Новая Зеландия, Таиланд и Тайвань) и проводилось с целью контроля за безопасностью вакцин. В связи с отсутствием единой базы регистрации нежелательных явлений, а также отказом ряда стран сотрудничать, из исследования были исключены Вьетнам, Бангладеш, Бруней-Даруссалам, Камбоджа, Индия, Индонезия, Лаос и Папуа – Новая Гвинея. Базы данных были в основном национальными по охвату, включали все возрасты и составляли от 1,4 до 52 миллионов человек.

Базы данных имели региональные особенности, например не во всех базах учитывались данные стационарных и амбулаторных электронных медицинских карт. В Австралии, Новой Зеландии, Тайване и Корее базы данных были представлены только реестром прививок. В последних двух странах дополнительно была регистрация заявлений о возмещении медицинских расходов. Код вакцины был доступен в 73 %. В разных источниках данных была выявлена вариабельность в отношении регистрации вакцин, маркировки, дозы, места инъекции. Например, в Тайване требуется регистрация только детских вакцин. Регистрация введения вакцин против дифтерии, кори, эпидемического паротита, коклюша, полиомиелита, краснухи и туберкулеза была зафиксирована в 73 % источников данных. Вакцины против тропических заболеваний, таких как брюшной тиф, желтая лихорадка и лихорадка денге, были зарегистрированы только в базах данных Китая и Кореи. Возраст и пол были записаны во всех базах данных, только четыре источника данных отражали этническую принадлежность. Медикаментозное лечение было зарегистрировано у 64 %, информация о весе, росте и социально-экономическом статусе была доступна почти в половине баз данных (46 %). Уровень образования и переменные образа жизни (курение, алкоголь, диета и физическая активность) регистрировались редко. Одна из проблем, с которой столкнулись исследователи, – это различия в получении разрешения на доступ к информации. Например, в Австралии, Китае и Тайване требовался наиболее широкий спектр разрешений, а в Гонконге доступ к базам данных требовал одобрения одного комитета по этике [25].

Проблема доступа также обсуждалась исследователями Канады, где в большинстве провинций созданы административные хранилища данных здравоохранения с целью облегчить доступ к этим данным для исследований. Однако исследователи описали задержки от 1 до 18 месяцев и существенные различия между провинциями в своевременности разрешений на доступ к данным. Например, в 2016 году исследование на нескольких площадках, проведенное Канадской сетью наблюдательных исследований воздействия лекарств (CNODES), показало, что сайт в Британской Колумбии не был включен из-за «длительных сроков» доступа к данным [26].

БД в фармакоэпидемиологии ЛС у пациентов с психическими заболеваниями

Проблему составляет использование БД для оценки безопасности лекарств у пациентов с психическими заболеваниями [27]. Популяционные исследования психических заболеваний имеют важное социальное значение, поскольку позволяют снизить стигматизацию психических больных. КИ с использованием БД демонстрируют не только эффективность и безопас-

ность ЛС, но и возможности психиатрического лечения для общественного здравоохранения, а также негативные последствия отсутствия доступа к необходимой медицинской помощи. В то же время стигматизация психических заболеваний приводит к усилению беспокойства по поводу конфиденциальности клинической информации и может ограничить проведение популяционных и клинических исследований [28].

Проблемы использования БД в здравоохранении

Объединение нескольких медицинских баз данных для постмаркетингового надзора за безопасностью лекарств и вакцин значительно увеличивает размер статистической выборки. Однако их использование сопровождается рядом технических проблем [29].

БД здравоохранения обычно различаются по базовым системам здравоохранения, типу собираемой информации, системам кодирования лекарств/вакцин и медицинских событий, а также языку.

Еще одна основная проблема заключается в выборе рабочих моделей для управления и анализа данных при соблюдении национальных правил в отношении их конфиденциальности и анонимности [30]. Необходимо унифицированное извлечение медицинских данных с помощью однородных алгоритмов кодирования в самых разных базах данных. Хранение и безопасный доступ к данным из разных баз требуют разработки соответствующей среды удаленных исследований [31].

В БД применяется система кодирования медицинской информации, такой как рецепты, диагнозы и процедуры, стандартизированные коды, подходящие для использования в больших наборах медицинских данных.

Важным условием возможности использования информации из разных баз данных в фармакоэпидемиологии является идентичность кодирования [32]. Недостаточное единообразие в практике кодирования ограничивает использование медицинских баз данных. Chen Y. и соавт. провели анализ 897 статей и 57 систем кодирования. Базы данных охватывали стационары, амбулаторные отделения, отделения неотложной помощи. Исследователи пришли к выводу, что системы кодирования различаются в зависимости от типа собираемой информации, клинических условий и региона использования [33].

Поскольку анализ БД становится все более распространенным, становятся актуальными вопросы обеспечения конфиденциальности и безопасности.

Защита конфиденциальности пациентов имеет первостепенное значение как по этическим, так и по юридическим причинам. Личные медицинские записи и электронные медицинские базы данных являются неотъемлемой частью систем электронного здравоохранения и поэтому должны быть защищены.

Одним из методов защиты является криптография – инструмент, который помогает обеспечить конфиденциальность медицинских данных. Шифрование является важнейшим аспектом сохранения конфиденциальности медицинских записей, особенно с учетом растущей интеграции наборов данных, электронного обмена данными. Современные криптографические методы имеют важное значение в обеспечении конфиденциальности данных в облачной среде и обеспечении безопасности работы медицинских систем. Для более сильной и надежной защиты конфиденциальности данных желательно хранить все их элементы в зашифрованном виде и разрешать доступ к исходным значениям только авторизованным пользователям. Такие методы могут облегчить использование больших распределенных медицинских баз данных в медицинских целях, например надежную проверку роли образа жизни и социальных факторов в профилактике заболеваний, эффективности вариантов лечения.

Наиболее эффективной стратегией предотвращения раскрытия конфиденциальной информации является удаление всей идентифицирующей информации из медицинских записей перед использованием в каких-либо исследованиях [34]. Там, где это практически осуществимо, использование в исследованиях должно быть ограничено такой «обезличенной» информацией. Однако в некоторых случаях идентифицирующая информация имеет важное значение. Примеры ситуаций, в которых существует явная потребность в идентификации информации, включают связь данных из клинических источников с другими записями (например, записями о рождении или смерти), связь между различными системами клинических данных (например, связь амбулаторных посещений с госпитализацией или посещением аптеки), а также необходимость связываться с пациентами для приглашения принять участие в последующих исследованиях (например, долгосрочном наблюдении за пациентами, выписанными из психиатрических больниц).

Для медицинских исследований может оказаться необходимым использовать некоторые части медицинских записей произвольно выбранных пациентов, не требуя от них подписания юридического согласия, что в противном случае могло бы внести предвзятость в исследование и сократить участие. Если публикуется только часть записи, для соблюдения законодательства о конфиденциальности важно гарантировать, что личность человека не может быть установлена.

Сокращение потерь информации является основной проблемой, с которой сталкивается подход,

основанный на искажении данных, при интеллектуальном анализе данных с сохранением конфиденциальности. Для управления и анализа данных пользователи должны использовать локальный исследовательский центр или объект с повышенными мерами безопасности. Пользователи должны иметь доступ только к минимальному количеству элементов данных, необходимых для исследования. На этапе публикации результатов анализа пользователи должны гарантировать, что отдельные пациенты не будут идентифицированы [35].

К сожалению, существуют данные как об утечке информации, так и о попытках несанкционированного доступа. В 2016 году компания SynergisTek (Консалтинговые услуги по кибербезопасности и управлению информацией) опубликовала седьмой ежегодный отчет Redspin «Защищенная медицинская информация», в котором сообщается, что хакерские атаки на поставщиков медицинских услуг увеличились на 320 % в 2016 году в сравнении с предыдущим. В результате 325 крупных вмешательств в конфиденциальную медицинскую информацию были нарушены 16 612 985 индивидуальных записей пациентов, украдены 3 620 000 записей пациентов, 40 % нарушений были связаны с несанкционированным доступом / раскрытием информации [36].

Юридическая защита конфиденциальной информации

Вопросы защиты конфиденциальной информации регламентируются многими международными законами. В первую очередь, это «Всеобщая декларация прав человека», в которой признана необходимость защиты прав и свобод человека, закреплено право на личную неприкосновенность и невмешательство в частную жизнь¹.

«Рекомендация об этических аспектах искусственного интеллекта», принятая Генеральной конференцией Организации Объединенных Наций в 2021 году², также подчеркивает, что использование искусственного интеллекта должно быть направлено во благо человека и сохранение его права на неприкосновенность частной жизни и защиту данных [37].

Защита конфиденциальной информации детей определена в «Конвенция о правах ребенка. Замечание общего порядка № 25 (2021) о правах детей в связи с цифровой средой», которая регулирует доступ информационных ресурсов к личной информации³. Право на неприкосновенность частной жизни,

¹ «Всеобщая декларация прав человека» (принята Генеральной Ассамблеей ООН 10.12.1948). URL: https://www.consultant.ru/document/cons_doc_LAW_120805/ (дата обращения: 12.03.2024).

² Генеральная конференция ООН «Рекомендация об этических аспектах искусственного интеллекта», 9–24 ноября 2021 г. URL: <https://4people.grfc.ru/analytics-and-legislation/international-and-foreign-acts/international-acts/rekomendaciya-ob-eticheskikh-aspektah-iskusstvennogo-intellekt-a/?ysclid=lfw5go9qbi466483606> (дата обращения: 12.03.2024).

³ Комитет ООН по правам ребенка. «Конвенция о правах ребенка. Замечание общего порядка № 25 (2021) о правах детей в связи с цифровой средой» от 21.11.2021. URL: <https://4people.grfc.ru/analytics-and-legislation/international-and-foreign-acts/international-acts/konvenciya-o-pravah-rebenka-zamechanie-obshego-poryadka-25-2021-o-pravah-detey-v-svyazi-s-cifrovoy-sredoy> (дата обращения: 12.03.2024).

подчеркивает, что неприкосновенность частной жизни имеет жизненно важное значение для свободы действий, достоинства и безопасности детей, а также для осуществления их прав. Персональные данные детей обрабатываются с целью предоставления им образовательных, медицинских и других услуг. Определенные комбинации персональных данных, включая биометрические, могут однозначно идентифицировать ребенка, что может привести к произвольному или незаконному ущемлению права детей на неприкосновенность частной жизни; она может иметь неблагоприятные последствия для детей, которые способны продолжать сказываться на них на более поздних этапах жизни. В тех случаях, когда запрашивается согласие на обработку данных ребенка, следует обеспечить, чтобы согласие было информированным и свободно дано ребенком, или родителем, или лицом, осуществляющим уход за ним, и получено до обработки этих данных

Проведение КИ в России регулируется Национальным стандартом РФ ГОСТ Р 52379-2005 «Надлежащая клиническая практика», который представляет собой международный этический и научный стандарт планирования и проведения исследований с участием человека в качестве субъекта, а также документального оформления и представления результатов таких исследований, и включает в том числе вопросы конфиденциальности: сохранения в тайне от неуполномоченных лиц информации, принадлежащей спонсору, или информации, позволяющей идентифицировать субъекта исследования и наличие добровольного информированного согласия, которое должно быть получено у каждого субъекта до его включения в исследование⁴.

В 2016 г. был принят Приказ Министерства здравоохранения РФ № 200н «Об утверждении правил надлежащей клинической практики», обеспечивающий систему защиты данных КИ, предотвращающую несанкционированный доступ к данным, в том числе посредством утверждения списка лиц, имеющих доступ к данным КИ с правом внесения в них изменений и резервного копирования данных⁵. Исследователь информирует участника КИ или его законного представителя о том, что записи, идентифицирующие этого участника, будут сохранены в тайне, раскрытие их допускается в соответствии с законодательством Российской Федерации и при публикации результатов КИ конфиденциальность данных участника КИ будет сохранена.

Федеральный закон от 12.04.2010 № 61 «Об обращении лекарственных средств»⁶ с поправками, принятыми в 2024 году, определил, что КИ лекарственных препаратов для медицинского применения проводятся в соответствии с правилами надлежащей клинической практики, утвержденными Евразийской экономической комиссией. Участнику должна быть гарантирована конфиденциальность участия в КИ лекарственного препарата для медицинского применения. Добровольное согласие пациента на участие в таком исследовании подтверждается его подписью или подписью его законного представителя на информационном листке пациента. В законе регулируется участие особых групп населения. Проведение КИ лекарственного препарата для медицинского применения с участием в качестве пациентов детей допускается только с согласия в письменной форме их родителей, усыновителей. Запрещается проведение КИ лекарственного препарата для медицинского применения с участием в качестве пациентов: детей-сирот и детей, оставшихся без попечения родителей; женщин в период беременности, период грудного вскармливания, за исключением случаев, если проводится КИ лекарственного препарата, предназначенного для указанных женщин, при условии необходимости получения информации только во время проведения соответствующих КИ и принятия всех необходимых мер по исключению риска нанесения вреда женщине в период беременности, женщине в период грудного вскармливания, плоду или ребенку. Допускается проведение КИ лекарственного препарата для медицинского применения, предназначенного для лечения психических заболеваний, с участием в качестве пациентов лиц с этими заболеваниями, признанных недееспособными в порядке, установленном законодательством Российской Федерации. КИ лекарственного препарата в этом случае проводится при наличии согласия в письменной форме законных представителей указанных лиц⁷.

Ряд принятых за последние 3 года законов Российской Федерации определяют меры защиты конфиденциальной информации при обработке БД: Федеральный закон «О внесении изменений в Федеральный закон “О персональных данных”» от 30.12.2020 № 519-ФЗ⁸; Федеральный закон «О внесении изменений в Федеральный закон “Об информации, информационных технологиях и о защите информации” и Федеральный закон

⁴ Национальный стандарт Российской Федерации. ГОСТ Р 52379-2005 «Надлежащая клиническая практика». URL: <https://docs.cntd.ru/document/1200041147> (дата обращения: 12.03.2024).

⁵ Приказ Министерства здравоохранения РФ от 1 апреля 2016 г. № 200н «Об утверждении правил надлежащей клинической практики». URL: <https://www.garant.ru/hotlaw/federal/1471885> (дата обращения: 12.03.2024).

⁶ Федеральный закон от 12.04.2010 № 61-ФЗ (ред. от 30.01.2024) «Об обращении лекарственных средств». URL: https://www.consultant.ru/document/cons_doc_LAW_99350/fb1fa6978ee517655735751c43bb6ef0670aade7 (дата обращения: 12.03.2024).

⁷ Там же (дата обращения: 12.03.2024).

⁸ Федеральный закон «О внесении изменений в Федеральный закон “О персональных данных”» от 30.12.2020 № 519-ФЗ. URL: <https://www.garant.ru/hotlaw/federal/1471885> (дата обращения: 12.03.2024).

“О связи”⁹; Федеральный закон «О внесении изменений в отдельные законодательные акты Российской Федерации в связи с принятием федерального закона “Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации”»¹⁰.

ЗАКЛЮЧЕНИЕ

Медицинские данные характеризуются разнообразием заболеваний, неоднородностью лечения и результатов, а также сложностью сбора, обработки и интерпретации данных, в результате чего ценность их не используется в полной мере. Рост цифровизации в медицине значительно расширил возможности использования БД в фармакоэпидемиологии. Это позволяет сократить расходы на КИ, увеличить скорость набора и расширить выборку, дает возможность оценить действие ЛС у беременных и эффективность при редких заболеваниях, оценить отдаленные эффекты лекар-

ственной терапии [38]. В то же время использование БД сопряжено с проблемой защиты конфиденциальности информации. Для расширения возможностей применения БД в здравоохранении необходимо совершенствование технологий по введению, обработке, хранению и использованию информации, а также совершенствование законодательной базы, регулирующей использование информации БД для проведения КИ.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interests. The authors declare that there is no conflict of interests.

Финансирование. Исследование не имело спонсорской поддержки (собственные ресурсы).

Financial support. The study was not sponsored (own resources).

ВКЛАД АВТОРОВ

Орлова Н.В., Суворов Г.Н. – идея исследования, обсуждение концепта, подборка и анализ материала, написание и редактирование текста.

Карселадзе Н.Д., Якушев Ю.Ю. – обсуждение концепта, сбор и обработка данных.

Гололобова Т.В. – концепция и дизайн исследования, подборка и анализ материала, написание и редактирование текста.

Все авторы утвердили окончательную версию статьи.

AUTHOR CONTRIBUTIONS

Natalya V. Orlova, Georgy N. Suvorov – research concept and design, selection and analysis of material, writing and editing text.

Natalia D. Karseladze, Yuri Yu. Yakushev – concept discussion, data collection and processing.

Tatyana V. Gololobova – concept and design of the study, selection and analysis of material, writing and editing text.

All the authors approved the final version of the article.

ЛИТЕРАТУРА/REFERENCES

- Issa A.M., Carleton B., Gerhard T., et al. Pharmacoepidemiology: A time for a new multi-disciplinary approach to precision medicine. *Pharmacoepidemiol Drug Saf.* 2021; 30(8): 985–992. <https://doi.org/10.1002/pds.5226>
- Montastruc J.L., Benevent J., Montastruc F., et al. What is pharmacoepidemiology? Definition, methods, interest and clinical applications. *Therapie.* 2019; 74(2): 169–174. <https://doi.org/10.1016/j.therap.2018.08.001>
- Raghupathi W., Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014; 2(3): 1–10. <https://doi.org/10.1186/2047-2501-2-3>
- Burden A.M. Pharmacoepidemiology and Big Data Analytics: Challenges and Opportunities when Moving towards Precision Medicine. *Chimia (Aarau).* 2019; 73(12): 1012–1017. <https://doi.org/10.2533/chimia.2019.1012>
- 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>
- Mehta N., Pandit A., Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *J Biomed Inform.* 2019; 100: 103311. <https://doi.org/10.1016/j.jbi.2019.103311>
- Benevent J., Araujo M., Hurault-Delarue C., et al. Pharmacoepidemiology in pregnancy. *Therapie.* 2019; 74(2): 289–300. <https://doi.org/10.1016/j.therap.2018.11.014>
- Belleudi V., Fortinguerra F., Poggi F.R., et al. The Italian Network for Monitoring Medication Use During Pregnancy (MoM-Net): Experience and Perspectives. *Front Pharmacol.* 2021; 12: 699062. <https://doi.org/10.3389/fphar.2021.699062>
- Fairlie L., Laves D., Kalk E., et al. Safety surveillance for PrEP in pregnant and breastfeeding women. *Front Reprod Health.* 2023; 5: 1221101. <https://doi.org/10.3389/frph.2023.1221101>
- Issa A.M., Carleton B., Gerhard T., et al. Pharmacoepidemiology: A time for a new multi-disciplinary approach to precision medicine. *Pharmacoepidemiol Drug Saf.* 2021; 30(8): 985–992. <https://doi.org/10.1002/pds.5226>
- Montastruc J.L., Benevent J., Montastruc F., et al. What is pharmacoepidemiology? Definition, methods, interest and clinical applications. *Therapie.* 2019; 74(2): 169–174. <https://doi.org/10.1016/j.therap.2018.08.001>
- Raghupathi W., Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014; 2(3): 1–10. <https://doi.org/10.1186/2047-2501-2-3>
- Burden A.M. Pharmacoepidemiology and Big Data Analytics: Challenges and Opportunities when Moving towards Precision Medicine. *Chimia (Aarau).* 2019; 73(12): 1012–1017. <https://doi.org/10.2533/chimia.2019.1012>
- Goldina T.A., Kolbin A.S., Belousov D.Yu., Borovskaya V.G. Review of real-world data study. *Good Clinical Practice.* 2021; 1: 56–63 (In Russian). <https://doi.org/10.37489/2588-0519-2021-1-56-63>
- Mehta N., Pandit A., Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *J Biomed Inform.* 2019; 100: 103311. <https://doi.org/10.1016/j.jbi.2019.103311>
- Benevent J., Araujo M., Hurault-Delarue C., et al. Pharmacoepidemiology in pregnancy. *Therapie.* 2019; 74(2): 289–300. <https://doi.org/10.1016/j.therap.2018.11.014>
- Belleudi V., Fortinguerra F., Poggi F.R., et al. The Italian Network for Monitoring Medication Use During Pregnancy (MoM-Net): Experience and Perspectives. *Front Pharmacol.* 2021; 12: 699062. <https://doi.org/10.3389/fphar.2021.699062>
- Fairlie L., Laves D., Kalk E., et al. Safety surveillance for PrEP in pregnant and breastfeeding women. *Front Reprod Health.* 2023; 5: 1221101. <https://doi.org/10.3389/frph.2023.1221101>

⁹ Федеральный закон от 31.07.2023 № 406-ФЗ «О внесении изменений в Федеральный закон “Об информации, информационных технологиях и о защите информации”» и Федеральный закон “О связи”». URL: https://www.consultant.ru/document/cons_doc_LAW_453265 (дата обращения: 12.03.2024).

¹⁰ Федеральный закон от 02.07.2021 № 331-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации в связи с принятием федерального закона “Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации”». URL: <https://www.garant.ru/hotlaw/federal/1471885> (дата обращения: 12.03.2024).

- 10 Charlton R.A., Neville A.J., Jordan S., et al. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiology and Drug Safety*. 2014; 23(6): 586–594. <https://doi.org/10.1002/pds.3613>
- 11 Yoshida S., Takeuchi M., Tanaka-Mizuno S., et al. Clinical epidemiology and pharmaco-epidemiology studies with real-world databases. *Proceedings of the Japan Academy, Series B*. 2022; 98(10): 517–528. <https://doi.org/10.2183/pjab.98.026>
- 12 Lapeyre-Mestre M. The challenges of pharmacoepidemiology of orphan drugs in rare diseases. *Therapie*. 2020; 75(2): 215–220. <https://doi.org/10.1016/j.therap.2020.02.013>
- 13 Matsubayashi K., Kawakami K. Prevalence, incidence, comorbidities, and treatment patterns among Japanese patients with acromegaly: a descriptive study using a nationwide claims database. *Endocrine Journal*. 2020; 67(10): 997–1006. <https://doi.org/10.1507/endocrj.EJ20-0129>
- 14 De Gernay S., Berdai D., Noize P. Pharmacoepidemiology for oncology clinical practice: Foundations, state of the art and perspectives. *Therapie*. 2022; 77(2): 229–240. <https://doi.org/10.1016/j.therap.2021.08.001>
- 15 Scosyrev E., Messing J., Noyes K., et al. Surveillance Epidemiology and End Results (SEER) program and population-based research in urologic oncology: an overview. *Urol Oncol*. 2012; 30(2): 126–132. <https://doi.org/10.1016/j.urolonc.2009.11.005>
- 16 Taylor J.S., He W., Harrison R., et al. Disparities in treatment and survival among elderly ovarian cancer patients. *Gynecol Oncol*. 2018; 151(2): 269–274. <https://doi.org/10.1016/j.ygyno.2018.08.041>
- 17 Ma H., Sun H., Sun X. Survival improvement by decade of patients aged 0–14 years with acute lymphoblastic leukemia: a SEER analysis. *Scientific Reports*. 2014; 4: 4227. <https://doi.org/10.1038/srep04227>
- 18 Yang J., Li Y., Liu Q., et al. Brief introduction of medical database and data mining technology in big data era. *Journal of Evidence-Based Medicine*. 2020; 13(1): 57–69. <https://doi.org/10.1111/jebm.12373>
- 19 Suh E., Stratton K.L., Leisenring W.M., et al. Late mortality and chronic health conditions in long-term survivors of early-adolescent and young adult cancers: a retrospective cohort analysis from the Childhood Cancer Survivor Study. *Lancet Oncol*. 2020; 21(3): 421–435. [https://doi.org/10.1016/S1470-2045\(19\)30800-9](https://doi.org/10.1016/S1470-2045(19)30800-9)
- 20 Tikellis G., Dwyer T., Paltiel O., et al. International Childhood Cancer Cohort Consortium. The International Childhood Cancer Cohort Consortium (I4C): A research platform of prospective cohorts for studying the aetiology of childhood cancers. *Paediatric Perinatal Epidemiology*. 2018; 32(6): 568–583. <https://doi.org/10.1111/ppe.12519>
- 21 Van Kalsbeek R.J., Hudson M.M., Mulder R.L., et al. International Childhood Cancer Outcome Project participants. A joint international consensus statement for measuring quality of survival for patients with childhood cancer. *Nat Med*. 2023; 29(6): 1340–1348. <https://doi.org/10.1038/s41591-023-02339-y>
- 22 Gault N., Castañeda-Sanabria J., De Rycke Y., et al. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. *BMC Medical Research Methodology*. 2017; 17(1): 25. <https://doi.org/10.1186/s12874-016-0278-0>
- 23 Pariente A., Bezin J. Evaluation of Covid-19 vaccines: Pharmacoepidemiological aspects. *Therapie*. 2021; 76(4): 305–309. <https://doi.org/10.1016/j.therap.2021.05.002>
- 24 Hillus D., Schwars T., Tober-Lau P., et al. Safety, reactogenicity, and immunogenicity of homologous and heterologous prime-boost immunisation with ChAdOx1 nCoV-19 and BNT162b2: a prospective cohort study. *The Lancet. Respiratory Medicine*. 2021; 9(11): 12551265. [https://doi.org/10.1016/S2213-2600\(21\)00357-X](https://doi.org/10.1016/S2213-2600(21)00357-X)
- 25 Duszynski K.M., Stark J.H., Cohet C., et al. Suitability of databases in the Asia-Pacific for collaborative monitoring of vaccine safety. *Pharmacoepidemiology and Drug Safety*. 2021; 30(7): 843–857. <https://doi.org/10.1002/pds.5214>
- 26 Kendell C., Levy A.R., Porter G., et al. Factors affecting access to administrative health data for research in Canada: a study protocol. *Int J Popul Data Sci*. 2021; 6(1): 1653. <https://doi.org/10.23889/ijpds.v6i1.1653>
- 27 Davis K.A.S., Farooq S., Hayes J.F., et al. Pharmacoepidemiology research: delivering evidence about drug safety and effectiveness in mental health. *Lancet Psychiatry*. 2020; 7(4): 363–370. [https://doi.org/10.1016/S2215-0366\(19\)30298-6](https://doi.org/10.1016/S2215-0366(19)30298-6)
- 28 Raspa M., Moultrie R., Wagner L., et al. Ethical, Legal, and Social Issues Related to the Inclusion of Individuals With Intellectual Disabilities in Electronic Health Record Research: Scoping Review. *J Med Internet Res*. 2020; 22(5): e16734. <https://doi.org/10.2196/16734>
- 29 Wylie J.E., Mineau G.P. Biomedical databases: protecting privacy and promoting research. *Trends Biotechnol*. 2003; 21(3): 113–116. [https://doi.org/10.1016/S0167-7799\(02\)00039-2](https://doi.org/10.1016/S0167-7799(02)00039-2)
- 10 Charlton R.A., Neville A.J., Jordan S., et al. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiology and Drug Safety*. 2014; 23(6): 586–594. <https://doi.org/10.1002/pds.3613>
- 11 Yoshida S., Takeuchi M., Tanaka-Mizuno S., et al. Clinical epidemiology and pharmaco-epidemiology studies with real-world databases. *Proceedings of the Japan Academy, Series B*. 2022; 98(10): 517–528. <https://doi.org/10.2183/pjab.98.026>
- 12 Lapeyre-Mestre M. The challenges of pharmacoepidemiology of orphan drugs in rare diseases. *Therapie*. 2020; 75(2): 215–220. <https://doi.org/10.1016/j.therap.2020.02.013>
- 13 Matsubayashi K., Kawakami K. Prevalence, incidence, comorbidities, and treatment patterns among Japanese patients with acromegaly: a descriptive study using a nationwide claims database. *Endocrine Journal*. 2020; 67(10): 997–1006. <https://doi.org/10.1507/endocrj.EJ20-0129>
- 14 De Gernay S., Berdai D., Noize P. Pharmacoepidemiology for oncology clinical practice: Foundations, state of the art and perspectives. *Therapie*. 2022; 77(2): 229–240. <https://doi.org/10.1016/j.therap.2021.08.001>
- 15 Scosyrev E., Messing J., Noyes K., et al. Surveillance Epidemiology and End Results (SEER) program and population-based research in urologic oncology: an overview. *Urol Oncol*. 2012; 30(2): 126–132. <https://doi.org/10.1016/j.urolonc.2009.11.005>
- 16 Taylor J.S., He W., Harrison R., et al. Disparities in treatment and survival among elderly ovarian cancer patients. *Gynecol Oncol*. 2018; 151(2): 269–274. <https://doi.org/10.1016/j.ygyno.2018.08.041>
- 17 Ma H., Sun H., Sun X. Survival improvement by decade of patients aged 0–14 years with acute lymphoblastic leukemia: a SEER analysis. *Scientific Reports*. 2014; 4: 4227. <https://doi.org/10.1038/srep04227>
- 18 Yang J., Li Y., Liu Q., et al. Brief introduction of medical database and data mining technology in big data era. *Journal of Evidence-Based Medicine*. 2020; 13(1): 57–69. <https://doi.org/10.1111/jebm.12373>
- 19 Suh E., Stratton K.L., Leisenring W.M., et al. Late mortality and chronic health conditions in long-term survivors of early-adolescent and young adult cancers: a retrospective cohort analysis from the Childhood Cancer Survivor Study. *Lancet Oncol*. 2020; 21(3): 421–435. [https://doi.org/10.1016/S1470-2045\(19\)30800-9](https://doi.org/10.1016/S1470-2045(19)30800-9)
- 20 Tikellis G., Dwyer T., Paltiel O., et al. International Childhood Cancer Cohort Consortium. The International Childhood Cancer Cohort Consortium (I4C): A research platform of prospective cohorts for studying the aetiology of childhood cancers. *Paediatric Perinatal Epidemiology*. 2018; 32(6): 568–583. <https://doi.org/10.1111/ppe.12519>
- 21 Van Kalsbeek R.J., Hudson M.M., Mulder R.L., et al. International Childhood Cancer Outcome Project participants. A joint international consensus statement for measuring quality of survival for patients with childhood cancer. *Nat Med*. 2023; 29(6): 1340–1348. <https://doi.org/10.1038/s41591-023-02339-y>
- 22 Gault N., Castañeda-Sanabria J., De Rycke Y., et al. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. *BMC Medical Research Methodology*. 2017; 17(1): 25. <https://doi.org/10.1186/s12874-016-0278-0>
- 23 Pariente A., Bezin J. Evaluation of Covid-19 vaccines: Pharmacoepidemiological aspects. *Therapie*. 2021; 76(4): 305–309. <https://doi.org/10.1016/j.therap.2021.05.002>
- 24 Hillus D., Schwars T., Tober-Lau P., et al. Safety, reactogenicity, and immunogenicity of homologous and heterologous prime-boost immunisation with ChAdOx1 nCoV-19 and BNT162b2: a prospective cohort study. *The Lancet. Respiratory Medicine*. 2021; 9(11): 12551265. [https://doi.org/10.1016/S2213-2600\(21\)00357-X](https://doi.org/10.1016/S2213-2600(21)00357-X)
- 25 Duszynski K.M., Stark J.H., Cohet C., et al. Suitability of databases in the Asia-Pacific for collaborative monitoring of vaccine safety. *Pharmacoepidemiology and Drug Safety*. 2021; 30(7): 843–857. <https://doi.org/10.1002/pds.5214>
- 26 Kendell C., Levy A.R., Porter G., et al. Factors affecting access to administrative health data for research in Canada: a study protocol. *Int J Popul Data Sci*. 2021; 6(1): 1653. <https://doi.org/10.23889/ijpds.v6i1.1653>
- 27 Davis K.A.S., Farooq S., Hayes J.F., et al. Pharmacoepidemiology research: delivering evidence about drug safety and effectiveness in mental health. *Lancet Psychiatry*. 2020; 7(4): 363–370. [https://doi.org/10.1016/S2215-0366\(19\)30298-6](https://doi.org/10.1016/S2215-0366(19)30298-6)
- 28 Raspa M., Moultrie R., Wagner L., et al. Ethical, Legal, and Social Issues Related to the Inclusion of Individuals With Intellectual Disabilities in Electronic Health Record Research: Scoping Review. *J Med Internet Res*. 2020; 22(5): e16734. <https://doi.org/10.2196/16734>
- 29 Wylie J.E., Mineau G.P. Biomedical databases: protecting privacy and promoting research. *Trends Biotechnol*. 2003; 21(3): 113–116. [https://doi.org/10.1016/S0167-7799\(02\)00039-2](https://doi.org/10.1016/S0167-7799(02)00039-2)

- 30 Kumamaru H., Togo K., Kimura T., et al. Inventory of real-world data sources in Japan: Annual survey conducted by the Japanese Society for Pharmacoepidemiology Task Force. *Pharmacoepidemiol Drug Saf.* 2023; 33(1): e5680. <https://doi.org/10.1002/pds.5680>
- 31 Trifirò G., Coloma P.M., Rijnbeek P.R., et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *Journal of Internal Medicine.* 2014; 275(6): 551–561. <https://doi.org/10.1111/joim.12159>
- 32 Franco J.S., Vizcaya D. Availability of secondary healthcare data for conducting pharmacoepidemiology studies in Colombia: A systematic review. *Pharmacol Res Perspect.* 2014; 8(5): e00661. <https://doi.org/10.1002/prp2.66>
- 33 Chen Y., Zivkovic M., Wang T., et al. A Systematic Review of Coding Systems Used in Pharmacoepidemiology and Database Research. *Methods Inf Med.* 2018; 57(1): 1–42. <https://doi.org/10.3414/ME17-05-0006>
- 34 Leufkens H.G. Privacy issues in pharmacoepidemiology: the importance of weighing costs and benefits. *Pharmacoepidemiol Drug Saf.* 2002; 10(7): 659–662. <https://doi.org/10.1002/pds.645>
- 35 Rivera D.R., Gokhale M.N., Reynolds M.W., et al. Linking electronic health data in pharmacoepidemiology: Appropriateness and feasibility. *Pharmacoepidemiol Drug Saf.* 2020; 29(1): 18–29. <https://doi.org/10.1002/pds.4918>
- 36 Abouelmehdi K., Beni-Hessane A., Khaloufi H. Big healthcare data: preserving security and privacy. *Journal of Big Data.* 2018; 5: 1. <https://doi.org/10.1186/s40537-017-0110-7>
- 37 Рицу С. Рекомендации по этическим аспектам искусственного интеллекта в приложении к сфере трудовых отношений. *Journal of Digital Technologies and Law.* 2023; 1(2): 498–519. <https://doi.org/10.21202/jdtl.2023.21>
- 38 Alonso-Betanzos A., Bolón-Canedo V. Big-Data Analysis, Cluster Analysis, and Machine-Learning Approaches. *Adv Exp Med Biol.* 2018; 1065: 607–626. https://doi.org/10.1007/978-3-319-77932-4_37
- 30 Kumamaru H., Togo K., Kimura T., et al. Inventory of real-world data sources in Japan: Annual survey conducted by the Japanese Society for Pharmacoepidemiology Task Force. *Pharmacoepidemiol Drug Saf.* 2023; 33(1): e5680. <https://doi.org/10.1002/pds.5680>
- 31 Trifirò G., Coloma P.M., Rijnbeek P.R., et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *Journal of Internal Medicine.* 2014; 275(6): 551–561. <https://doi.org/10.1111/joim.12159>
- 32 Franco J.S., Vizcaya D. Availability of secondary healthcare data for conducting pharmacoepidemiology studies in Colombia: A systematic review. *Pharmacol Res Perspect.* 2014; 8(5): e00661. <https://doi.org/10.1002/prp2.66>
- 33 Chen Y., Zivkovic M., Wang T., et al. A Systematic Review of Coding Systems Used in Pharmacoepidemiology and Database Research. *Methods Inf Med.* 2018; 57(1): 1–42. <https://doi.org/10.3414/ME17-05-0006>
- 34 Leufkens H.G. Privacy issues in pharmacoepidemiology: the importance of weighing costs and benefits. *Pharmacoepidemiol Drug Saf.* 2002; 10(7): 659–662. <https://doi.org/10.1002/pds.645>
- 35 Rivera D.R., Gokhale M.N., Reynolds M.W., et al. Linking electronic health data in pharmacoepidemiology: Appropriateness and feasibility. *Pharmacoepidemiol Drug Saf.* 2020; 29(1): 18–29. <https://doi.org/10.1002/pds.4918>
- 36 Abouelmehdi K., Beni-Hessane A., Khaloufi H. Big healthcare data: preserving security and privacy. *Journal of Big Data.* 2018; 5: 1. <https://doi.org/10.1186/s40537-017-0110-7>
- 37 Ritsu S. Recommendations on the ethical aspects of artificial intelligence in the application to the field of labor relations. *Journal of Digital Technologies and Law.* 2023; 1(2): 498–519 (In Russian). <https://doi.org/10.21202/jdtl.2023.21>
- 38 Alonso-Betanzos A., Bolón-Canedo V. Big-Data Analysis, Cluster Analysis, and Machine-Learning Approaches. *Adv Exp Med Biol.* 2018; 1065: 607–626. https://doi.org/10.1007/978-3-319-77932-4_37

Информация об авторах

Орлова Наталья Васильевна — д-р мед. наук, профессор кафедры факультетской терапии педиатрического факультета ФГАОУ ВО «Российский национальный исследовательский медицинский университет им. Н.И. Пирогова» Министерства здравоохранения Российской Федерации.

ORCID: <https://orcid.org/0000-0002-4293-3285>

Суворов Георгий Николаевич — канд. юрид. наук, доцент кафедры госпитальной эпидемиологии, медицинской паразитологии и тропических болезней ФГБОУ ДПО «Российская медицинская академия непрерывного профессионального образования» Министерства здравоохранения Российской Федерации; заместитель директора по науке и сетевому взаимодействию ФГАНУ «Центр социологических исследований».

ORCID: <https://orcid.org/0000-0001-8452-5522>

Карселадзе Наталья Джимшеровна — канд. мед. наук, врач-терапевт Медицинского центра DMEMED аэропорта Домодедово имени М.В. Ломоносова.

ORCID: <https://orcid.org/0000-0001-8124-109X>

Якушев Юрий Юрьевич — студент 6-го курса ФГАОУ ВО «Российский национальный исследовательский медицинский университет им. Н.И. Пирогова» Министерства здравоохранения Российской Федерации.

ORCID: <https://orcid.org/0009-0007-2537-8658>

Гололобова Татьяна Викторовна — д-р мед. наук, доцент, заведующая кафедрой госпитальной эпидемиологии, медицинской паразитологии и тропических болезней ФГБОУ ДПО «Российская медицинская академия непрерывного профессионального образования» Министерства здравоохранения Российской Федерации; заместитель директора по научной работе ФГУП «Государственный научно-исследовательский институт биологического приборостроения» ФМБА России.

ORCID: <https://orcid.org/0000-0001-9033-5223>

Information about the authors

Natalya V. Orlova — Dr. of Sci. (Medicine), Professor, Department of Faculty Therapy, Faculty of Pediatrics, N.I. Pirogov Russian National Research Medical University.

ORCID: <https://orcid.org/0000-0002-4293-3285>

Georgiy N. Suvorov — Cand. of Sci. (Jurid.), Associate Professor, Department of Hospital Epidemiology, Medical Parasitology and Tropical Diseases, Russian Medical Academy of Continuing Professional Education; Deputy Director for Science and Network Interaction, Center for Sociological Research.

ORCID: <https://orcid.org/0000-0001-8452-5522>

Natalia D. Karseladze — Cand. of Sci. (Medicine), therapist, Medical Center DMEMED, M.V. Lomonosov Airport Domodedovo.

ORCID: <https://orcid.org/0000-0001-8124-109X>

Yuri Yu. Yakushev — 6th year student, N.I. Pirogov Russian National Research Medical University.

ORCID: <https://orcid.org/0009-0007-2537-8658>

Tatyana V. Gololobova — Dr. of Sci. (Medicine), Associate Professor, Head of the Department of Hospital Epidemiology, Medical Parasitology and Tropical Diseases, Russian Medical Academy of Continuous Professional Education; Deputy Director for Scientific Work, State Scientific Research Institute of Biological Instrumentation FMBA of Russia.

ORCID: <https://orcid.org/0000-0001-9033-5223>