

Анализ подходов FDA по использованию данных электронных медицинских карт и данных о медицинских претензиях для поддержки принятия регуляторных решений

Радаева К. С. , Пчелинцев М. В. 

ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова»
МЗ РФ, Санкт-Петербург, Российская Федерация

Аннотация

В 2021 году Управление по контролю качества пищевых продуктов и лекарственных средств США (FDA, Food and Drug Administration) выпустило проект руководства по использованию электронных медицинских карт (ЭМК) и данных о медицинских претензиях для принятия регуляторных решений. В проекте руководства даются рекомендации спонсорам в отношении использования данных реальной клинической практики при проведении исследований. Также в нём рассматриваются проблемы и ограничения, связанные с использованием этих данных, и даются рекомендации по их преодолению. Целью данной статьи является представить краткий обзор ключевых аспектов рекомендаций FDA по использованию данных ЭМК при принятии регуляторных решений.

Ключевые слова: рекомендации; данные реальной клинической практики; электронные медицинские карты; ЭМК; RWD; FDA

Для цитирования:

Радаева К. С., Пчелинцев М. В. Оценка данных электронных медицинских карт и данных о медицинских претензиях для поддержки принятия регуляторных решений. *Реальная клиническая практика: данные и доказательства*. 2023;3(3):28–35. <https://doi.org/10.37489/2782-3784-myrd-39>

Поступила: 09 октября 2023 г. **Принята:** 10 августа 2023 г. **Опубликована:** 12 октября 2023 г.

Analyzing FDA's approaches to using electronic health record and medical claims data to support regulatory decision making

Ksenia S. Radaeva , Mikhail V. Pchelintsev 

First St. Petersburg State Medical University named after I. P. Pavlov, St. Petersburg, Russian Federation

Abstract

In 2021, the U.S. Food and Drug Administration (FDA) issued draft guidance on the use of electronic health records and medical claims data for regulatory decisions. The draft guidance provides recommendations for sponsors regarding the use of real-world data in conducting studies. It also addresses the challenges and limitations associated with the use of these data and provides recommendations to overcome them. The purpose of this article is to provide a brief overview of key aspects of the FDA's guidance on the use of electronic health record data in regulatory decision making.

Keywords: guidelines; real-world data; electronic health records; EHR; RWD; FDA

For citation:

Radaeva KS, Pchelintsev MV. ICH Analyzing FDA's approaches to using electronic health record and medical claims data to support regulatory decision making. *Real-World Data & Evidence*. 2023;3(3):28–35. <https://doi.org/10.37489/2782-3784-myrd-39>

Received: October 9, 2023. **Accepted:** October 10, 2023. **Published:** October 12, 2023.

Введение

Данные реальной клинической практики (РКП) могут быть полезны, как при оценке эффективности и безопасности лекарств в течение всего цикла их «жизни», так и при принятии регуляторных

решений. Однако, использование данных РКП сопровождается определёнными сложностями. К ним относятся следующие:

- неполные данные;

- сопоставимость и практическая реализуемость данных;
- неоднородность баз данных и отсутствие единого методологического подхода к сбору и анализу медицинских записей;
- наличие различий между традиционными рандомизированными клиническими исследованиями (РКИ) и обсервационными исследованиями, с использованием данных РКИ.

Тем не менее, несмотря на указанные сложности, данные РКИ являются ценным дополнением к результатам проведённых клинических исследований, так как они отражают реальную практику применения лекарственных средств у большого количества пациентов. Также следует подчеркнуть важность использования этих данных, в тех случаях, когда проведение традиционных клинических исследований не представляется возможным или этичным.

В 2021 году Управление по контролю качества пищевых продуктов и лекарственных средств США (англ. Food and Drug Administration; FDA) выпустило проект руководства по использованию электронных медицинских карт (ЭМК) и данных о медицинских претензиях для принятия регуляторных решений. Согласно проекту руководства, FDA рекомендует спонсорам клинических исследований, использующих данные РКИ, учитывать адекватность и релевантность источника данных, недостающие данные, связь данных, а также подтверждение и валидацию переменных исследования наряду с обеспечением качества. В руководстве также обсуждаются конкретные вопросы, вызывающие обеспокоенность FDA. Например, достаточно ли последующего наблюдения в источниках данных РКИ для установления исходов и не ограничивают ли недостающие данные интерпретацию результатов. FDA рекомендует спонсорам обсуждать конкретные вопросы с соответствующим отделом экспертизы, однако процесс и сроки проведения таких обсуждений требуют чёткого определения [1].

Далее мы сосредоточим внимание на ключевых компонентах руководства, которое обеспечивает методологическую основу для ответа на контрольные вопросы регуляторных органов при оценке качества данных РКИ: качество источника данных (его адекватность и релевантность), полнота и сопоставимость данных, элементы дизайна исследования (характеристика и валидация воздействий, выбор конечных точек, популяции и учёт ковариат), контроль и оценка качества.

Адекватность, релевантность и качество данных реальной клинической практики

До начала проведения исследования спонсоры, проводящие исследования с использованием

электронных медицинских карт (ЭМК) или данных о медицинских претензиях и намеревающиеся представить их результаты в FDA для принятия регуляторного решения, должны предоставить FDA подробные планы с описанием дизайна исследования и методов статистического анализа. Если спонсоры хотят получить отзывы или рекомендации от FDA до начала исследования, им следует запросить комментарии или встречу с отделом FDA, отвечающим за оценку исследования. Это позволит спонсорам обсудить с FDA план исследования и все возникающие у них проблемы и вопросы. Для спонсоров важно определить и описать все ключевые аспекты своего исследования, включая план исследования, методы анализа данных, порядок проведения исследования и представления результатов. В протоколе и окончательном отчёте об исследовании необходимо объяснить, как каждый из этих элементов был определён с помощью выбранного источника данных РКИ, включая все проведённые валидационные исследования.

Исследования РКИ по применению, эффективности и безопасности медицинских изделий сталкиваются с проблемами, связанными с недостаточной детализацией и неоднозначностью сложных вариантов разработки и анализа. Для решения этой проблемы в рамках государственно-частного сотрудничества был разработан структурированный шаблон под названием STaRT-RWE [2]. Этот шаблон призван помочь в разработке, проведении и рецензировании исследований, предоставляя рекомендации и общую структуру для исследователей, спонсоров, рецензентов и лиц, принимающих решения. Он дополняет существующие контрольные списки и сводит к минимуму двусмысленные формулировки за счёт использования табличных и визуальных форматов. STaRT-RWE совместим с различными схемами исследований и реальными источниками данных, а также соответствует предыдущим рекомендациям и инструментам оценки смещения. Шаблон включает таблицы параметров исследования и рисунок для представления методов и результатов исследования. Он разработан с учётом гибкости и может быть включён в существующие процессы разработки протоколов и отчётности. Ожидается, что практика использования шаблона станет рутинной, а для облегчения его применения могут быть разработаны программные решения. Тем не менее шаблон STaRT-RWE имеет ряд ограничений. Во-первых, структура, заложенная в таблицах, может не подойти для некоторых случаев использования. В зависимости от исследования может оказаться актуальной только часть таблиц в STaRT-RWE. Во-вторых, использование шаблона для планирования исследования и принятия решений не гарантирует объективности результатов. Од-

нако чёткая и недвусмысленная информация о том, как были получены результаты и какие стратегии использовались для смягчения предвзятости, может значительно облегчить интерпретацию результатов исследования. В-третьих, хотя шаблон включает раздел, посвящённый источникам данных, в нём не отражена вся необходимая информация для оценки пригодности данных. Соответствие источников данных должно быть документировано, включая описание доступных полей данных и полноту их сбора. Кроме того, должны быть представлены чёткие сведения о происхождении данных, задокументированы все преобразования, выполненные в потоках данных, использованных для создания исследовательской базы данных [2].

Основными ограничениями для использования данных РКП являются [3, 4]:

- **качество данных** — ключевой фактор, определяющий целесообразность и полезность использования источника данных. Если качество источника вызывает подозрения, следует отказаться от использования данного источника;
- **адекватность сбора данных** — источники могут не содержать достаточного количества данных для целей исследования, для ответа на конкретный исследовательский вопрос могут потребоваться исследования различной продолжительности;
- **отсутствующие данные** — некоторые данные могут быть некорректно закодированы или отсутствовать в источниках, например, при смене пациентом поставщика услуг или страхового плана данные о прошлых эпизодах обращения за медицинской помощью могут отсутствовать;
- **размер выборки** — размер выборки должен быть достаточным для обеспечения аналитической точности с учётом ожидаемого эффекта лечения или статистической дисперсии, для ответа на этот запрос может понадобиться использование альтернативных источников данных;
- **потенциальные смещения и ошибки измерения**, в том числе неизмеряемые сопутствующие факторы, должны быть учтены. Интерпретация данных РКП может быть сложной, особенно если ожидаемый эффект лечения невелик. Даже при большом объёме выборки потенциальная погрешность может перевесить значимость результатов. Поэтому важно включать количественную оценку смещения или анализ чувствительности для учёта небольших изменений в алгоритмах, устойчивости предположений и влияния отсутствующих данных.

В руководстве также подчёркивается, что в разных странах мира существуют различия в медицинской практике и системах здравоохранения, что

может повлиять на соответствие источников данных вопросу исследования и целесообразность их использования. Пациенты, участвующие в различных программах оплаты медицинских услуг, могут иметь различные характеристики. Такие факторы, как распределение лекарственных препаратов по уровням, формулярные решения и охват пациентов, также могут влиять на различия в тяжести заболевания или других характеристиках болезни у пациентов, получающих одну и ту же терапию в разных системах здравоохранения. Важно убедиться, что источники данных охватывают все популяции, имеющие отношение к исследованию. FDA рекомендует указывать причину выбора конкретных источников данных, справочную информацию о системе здравоохранения, а также описание практики назначения и использования препаратов. Для источников данных, не относящихся к США, следует дать объяснение того, как эти факторы могут повлиять на обобщение результатов исследования на американскую популяцию.

Одним из инструментов при оценке целесообразности использования источников данных из формуляров и электронных медицинских карт для решения конкретных исследовательских задач является система PICOT [3]. Данная система включает в себя оценку по следующим параметрам:

- **P (population)**, популяция — исследователи должны оценить включает ли источник данных интересующую популяцию пациентов и может ли эта популяция быть достоверно и надёжно определена;
- **I (intervention)**, вмешательство — в источнике данных должны содержаться данные об изучаемом вмешательстве или лекарстве-компараторе и, исследователи должны оценить, есть ли возможность точно определить эти данные;
- **C (comparison)**, сравнение — по сравнению с каким воздействием оцениваются преимущества;
- **O (outcome)**, результаты — источник данных должен содержать информацию об интересующих исходах;
- **T (time)**, сроки — источник должен предоставлять данные за желаемый период исследования, и исследователи должны определить, имеются ли данные за требуемый им период времени.

Чтобы оценить целесообразность использования конкретного источника данных РКП, исследователи должны рассмотреть, можно ли идентифицировать каждый элемент PICOT в источнике, надёжно ли закодированы данные, имеются ли проверенные алгоритмы или структурированные данные для определения этих элементов. Кроме того, следует оценить доступность данных последующего наблюдения и адекватность их сбора.

FDA признаёт, что ни один источник данных не подходит для решения всех исследовательских вопросов, и все источники данных имеют свои ограничения [1, 4]. Исследователи должны рассмотреть адекватность сбора данных для каждого элемента конкретного вопроса исследования. Потенциальные ограничения часто можно устранить с помощью дизайна исследования, анализа чувствительности или ссылкой на дополнительные данные. Однако такие проблемы, как качество данных, практика медицинского кодирования и полнота сбора данных, могут привести к получению не интерпретируемых результатов [3, 5]. В руководстве FDA выражает озабоченность по поводу отсутствия в источниках важных для анализа данных. Агентство считает, что привязка к источникам данных может повысить эффективность сбора информации и уменьшить количество отсутствующих элементов [1]. *Girman CJ et al.* (2022 г.) считают, что должно поощряться использование дополнительных анализов чувствительности, аналогичных тем, что используются при анализе данных полученных в ходе классических РКИ в случае отсутствия данных о воздействии или исходе, с целью понять, изменились бы результаты при наличии этих недостающих данных [5]. Также в проекте руководства описана важная информация, учитываемая при анализе фармакоэпидемиологами, которая отсутствует в ЭМК [1]. К такой информации относятся формулярные ограничения, данные о ступенчатой терапии, предварительные разрешения. Ещё одним важным фактором, обуславливающим отсутствие данных в США, является особенность системы страхования здоровья граждан страны. Граждане США часто меняют медицинскую страховку при смене работодателя, что нарушает непрерывность данных о случаях медицинского обслуживания пациента. Для решения этой проблемы FDA предлагает использовать данные из других источников. Однако использование других источников не всегда представляется возможным, в связи с частым отсутствием достаточной связи между различными источниками данных, с юридическими аспектами обмена данных и вопросами конфиденциальности [1]. В руководстве FDA признаётся, что информация о некоторых элементах данных может содержаться в неструктурированных полях внутри ЭМК (например, в клинических записях), и рекомендуется тщательно описывать процесс извлечения и проверки таких данных, будь то извлечение вручную или с использованием автоматизированной технологий [1]. Однако этот процесс не детализирован в руководстве.

Использование таких связей позволяет увеличить объём и глубину данных об отдельных пациентах с течением времени и получить дополнительные данные для целей валидации. В исследованиях,

предусматривающих установление новых связей между внутренними или внешними источниками данных, в протоколе должны быть описаны каждый источник данных, информация, которая будет получена, методы установления связей, а также точность и полнота связей с течением времени. Если исследование предполагает получение дополнительных данных с помощью различных методов, таких как интервью, опросы или цифровые медицинские технологии, в протоколе также должны быть описаны методы сбора данных и их интеграции с электронными медицинскими данными. Для установления связи между данными могут использоваться различные подходы, такие как вероятностный и детерминированный, каждый из которых имеет свои преимущества в зависимости от сценария. При использовании вероятностного подхода план анализа должен включать проверку влияния степени совпадения и устойчивости полученных результатов. Для исследований, требующих объединения данных из нескольких источников или сайтов, важно продемонстрировать, как можно получить и интегрировать данные из разных источников с приемлемым качеством, учитывая возможную неоднородность характеристик населения, клинической практики и кодирования в разных источниках данных. Внимание к курированию данных и понимание сценариев связывания данных имеет решающее значение для обеспечения уместности новых связей данных. Необходимо документировать тип обработки данных для решения проблем дублирования или фрагментации, а также подходы, применяемые для решения проблем, которые не могут быть полностью устранены путём обработки [1].

Значительные объёмы клинических данных в ЭМК являются неструктурированными, включая текстовые поля данных, например, заметки врача, и нестандартизированную информацию в компьютерных документах, например, радиологических отчётах в формате PDF. Преобразование неструктурированных данных с целью повышения эффективности абстрагирования данных используются различные подходы, как существующие, так и разрабатываемые технологии, позволяющие преобразовывать неструктурированные данные в пригодный для вычислений формат.

Для извлечения данных из ЭМК могут быть использованы современные технологии искусственного интеллекта (ИИ). ИИ может обрабатывать неструктурированные документы, такие как записи врача или отчёты о лабораторных исследованиях, для выявления и извлечения необходимой информации. Это помогает создавать структурированные наборы данных с клинически значимыми переменными, необходимыми для анализа. ИИ может повысить реалистичность и адекватность данных из ЭМК за

счёт автоматизации процесса извлечения данных. Традиционные методы, такие как ручной просмотр карт пациентов, требуют много времени и ресурсов, что ограничивает число пациентов, доступных для исследования. ИИ может эффективно обрабатывать большие объёмы данных, позволяя исследователям получить доступ к более широкой популяции пациентов и более эффективно генерировать доказательства, полученные на основе данных РКИ [6].

Плюсы использования ИИ для оценки и извлечения данных ЭМК [6]:

- **Эффективность и масштабируемость:** методы ИИ позволяют быстро и точно обрабатывать большие объёмы неструктурированных данных, обеспечивая эффективное извлечение клинически значимой информации. Такая масштабируемость позволяет исследователям анализировать данные большего числа пациентов, что ведёт к получению более надёжных и репрезентативных доказательств.
- **Повышенная точность:** модели ИИ могут достигать высокой производительности при извлечении переменных из неструктурированных документов ЭМК, часто сравнимой с человеческим абстрагированием или даже превосходящей его. Такая точность снижает риск ошибок и несоответствий, которые могут возникнуть при ручном сборе данных, обеспечивая надёжность извлечённых данных для целей исследования.
- **Доступ к большему количеству клинической информации:** данные ЭМК часто содержат ценную клиническую информацию, записанную в виде примечаний или резюме в свободном формате. Модели ИИ, особенно основанные на архитектуре глубокого обучения, способны улавливать нюансы естественного языка и извлекать важнейшие клинические детали, которые могут быть упущены при использовании традиционных структурированных элементов данных. Такой доступ к клиническим деталям повышает глубину и детализацию извлекаемых данных.

Ограничения использования методов ИИ [6]:

- **Качество данных:** неточности в обучающих данных могут привести к необъективным или неточным результатам в извлекаемых переменных. Важно обеспечить репрезентативность и разнообразие обучающих данных, чтобы свести к минимуму предвзятость и улучшить обобщаемость моделей.
- **Отсутствие понимания контекста:** модели ИИ могут с трудом понимать контекст и нюансы клинической информации, особенно в сложных случаях или при работе с неоднозначными или неполными данными. Это ограничение может повлиять на точность и надёжность извлекаемых переменных, что подчёркивает необходимость постоянного человеческого контроля и валидации.

- **Прозрачность и объяснимость:** использование ИИ для извлечения данных из ЭМК вызывает озабоченность по поводу прозрачности и воспроизводимости. Для обеспечения прозрачности и воспроизводимости важно чётко описать методы и процессы, используемые при извлечении данных на основе ИИ. Такая прозрачность необходима для того, чтобы исследователи, регулирующие органы и медицинские учреждения понимали и доверяли результатам, полученным в результате анализа с помощью ИИ.

В целом использование ИИ для оценки и извлечения данных из ЭМК даёт значительные преимущества с точки зрения эффективности, точности и доступа к богатой клинической информации. Однако для обеспечения надёжности и валидности извлечённых переменных в исследовательских целях важно решить проблемы, связанные с качеством данных, предвзятостью, пониманием контекста и прозрачностью.

Валидация воздействий, результатов и контроль качества

В проекте руководства FDA отметило, что алгоритмы определения воздействия, исходов и важных ковариат должны быть определены на основе таких данных, как медицинские диагнозы, коды процедур или отпущенные лекарственные препараты [1]. Как и в случае клинических оценок в РКИ, алгоритмы не всегда могут быть точными, что может привести к неправильной классификации, а неправильная классификация может привести к смещённым оценкам эффекта лечения. Для предотвращения этой проблемы следует поощрять проведение анализа чувствительности и количественный анализ систематических ошибок (*англ.* Quantitative Bias Analysis; QBA) для информирования о потенциальном влиянии таких смещений, особенно когда размер эффекта незначителен [7, 8].

Оперативные определения разрабатываются на основе концептуальных определений для извлечения точных и полных данных из источника данных. В исследованиях, использующих ЭМК или данные о медицинских претензиях, оперативные определения часто представляют собой электронные алгоритмы на основе кодов, использующие структурированные элементы данных. В других исследованиях оперативные определения могут быть получены в результате извлечения релевантной информации из неструктурированных данных или комбинирования структурированных и неструктурированных элементов данных. Оперативные определения могут также включать дополнительные методы сбора данных, например опросы пациентов, если это целесообразно [1].

Оперативные определения, как правило, не совершенны и могут приводить к неправильной классификации интересующей переменной, что приводит к ложноположительным и ложноотрицательным результатам. Такая неправильная классификация может исказить ассоциацию между воздействием и исходом в определённом направлении и степени. Хотя полная верификация переменной сводит к минимуму неправильную классификацию и максимизирует внутреннюю валидность исследования, понимание последствий потенциальной неправильной классификации имеет решающее значение для определения того, какие переменные требуют валидации и в каком объёме. В некоторых исследованиях допустимая ошибка классификации может не оказывать существенного влияния на интерпретацию результатов, особенно если размер гипотетического эффекта велик [1].

При оценке влияния потенциальной ошибки классификации на величину ассоциации и интерпретацию результатов спонсоры должны учитывать несколько факторов. К ним относятся степень неправильной классификации, её дифференциальный или недифференциальный характер (например, неправильная классификация результата по воздействию), зависимый или независимый характер (например, коррелирующие неправильные классификации воздействия и результата), а также направление смещения связи между воздействием и результатом [1].

Хотя полная верификация переменной исследования считается наиболее строгим подходом, существуют сценарии, в которых она может оказаться невыполнимой, например, в больших популяциях или при отсутствии данных о референтных стандартах для всех испытуемых. В таких случаях может быть достаточно оценить эффективность операционного определения. Это предполагает оценку точности и полноты операционного определения при выявлении случаев и не случаев интересующей переменной.

В руководстве указывается на необходимость документирования всех процедур контроля качества в процессе сбора, обработки и преобразования данных в окончательный аналитический набор данных [1]. Несмотря на важность валидации происхождения данных, при использовании коммерческих баз данных спонсорам крайне сложно получить подробную документацию и сверить данные с первоисточниками. Руководство должно быть направлено на обеспечение качества, которое спонсоры могут легко контролировать и документировать с момента получения данных, а процедуры обеспечения качества должны определяться тем, как потенциальная неправильная классификация, ошибки и предвзятость могут повлиять на выводы исследо-

вания. С другой стороны, слишком жёсткие требования к контролю качества значительно повышают неэффективность, что противоречит поставленным FDA целям. Вместо этого следует использовать риск-ориентированный мониторинг с учётом рисков и поощрять прозрачность процедур контроля качества и воспроизведение анализов [9].

В проекте руководства описываются процессы контроля качества при сборе и обработке данных, которые аналогичны или превосходят требования к сбору первичных данных в РКИ. Рекомендации по оценке качества ЭМК и данных о медицинских претензиях в первую очередь ориентированы на распределённые сети данных, которые объединяют, связывают и обрабатывают разрозненные источники данных для создания комплексного хранилища данных. Исследователи должны оценивать полноту, точность и правдоподобность данных, сверять их с первоисточником и соответствовать стандартам данных, принятым на основе консенсуса. В протоколе исследования и плане анализа необходимо указать происхождение данных и описать, как процедуры сбора, обработки и преобразования данных могут повлиять на целостность данных и достоверность исследования. В руководстве FDA ссылается на полную валидацию всех переменных исследования с использованием данных ЭМК, признавая, что не может быть выполнено для всех переменных в любом исследовании РКП. В руководстве рекомендуется «ослеплять» референта для уменьшения предвзятости и использовать стандартизированные и воспроизводимые процессы реферирования и вынесения решений для минимизации внутри- и межреферентных ошибок.

Однако в руководстве не уточнены, какие подходы к валидации следует использовать. Не проводится различие между критическими переменными и не устанавливается их приоритет при составлении рекомендаций по валидации [1]. Для обеспечения качества данных спонсоры должны обосновать адекватность критических элементов исследования и провести анализ чувствительности результатов к различной степени потенциальной неправильной классификации [5]. В руководстве также говорится о том, что алгоритмы или операционные определения для ключевых переменных должны быть продемонстрированы с использованием достаточно больших выборок, соответствующих методов отбора проб и разумных эталонов. Однако проект не содержит конкретных указаний относительно того, как определить достаточно большую выборку или соответствующие методы формирования выборки. В нём также отсутствуют рекомендации по использованию конкретных эталонных методов для того или иного этапа оценки качества. FDA утверждает, что можно проводить

исследования для подтверждения интересующих результатов, а также использовать результаты предыдущих исследований или публикаций, если они проводились на аналогичной популяции или приводились в аналогичном источнике данных. Однако точность алгоритмов идентификации случаев может варьироваться в разных популяциях, медицинских учреждениях, системах кодирования или календарных периодах времени, что может повлиять на их эффективность и переносимость в другие базы данных здравоохранения. Поэтому при применении алгоритмов в других условиях или базах данных исследователи должны учитывать, различается ли распространённость исходов и сопоставимы ли ограничения в доступе к лекарственным препаратам между валидируемой и исследуемой популяциями [1].

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Конфликт интересов.

Автор декларирует отсутствие конфликта интересов.

Участие авторов. Все соавторы подтверждают окончательный вариант статьи, ответственность за целостность всех частей статьи. Радаева К.С. — написание текста, редактирование, оформление статьи; Пчелинцев М.В. — редактирование, оформление статьи.

СВЕДЕНИЯ ОБ АВТОРАХ

Радаева Ксения Сергеевна — ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет имени И.П. Павлова», Санкт-Петербург, Российская Федерация

Автор, ответственный за переписку

e-mail: xenrada@gmail.com

 <https://orcid.org/0000-0003-2121-094X>

РИНЦ AuthorID: 1214830

Пчелинцев Михаил Владимирович — к. м. н., доцент кафедры клинической фармакологии и доказательной медицины ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет им. академика И. П. Павлова» Минздрава России, Санкт-Петербург, Российская Федерация

e-mail: mvpchelintsev@inbox.ru

 <https://orcid.org/0000-0002-3878-8265>

РИНЦ AuthorID: 417217

Литература/References

1. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) and Oncology Center for Excellence (OCE). Real-World Data: Assessing electronic health records and medical claims to support reg-

Выводы

Проект руководства отражает подходы FDA к стандартизации использования данных РКИ для принятия регуляторных решений и является необходимым и полезным инструментом для этого. Однако есть опасения, что рекомендуемые методы могут ограничить использование данных РКИ и не решить проблему неэффективности стандартных РКИ.

В руководстве предлагается применять риск-ориентированный подход к валидации и обеспечению качества, а также обеспечить большую прозрачность процедур использования данных. Также рекомендуется обосновывать адекватность неструктурированных данных в исследованиях РКИ. Это соответствует цели FDA по более эффективному и результативному использованию этих данных.

ADDITIONAL INFORMATION

Conflict of interests.

The author declares no conflict of interests.

Participation of authors. All authors are responsible for the integrity of all parts of the manuscript and approval of its final version.

Radaeva KS — text writing, editing; article design; Pchelintsev MV — editing; article design.

ABOUT THE AUTHORS

Kseniia S. Radaeva — First St. Petersburg State Medical University, named after I.P. Pavlov, St. Petersburg, Russian Federation

Corresponding author

e-mail: xenrada@gmail.com

 <https://orcid.org/0000-0003-2121-094X>

RSCI AuthorID: 1214830

Mikhail V. Pchelintsev — Cand. Sci. Med., Assistant of professor Department of Clinical Pharmacology and Evidence-Based Medicine, FSBEI HE named after I. P. Pavlov SPbSMU MOH Russia, St. Petersburg, Russian Federation

e-mail: mvpchelintsev@inbox.ru

 <https://orcid.org/0000-0002-3878-8265>

RSCI AuthorID: 417217

- ulatory decision- making for drug and biological products: Guidance for Industry, Draft Guidance. Sept 2021. (<https://www.fda.gov/media/152503/download>)
2. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, Bartels DB, Kahler KH, Bessette LG, Schneeweiss S. STaRT-RWE: structured template for

- planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021 Jan 12;372:m4856. doi: 10.1136/bmj.m4856. PMID: 33436424; PMCID: PMC8489282.
3. Ritchey ME, Girman CJ. Evaluating the Feasibility of Electronic Health Records and Claims Data Sources for Specific Research Purposes. *Ther Innov Regul Sci*. 2020 Nov;54(6):1296-1302. doi: 10.1007/s43441-020-00139-x. Epub 2020 May 7. PMID: 33258098.
4. Dreyer NA. Advancing a Framework for Regulatory Use of Real-World Evidence: When Real Is Reliable. *Ther Innov Regul Sci*. 2018 May;52(3):362-368. doi: 10.1177/2168479018763591. Epub 2018 Mar 19. PMID: 29714575; PMCID: PMC5944086.
5. Girman CJ, Ritchey ME, Lo Re V 3rd. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. *Pharmacoepidemiol Drug Saf*. 2022 Jul;31(7):717-720. doi: 10.1002/pds.5444. Epub 2022 May 3. PMID: 35471704; PMCID: PMC9320939.
6. Adamson B, Waskom M, Blarre A, et al. Approach to machine learning for extraction of real-world data variables from electronic health records. *Frontiers in Pharmacology*. 2023;14:1180962. DOI: 10.3389/fphar.2023.1180962. PMID: 37781703; PMCID: PMC10541019.
7. Howe CJ, Cole SR. Applying Quantitative Bias Analysis to Epidemiologic Data. *American Journal of Epidemiology*. 2009;170(10):1316–1317. <https://doi.org/10.1093/aje/kwp317>
8. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep*. 2014 Dec;1(4):175-185. doi: 10.1007/s40471-014-0027-z. PMID: 26085977; PMCID: PMC4465810.
9. Hurley C, Shiely F, Power J, Clarke M, Eustace JA, Flanagan E, Kearney PM. Risk based monitoring (RBM) tools for clinical trials: A systematic review. *Contemp Clin Trials*. 2016 Nov;51:15-27. doi: 10.1016/j.cct.2016.09.003. Epub 2016 Sep 15. PMID: 27641969. [myRWD](#)