

Методы обнаружения сигналов безопасности лекарств с использованием регулярно собираемых данных наблюдений в электронном здравоохранении: систематический обзор

Мотринчук А. Ш. ¹, Логиновская О. А. ^{1,2}, Колбатов В. П. ²

¹ – ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет имени академика И.П. Павлова» Министерства здравоохранения Российской Федерации, Санкт-Петербург, Российская Федерация

² – Компания Flex Databases LLC, Санкт-Петербург, Российская Федерация

Аннотация

Обнаружение сигналов является важным этапом в обнаружении побочных реакций на лекарственные препараты в пострегистрационном периоде. Увеличивается интерес к использованию регулярно собираемых данных в дополнение к спонтанным сообщениям об обнаружении сигналов безопасности лекарств. **Целью** этой работы является систематический обзор методов определения сигналов безопасности лекарственных средств с использованием регулярно собираемых медицинских данных. **Методология.** Был проведён систематический обзор в соответствии с рекомендациями PRISMA, протокол исследования зарегистрирован в PROSPERO. **Результат.** В обзор вошла 101 статья, среди которых было 39 методологических работ, 25 документов по оценке эффективности и 24 наблюдательных исследования. Методы включали: адаптацию методов, которые использовались при спонтанных сообщениях, традиционные эпидемиологические схемы, методы, специфичные для обнаружения сигналов с использованием реальных данных. Двадцать пять исследований оценивали эффективность методов, в 16 из них в качестве основного показателя использовалась площадь под кривой (AUC) для ряда положительных и отрицательных контролей. Воспроизводимость результатов оценки эффективности была ограничена из-за отсутствия прозрачности в отчётности и отсутствия «золотого стандарта».

Ключевые слова: сигналы; методы обнаружения сигналов безопасности лекарственных средств; данные реальной клинической практики

Для цитирования:

Мотринчук А. Ш., Логиновская О. А., Колбатов В. П. Методы обнаружения сигналов безопасности лекарств с использованием регулярно собираемых данных наблюдений в электронном здравоохранении: систематический обзор. *Реальная клиническая практика: данные и доказательства*. 2023;3(2):42-55. <https://doi.org/10.37489/2782-3784-myrd-35>

Поступила: 25 мая 2023 г. Принята: 30 мая 2023 г. Опубликовано: 30 августа 2023 г.

Methods for drug safety signal detection using routinely collected observational electronic health care data: a systematic review

Motrinchuk A. Sh. ¹, Loginovskaya O.A. ^{1,2}, Kolbatov V.P. ²

¹ – Federal State Budgetary Educational Institution of Higher Education "First St. Petersburg State Medical University named after Academician I.P. Pavlov" of the Ministry of Health of the Russian Federation, St. Petersburg, Russian Federation

² – Flex Databases LLC, St. Petersburg, Russian Federation

Abstract

Signal detection is a crucial step in the discovery of post-marketing adverse drug reactions. There is a growing interest in using routinely collected data to complement established spontaneous report analyses. **The aim.** This work aims to systematically review the methods for drug safety signal detection using routinely collected healthcare data and their performance, both in general and for specific types of drugs and outcomes. **Methodology.** We conducted a systematic review following the PRISMA guidelines, and registered a protocol in PROSPERO. **Results.** The review included 101 articles, among which there were 39 methodological works, 25 performance assessment papers, and 24 observational studies. Methods included adaptations from those used with spontaneous reports, traditional epidemiological designs, methods specific to signal detection with real-world data. More recently, implementations of machine learning have been studied in the literature. Twenty-five studies evaluated method performances, 16 of them using the area under the curve (AUC) for a range of positive and negative controls as their main measure. Despite the likelihood that performance measurement could vary by drug-event pair, only 10 studies reported

performance stratified by drugs and outcomes, in a heterogeneous manner. The replicability of the performance assessment results was limited due to lack of transparency in reporting and the lack of a gold standard reference set.

Keywords: signal detection; methods for drug safety signal detection; real-world data

For citations:

Motrinchuk ASH, Loginovskaya OA, Kolbatov VP. Methods for drug safety signal detection using routinely collected observational electronic health care data: a systematic review. *Real-World Data & Evidence*. 2023;3(2):42–55. <https://doi.org/10.37489/2782-3784-myrd-35>

Received: May 25, 2023. **Accepted:** May 30, 2023. **Published:** August 30, 2023.

Введение

На протяжении десятилетий спонтанные сообщения (*англ.* spontaneous reports; SRs) были основным методом выявления нежелательных лекарственных реакций (НР; *англ.* adverse drug reactions; ADR), которые не были выявлены в ходе клинических исследований [1], и они остаются таковыми, несмотря на общепризнанные ограничения [2, 3]. Растёт интерес к использованию данных реальной клинической практики (*англ.* real-world data; RWD), данным о жалобах пациентов и записям электронных медицинских карт (*англ.* electronic health records; EHRs). Их потенциал для обнаружения сигналов был признан надеждой на потенциально более быстрый и эффективный пострегистрационный надзор [4].

Цель систематического обзора: обновить список методов обнаружения сигналов безопасности лекарств с использованием регулярно собираемых данных; обобщить и сравнить эффективность методов; оценить эффективность каждого метода.

Систематический обзор проводился в соответствии с протоколом, зарегистрированным в PROSPERO (регистрационный номер CRD42021 267610). Поиск проводился в MEDLINE и EMBASE через OVID, Web of science, Scopus, PubMed и Библиотеку Кокрейна без ограничений на период до 13 июля 2021 года. Статьи отбирались по названию и аннотациям, проводилась полнотекстовая оценка подходящих статей.

Извлечение данных. Данные извлекались на основе фармакоэпидемиологического контрольного списка (*англ.* Pharmacoepidemiology Checklist) [5], внимание было сосредоточено на деталях методов: дизайне, статистических результатах; воздействиях, исходах и эффективности методов.

Также были оценены риски предвзятости и недоразумений, оценивалась целесообразность тестирования НР и степень, в которой база данных фиксирует результаты.

Анализ данных. Были представлены характеристики включённых исследований и методы определения сигналов безопасности лекарственных средств, описаны методы обнаружения сигналов безопасности лекарственных средств с использова-

нием RWD и подсчитано количество раз, когда они использовались. Эффективность этих методов оценивалась с использованием показателей, представленных в литературе, как в целом по всем парам «препарат-результат», так и по препарату и исходу, когда это было доступно.

Результат

Было проверено 1765 наименований и рефератов. После применения критериев включения и исключения 351 статья была классифицирована как потенциально приемлемая. Из них в обзор были включены 116 соответствующих исследований, в том числе 101 оригинальное исследование и 15 обзоров.

Из включённых исследований 38,6% описывали исключительно методы (см. табл. 1), 24,8% касались оценки эффективности и 23,8% были наблюдательными исследованиями без оценки эффективности. Среди исследований 5,9% из них сравнивали использование электронных медицинских карт (ЭМК; *англ.* electronic health records; EHRs) и спонтанных сообщений (СС; *англ.* spontaneous reports; SRs) для обнаружения сигналов [1, 6-10]. Остальные 6,9% включали недавнюю докторскую диссертацию [11], два комментария [12, 13], исследование, направленное на установление эталонного стандарта для обнаружения сигналов [14] и 3 исследования, рассматривающие значимость обнаружения сигналов результатов [15, 16, 17]. В большинстве исследований (88,1%) использовались традиционные данные ЭМК, в то время как в 6,9% использовались ненормальные лабораторные результаты [18, 19-24] и набор данных только для рецептов (5%), где рецепты используются в качестве прокси для постановки диагноза [25-29]. Поскольку целью систематического обзора было выявление оригинальных исследований, поэтому любые обзорные статьи, которые были определены в рамках области, использовались только для предоставления потенциальных дополнительных оригинальных исследовательских публикаций для включения, и их содержание не было извлечено [4, 30, 15, 31-41].

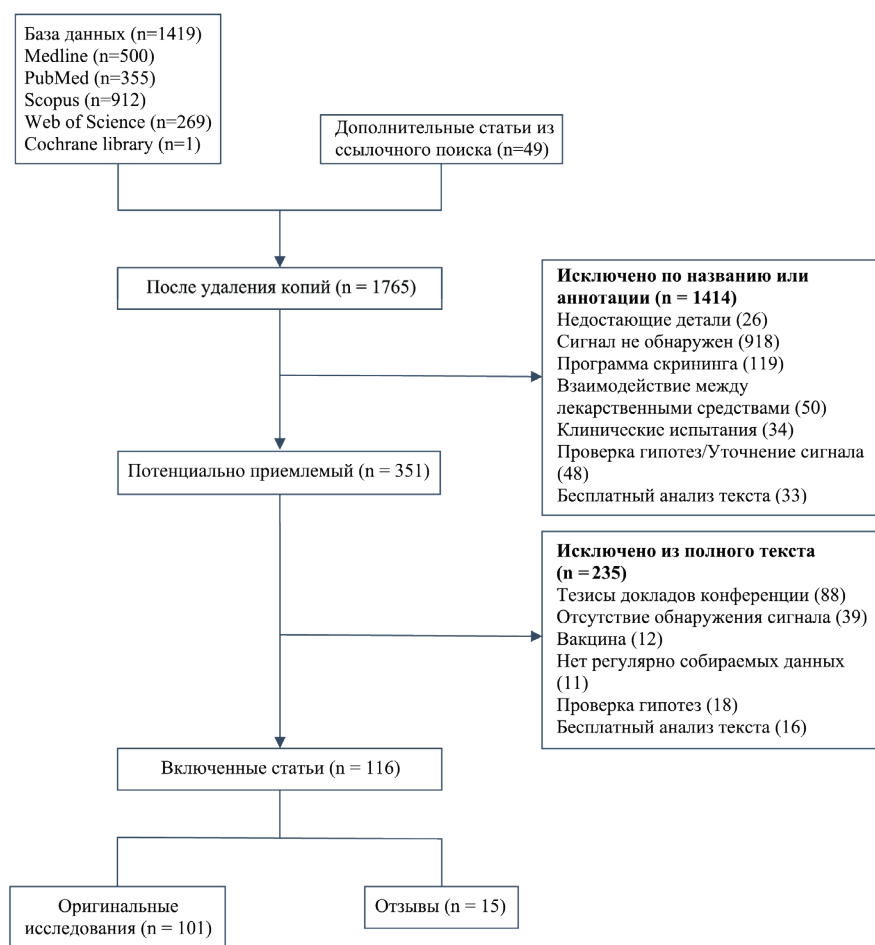


Рис. 1. Блок-схема включения статей

Таблица 1. Краткие характеристики публикаций, включённых в обзор			
Характеристика	Количество публикаций	Характеристика	Количество публикаций
Основная цель статьи		Подход	
Описание метода	39 (38,6%)	На основе результатов	40 (39,6%)
Оценка эффективности	25 (24,8%)	На основе воздействия	26 (25,7%)
Сравнение данных между EHRs 1 и SRs2	6 (5,9%)	Указаны как лекарства, так и результаты	6 (5,9%)
Применение метода без оценки эффективности	24 (23,8%)	Все лекарства и результаты в базе данных (базах данных)	4 (4,0%)
Другое ³	7 (6,9%)	Нет (чисто методологический)	25 (24,8%)
Расположение данных		Тип данных, используемых методом	
США	43 (42,6%)	Метод, основанный на кодах рецептов и диагнозов	89 (88,1%)
Европа	37 (36,7%)	Метод, основанный только на данных о рецепте	5 (5,0%)
Азия / Австралия	18 (17,8%)	Метод, основанный на сравнении результатов лабораторных испытаний	7 (6,9%)
Иностраным студентам	3 (3,0%)		
Примечания: EHRs – записи электронных медицинских карт (electronic health records); 2SRs – спонтанные сообщения (spontaneous reports); 3Другое – докторская диссертация, комментарии, справочный стандарт			

Методы обнаружения сигналов безопасности лекарственных средств. Широкий спектр методов был описан во включённых исследованиях и обобщён в таблицах 2 и 3 в соответствии с классификацией, используемой *Arnaud M. и соавт.* (2017 г.) [30]. В целом, литература сосредоточена на адаптации методов анализа диспропорциональности для обнаружения сигналов и реализации традиционных эпидемиологических схем. Другие методы, использующие байесовские сетевые модели, параметры формы

Weibull или тесты отношения правдоподобия, были предложены в методических статьях, но использовались в единственном наблюдательном исследовании или вообще не использовались, поэтому не включены в следующие таблицы [42, 43]. Распределение Weibull в теории вероятностей — двухпараметрическое семейство абсолютно непрерывных распределений. Оно моделирует широкий диапазон случайных величин, в основном по характеру времени до сбой или времени между событиями.

Таблица 2. Количество раз, когда каждый метод, примененный дважды или более, использовался во всех публикациях обзора

Метод	Количество работ, использующих дизайн ^a	Метод	Количество работ, использующих дизайн ^a
Анализ диспропорциональности		Временная ассоциация	
PRR	9 (17,3%)	Обнаружение временных закономерностей	10 (50,0%)
ROR	8 (15,4%)	MUTARA/HUNT	6 (30,0%)
BCPNN	9 (17,3%)	Fuzzy-based logic	4 (20,0%)
GPS / MGPS	6 (11,5%)	Промежуточный итог	20 (100,0%)
LGPS / LEOPARD	12 (23,1%)	Анализ симметрии последовательности	6 (100,0%)
Другое	8 (15,4%)	Последовательное тестирование	
Промежуточный итог	52 (100,0%)	MaxSPRT	4 (66,7%)
Традиционные эпидемиологические схемы		CSSP	2 (33,3%)
Серия самоконтролируемых случаев (SCCS)	15 (34,1%)	Промежуточный итог	6 (100,0%)
Самоконтролируемая когорта (SCC)	5 (11,4%)	Tree-based scan statistic ²	9 (100,0%)
Новая когорта пользователей (NUC)	5 (11,4%)	Другие проекты, включая машинное обучение	13 (100,0%)
Случай-контроль	13 (29,5%)	Результаты лабораторных исследований	9 (100,0%)
Случай-кроссовер	3 (6,8%)	Методы, доступные только по рецепту врача	5 (100,0%)
Случай-популяция	3 (6,8%)		
Промежуточный итог	44 (100,0%)		

Сокращения: PRR — Пропорциональный коэффициент отчётности (Proportional Reporting Ratio; PRR); ROR — Отношение шансов отчётности (Reporting Odds Ratio; ROR); BCPNN — Байесовская нейронная сеть распространения достоверности (Bayesian Confidence Propagation Neural Network; BCPNN); GPS — Гамма-пуассоновский анализатор (Gamma Poisson Shrinker; GPS); MGPS — Многопозиционный гамма-пуассоновский анализатор (Multi-Item Gamma Poisson Shrinker; MGPS); LGPS — Лонгитудинальный гамма-пуассоновский анализатор (Longitudinal Gamma Poisson Shrinker; LGPS); LEOPARD — Лонгитудинальная оценка наблюдаемых профилей нежелательных явлений, связанных с лекарствами (Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs; LEOPARD); TARs — Правила времени ассоциации (Temporal Association Rules; Tars); MUTARA — Анализ неожиданных правил временной ассоциации (Temporal Association Rules; TARs) с учётом предшествующего (Mining Unexpected Temporal Association Rules (TARs) Given the Antecedent; MUTARA); HUNT — Выделение неожиданных игнорирующих TARs (Temporal Association Rules; Tars), игнорирующих TARs (Temporal Association Rules; TARs) — (Highlighting Unexpected TARs Neglecting TARs; HUNT); Нечёткая логика это форма многозначной логики, в которой значением истинности переменных может быть любое действительное число в диапазоне от 0 до 1; ²TreeScan™ — это бесплатное программное обеспечение для интеллектуального анализа данных, которое реализует статистику сканирования на основе дерева, метод интеллектуального анализа данных, который одновременно ищет избыточный риск в любой из большого количества отдельных ячеек в базе данных, а также в группах тесно связанных ячеек, с учётом многократного тестирования, присущего большому количеству перекрывающихся оцениваемых групп; MaxSPRT — Тест на максимальное последовательное отношение вероятностей (Maximized Sequential Probability Ratio Test; MaxSPRT); CSSP — Процедура условной последовательной выборки (Conditional Sequential Sampling Procedure; CSSP); SCCS — Серия самоконтролируемых случаев (Self-controlled case series; SCCS); SCC — Самоконтролируемая когорта (Self-controlled cohort; SCC); NUC — Новая когорта пользователей (New user cohort; NUC); CC — Случай контроль (Case — control; CC); Случай кроссовер — Case-crossover; Случай-популяция — Case-population; ^a Исследования, в которых изучалось более одного метода, были подсчитаны для каждого из рассмотренных методов, так что общее количество статей в этой таблице не соответствует количеству включённых исследований.

Таблица 3. Обзор методов определения сигналов безопасности лекарственных средств с использованием RWD, упомянутых или применённых более чем в одном исследовании, о чём сообщается во включённых статьях				
Метод	Общая концепция	Преимущества	Недостатки	Дополнительные комментарии
Анализ диспропорциональности (DR)				
PRR и ROR	Первоначально применялся к SRs для определения степени диспропорциональности между сообщениями о состоянии для данного лекарственного средства [44].	Легко рассчитать и реализовать.	Нестабильность с небольшим количеством событий и большими доверительными интервалами приводит к высоким ложноположительным показателям для редких событий [39]. Предназначен для перекрестных данных, поэтому общее количество случаев, доступных в базах данных EHRs, не используется.	
Информационный компонент (BCPNN)	Полностью байесовский (заранее определённый) метод диспропорциональности, использующий сокращение наблюдаемых к ожидаемым (O / E) баллов для вычисления информационного компонента (Information Component, IC) [45]. Иногда реализуется в нейронной сети (BCPNN), где IC — это вес	Устраняет проблему высокого уровня ложноположительных результатов методов на основе SRs — за счёт большего сокращения, когда мало данных поддерживают снижение ложных случайных результатов [44].	Как указано выше, кроме того, более сложная интерпретируемость из-за влияния байесовского сжатия добавляет дополнительную сложность к интерпретации результатов.	
GPS / MGPS	GPS — это ещё один Bayesian метод уменьшения наблюдаемого / ожидаемого соотношения. MGPS — это расширение для анализа взаимодействий между лекарствами и лекарственными препаратами [39].	То же, что и BCPNN, эмпирическая усадка, поэтому сила усадки адаптируется к конкретному набору данных.	В основном такие же, как BCPNN, и влияние переменной усадки. Требует больших вычислительных затрат.	
LGPS	Адаптация метода GPS к продолжительным данным. LGPS сравнивает частоту возникновения исходов в период риска воздействия с фоновым показателем для всех людей.	Использовался в сочетании с методом LEOPARD для теоретической обработки протопатических и позитивных искажений. LEOPARD сравнивает частоту назначения лекарств до событий и после [46].	Влияние байесовского подхода, как указано выше.	
Традиционные эпидемиологические схемы				
Серия самоконтролируемых случаев (Self-controlled case series, SCCS)	Сравнение частоты событий во время облучения и неэкспонированного времени у одного и того же индивидуума [47].	Устраняет погрешности, не зависящие от времени, что выгодно, когда исходные ковариаты не изменены с достаточной точностью [47].	Точная датировка результатов имеет решающее значение, лучше всего это относится к периодическим воздействиям и преходящим или острым состояниям [48].	Было предложено несколько дополнительных модификаций, но они ещё не реализованы в исследованиях обнаружения сигналов [49, 50, 51].

Таблица 3. Обзор методов определения сигналов безопасности лекарственных средств с использованием RWD, упомянутых или применённых более чем в одном исследовании, о чём сообщается во включённых статьях				
Метод	Общая концепция	Преимущества	Недостатки	Дополнительные комментарии
Традиционные эпидемиологические схемы				
Самоконтролируемая когорта (Self-controlled cohort, SCC)	Включающий как когорту, так и саморегулируемую коррекцию.	Использует внешнюю контрольную группу для корректировки оставшихся изменений во времени смещения после самоконтролируемого компонента [13].	Чувствителен к различиям между риском и контрольным периодом, специфичным для группы, подвергшейся воздействию, таким как прототипический уклон [30].	
Новая когорта пользователей (New user cohort, NUC)	Сравнивает частоту событий в когорте, вводящей интересующий препарат, с когортой, не вводящей этот препарат [52, 53].	Широко применимый метод [57], подход активного компаратора для устранения недоразумений по показателям [54]. Вычисление абсолютного риска событий [30].	Более высокие вычислительные требования, чем методы с самоконтролем. Между людьми смещение. Необходимость в предопределённых компараторах, поскольку их соответствие трудно оценить в реальных условиях.	
Метод случай – контроль (Case – control, CC)	Сравнивает частоту воздействия «пациентов», которые испытали результат, с частотой сопоставимых контрольных, которые не испытали результата [54].	Эффективен для редких исходов [55].	Возможно, сложный выбор контроля, подверженный путанице между людьми. Высокие требования к вычислениям, самый медленный метод в ОМОР [56].	
Случай-кроссовер	Используется сравнение воздействия внутри человека в период наблюдения по сравнению с контрольным периодом [57].	Аналогично другим самоконтролируемым конструкциям.	При наличии тенденции времени воздействия возможны отклонения [57].	
Случай-популяция	Аналогично схеме «случай – контроль», но в качестве контрольной группы используется всё население [58].	Повышенная статистическая точность по сравнению с случай-контролем [58].	Более высокие вычислительные требования.	
Временная ассоциация				
Обнаружение временных закономерностей (Temporal pattern discovery, TPD)	На основе отношения наблюдаемого к ожидаемому. DR, добавляя сравнение времени воздействия с контрольным периодом до первого воздействия препарата для выявления временной ассоциации [59]. Аналогично GPS.	Bayesian shrinkage (байесовская усадка) для защиты от ложных ассоциаций, особенно для редких случаев [60]. Использование хронографов для графической интерпретации, возможность обнаружения большего разнообразия закономерностей	Сложнее выявить закономерности, связанные с очень редкими пространственными событиями.	

Таблица 3. Обзор методов определения сигналов безопасности лекарственных средств с использованием RWD, упомянутых или применённых более чем в одном исследовании, о чём сообщается во включённых статьях				
Метод	Общая концепция	Преимущества	Недостатки	Дополнительные комментарии
MUTARA и HUNT	MUTARA — это алгоритм интеллектуального анализа данных, который ищет неожиданные временные правила ассоциации, находя любые события, неожиданно возникающие после интересующих препаратов в течение заранее определённого периода риска, и включая общие события, которые вряд ли будут ADR, используя контрольный период для сопоставления списка важных ADRs. Сбой лечения может привести к регистрации событий после интересующего препарата, вызывая ложные сигналы. HUNT, модифицированная версия MUTARA, ранжирует сигналы с учётом неудачного лечения [61, 62].	Поскольку MUTARA ориентирована на события, она должна обладать теоретической способностью сигнализировать о нечастых ADRs [63].	MUTARA было трудно отличить побочные эффекты от неудач лечения. [64]. HUNT был разработан для учёта этих неудач в лечении.	
Fuzzy-based logic with causal-leverage measure	Это вычислительная модель, называемая моделью принятия решений с нечётким распознаванием. В нём используются такие сигналы, как временная связь между лекарствами и результатами, повторный вызов, снятие вызова, чтобы выявить потенциальные причинно-следственные связи у каждого человека, подвергнутого воздействию наркотиков. Затем потенциальные причинно-следственные связи включаются в новые показатели (причинно-следственная связь и обратная причинно-следственная связь), которые используются для ранжирования пары лекарство-результат и удаления ложных сигналов [65, 66].	Попытайтесь ранжировать важные сигналы путём количественной оценки степени связи пары лекарство-результат и устранения фонового шума (ложных сигналов) с использованием новых мер. Разработан специально для нечастых ассоциаций [67].	Параметр, например, период опасности для определения результатов, и нечёткие переменные (например, временная ассоциация, повторный вызов и повторный вызов) могут быть сложными для оптимизации.	
Анализ симметрии последовательности (Sequence Symmetry Analysis, SSA)	Исследуйте последовательность событий, связанных с введением препарата. Тестирование на асимметрию на основе нулевой гипотезы о том, что при отсутствии связи можно ожидать симметричного распределения результатов до и после начала приёма лекарства [35].	Вычислительно эффективный, устойчивый к помехам, стабильный с течением времени [35]. Возможно оценить все лекарства и все результаты в базе данных [68].	Несимметричный шаблон не обязательно указывает на сигнал [40]. Сортировка, необходимая для поиска потенциально интересных ассоциаций, большое количество ложноположительных результатов. Предполагается наличие соответствующей единой точки симметрии.	В литературе в основном применяется в крупномасштабных условиях [69–71].

Таблица 3. Обзор методов определения сигналов безопасности лекарственных средств с использованием RWD, упомянутых или применённых более чем в одном исследовании, о чём сообщается во включённых статьях

Метод	Общая концепция	Преимущества	Недостатки	Дополнительные комментарии
Последовательное тестирование: MaxSPRT — Тест на максимальное последовательное отношение вероятностей (Maximized Sequential Probability Ratio Test, MaxSPRT) CSSP — Процедура условной выборочной (Conditional Sequential Sampling Procedure, GSSP) LLMP — Логарифмическая линейная модель для данных Пуассона (Log Linear Model for Poisson Data, LLMP)	Набор методов для наблюдения в режиме, близком к реальному времени, или проспективного наблюдения, основанного на повторном тестировании для выявления ассоциаций. Пациенты добавляются постепенно [72]. MaxSPRT — это статистический тест, учитывающий многократное тестирование [73, 74]. Метод CSSP использует условную вероятность того, что результат будет более экстремальным, чем наблюдаемая частота событий, и стратегически фиксирует популяцию населения [76]. LLMP представляет собой логарифмически-линейный метод, основанный на Пуассоне.	Способность обрабатывать редкие события [74]. Наблюдение в режиме реального времени делает возможным раннее выявление новых ADR, повышая своевременность обнаружения сигнала [75]. CSSP был разработан для обработки хронических воздействий [76].	MaxSPRT требует большого объема исторических данных для получения точных оценок ожидаемого количества событий. Эта проблема решается методом CSSP [77]. CSSP может испытывать трудности с сохранением ошибки типа I при частом тестировании или многочисленных областях [76]. Ограничено для обработки непрерывных помех [72].	
Tree-based scan statistic	События в области здравоохранения классифицируются в древовидной иерархической системе на основе кодов диагнозов на разных уровнях. Метод интеллектуального анализа данных ищет избыточный риск в отдельных ячейках дерева, а также в тесно связанных ячейках [78]. Проверка гипотез выполняется с использованием теста отношения правдоподобия [79].	Одновременное сканирование сигналов на разных уровнях детализации. Минимальное предвзятое предположение о типе интересующего результата для здоровья [79]. Корректировка для проверки нескольких гипотез [79].	Потенциальная опасность ложноположительных сигналов из-за контроля ошибки типа I [80].	Гибко применяется с новым пользователям, дизайном активного компаратора, методами оценки склонности и самоконтролируемыми конст-рукциями [81, 82].
Другие подходы, основанные на машинном обучении (Other machine learning (ML) based approaches, ML)	Были предложены различные подходы. Некоторые стремятся обучить классификатор целому ряду положительных и отрицательных контролей [83]. Другие используют стратегии, основанные на данных, для поиска очевидных выбросов [58, 83].	Способность работать с большими, многомерными и разреженными данными [85], хотя производительность вычислений может быть снижена.	Необходимость в большом обучающем наборе данных. Неспециалистам в области ОД труднее реализовать. Предполагается на способность этапа обучения работать с зашумленными медицинскими данными.	Используется различными способами для обнаружения сигналов [18, 29, 83, 27, 87–92] включая расчёт оценки склонности [92] или в целом подход с использованием соотношений причинно-следственной связи Брэдфорда Хилла [84].
Другое Анализ симметрии последовательности рецептов (PSSA) Методы с использованием лабораторных результатов	Вариация SSA, в которой лекарства принимаются в качестве прокси для интересующих результатов в отношении здоровья – (Health Outcomes of Interest; HOIs) [25, 28]. Методы сравнивали аномальные или экстремальные лабораторные результаты до и после воздействия лекарства [24].			ML также использовался только для данных о рецептах [27, 29]. ML использовался для объединения сравнения результатов экстремальных лабораторных тестов (CERT), сравнения коэффициента экстремальных отклонений (CLEAR) и алгоритмов назначения рецептов в зависимости от клинических событий (PASC) в одном значении [18].

Из таблицы 3 видно, что ни один метод теоретически не может одинаково хорошо работать для всех лекарств и результатов (например, некоторые методы больше подходят для острых или редких исходов).

Выводы

В целом, методы с самоконтролем, как правило, достигали более высоких AUC (*англ.* area under the curve; AUC), чем другие методы, включая методы случай–контроль и диспропорциональность. Методы HDPS (*англ.* High Dimensional Propensity Score; HDPS) и TPD (*англ.* Temporal pattern discovery; TPD) также достигли более высоких AUC, как в среднем, так и в определённых подгруппах. Однако они не оценивались во многих исследованиях, и время их выполнения было больше, чем для других методов [63]. Методы диспропорциональности, широко используемые при спонтанных сообщениях, по-видимому, не способны различать положительный и отрицательный контроль, поскольку они сообщали о AUCs, близких к случайному угадыванию [91]. Хотя статистика сканирования на основе дерева не подвергалась официальной оценке эффективности, она смогла уловить известные сигналы [78] и может быть полезна для оценки результатов на разных уровнях детализации, особенно при подходе, основанном на лекарствах. Аналогичным образом, эффективность машинного обучения (*англ.* machine learning; ML), оценивалась неоднородно, но предварительные результаты подчёркивают её потенциал для обнаружения сигналов.

Показатели эффективности, как правило, сообщались в среднем по всем препаратам и результатам в контрольном наборе, даже несмотря на то, что каждый проект эпидемиологического исследования работает лучше с некоторыми типами воздействия и результатов, чем с другими. Таким образом, сообщаемые общие показатели могут скрывать особенно высокие или слабые показатели для наборов аналогичных комбинаций воздействия и результатов.

Эффективность в основном оценивалась и представлялась с помощью AUC, который является еди-

ным показателем и не включает такие аспекты, как предвзятость [92]. Предполагается, что каждый порог чувствительности и специфичности одинаково важен, что на практике не относится к обнаружению сигнала, и, хотя на практике объективно, может дать неверное представление о значении обнаружения сигнала. О других мерах сообщалось редко, и их нельзя было сравнить.

Большинство статей были неспецифичны в выборе результата и его характеристике или были сосредоточены на быстром начале НР. Наилучший метод выявления долгосрочных НР, если таковой имеется на сегодняшний день, остаётся недостаточно изученным и, следовательно, неясным. В этой области необходима дальнейшая работа, поскольку регулярно собираемые данные могут иметь большое преимущество в регистрации долгосрочных результатов по сравнению со спонтанными сообщениями. Поскольку они могут возникать спустя годы после воздействия, очевидно, что ещё более сложной проблемой обнаружения сигналов является связь результата с воздействием лекарственного препарата со спонтанным сообщением.

Необходимы дальнейшие исследования эффективности методов для конкретных типов лекарств и результатов с упором на сильные и слабые стороны каждого метода. Необходима также полная отчётность об эффективности для отдельных пар или подгрупп пар «лекарство–результат». Хотелось бы больше видеть прямых сравнений методов для более широкого спектра пар «лекарство–результат», включая результаты с более медленным началом. Поскольку все эталонные наборы имеют присущие им ограничения, поощряется разработка множества разнообразных эталонных наборов, общедоступных для повторного использования. В идеале можно было бы разработать общие и доступные коды, которые могут быть реализованы в любой базе данных, с использованием общих моделей данных. Хотелось бы увидеть результаты по своевременности обнаружения сигналов при RWD, которые были исследованы только в одной статье, включённой в этот обзор.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Конфликт интересов.

Авторы декларируют отсутствие конфликта интересов.

Участие автора. Мотринчук А. Ш. — написание текста; Логиновская О.А. — редактирование, оформление статьи; Колбатов В.П. — редактирование, оформление статьи.

ADDITIONAL INFORMATION

Conflict of interests.

The authors declare no conflict of interest.

Participation of author. Motrinchuk AS — text writing; Loginovskaya OA — editing, article design; Kolbatov VP — editing, article design.

СВЕДЕНИЯ ОБ АВТОРАХ

Мотринчук Айтэн Шерифовна — Ординатор кафедры клинической фармакологии и доказательной медицины ФГБОУ ВО ПСПбГМУ им. акад. И. П. Павлова Минздрава России, Санкт-Петербург, Российская Федерация

Автор, ответственный за переписку

e-mail: aitesha555@yandex.ru

 <https://orcid.org/0000-0003-2863-792X>

Логиновская Ольга Александровна — Директор по качеству и корпоративному развитию Flex Databases; ассистент кафедры клинической фармакологии и доказательной медицины ФГБОУ ВО ПСПбГМУ им. акад. И. П. Павлова Минздрава России, Санкт-Петербург, Российская Федерация

e-mail: olga.loginovskaya@flexdatabases.com

 <https://orcid.org/0000-0003-3683-1300>

Колбатов Владимир Павлович — Ведущий специалист по сигналам

e-mail: vladimir.kolbatov@flexdatabases.com

 <https://orcid.org/0000-0002-6345-2341>

Литература/References

1. Patadia VK, Coloma P, Schuemie MJ, et al. Using real-world healthcare data for pharmacovigilance signal detection — the experience of the EU-ADR project. *Expert Rev Clin Pharmacol*. 2015 Jan;8(1):95-102. doi: [10.1586/17512433.2015.992878](https://doi.org/10.1586/17512433.2015.992878).
2. CIOMS. Working Group VIII. Practical Aspects of Signal Detection in Pharmacovigilance. Council for International Organizations of Medical Sciences (CIOMS); 2010.
3. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009 Jun;18(6):427-36. doi: [10.1002/pds.1742](https://doi.org/10.1002/pds.1742).
4. Moore TJ, Furberg CD. Electronic Health Data for Postmarket Surveillance: A Vision Not Realized. *Drug Saf*. 2015 Jul;38(7):601-10. doi: [10.1007/s40264-015-0305-9](https://doi.org/10.1007/s40264-015-0305-9).
5. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018 Nov 14;363:k3532. doi: [10.1136/bmj.k3532](https://doi.org/10.1136/bmj.k3532).
6. Kim J, Kim M, Ha JH, et al. Signal detection of methylphenidate by comparing a spontaneous reporting database with a claims database. *Regul Toxicol Pharmacol*. 2011 Nov;61(2):154-60. doi: [10.1016/j.yrtph.2011.03.015](https://doi.org/10.1016/j.yrtph.2011.03.015).
7. A Wahab I, Pratt NL, Kalisch LM, Roughead EE. Comparing time to adverse drug reaction signals in a spontaneous reporting database and a claims database: a case study of rofecoxib-induced myocardial infarction and rosiglitazone-induced heart failure signals in Australia. *Drug Saf*. 2014 Jan;37(1):53-64. doi: [10.1007/s40264-013-0124-9](https://doi.org/10.1007/s40264-013-0124-9).
8. Repts J, Feyereisl J, Garibaldi J, et al. Investigating the detection of adverse drug events in a UK general practice electronic health-care database. Paper presented at: UKCI 2011 — Proceedings of the 11th UK Workshop on Computational Intelligence; 2011; Manchester UK, 167–173. doi: [10.48550/arXiv.1307.1078](https://doi.org/10.48550/arXiv.1307.1078).
9. Pacurariu AC, Straus SM, Trifirò G, et al. Useful Interplay Between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection. *Drug Saf*. 2015 Dec;38(12):1201-10. doi: [10.1007/s40264-015-0341-5](https://doi.org/10.1007/s40264-015-0341-5).
10. Patadia VK, Schuemie MJ, Coloma P, et al. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. *Int J Clin Pharm*. 2015 Feb; 37(1):94-104. doi: [10.1007/s11096-014-0044-5](https://doi.org/10.1007/s11096-014-0044-5).
11. Demailly R. Détection automatisée de signaux en pharmacovigilance chez la femme enceinte à partir de bases médico-administratives. HAL; 2021.
12. Schuemie MJ. Safety surveillance of longitudinal databases: further methodological considerations. *Pharmacoepidemiol Drug Saf*. 2012 Jun;21(6):670-2; author reply 673-5. doi: [10.1002/pds.3259](https://doi.org/10.1002/pds.3259).
13. Norén GN, Hopstadius J, Bate A, Edwards IR. Safety surveillance of longitudinal databases:

ABOUT THE AUTHORS

Motrinchuk Aiten S. — resident of the department of Clinical Pharmacology and Evidence-Based Medicine Pavlov First Saint Petersburg State Medical University, St. Petersburg, Russian Federation

Corresponding author

e-mail: aitesha555@yandex.ru

 <https://orcid.org/0000-0003-2863-792X>

Loginovskaya Olga A. — Director of Quality and Corporate Development Flex Databases; Assistant of the Department Of Clinical Pharmacology and Evidence-Based Medicine Pavlov First Saint Petersburg State Medical University, St. Petersburg, Russian Federation

e-mail: olga.loginovskaya@flexdatabases.com

 <https://orcid.org/0000-0003-3683-1300>

Kolbatov Vladimir P. — The best signal specialist on Flex Databases

e-mail: vladimir.kolbatov@flexdatabases.com

 <https://orcid.org/0000-0002-6345-2341>

- methodological considerations. *Pharmacoepidemiol Drug Saf.* 2011 Jul;20(7):714-7. doi: [10.1002/pds.2151](https://doi.org/10.1002/pds.2151).
14. Coloma PM, Avillach P, Salvo F, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Saf.* 2013 Jan;36(1):13-23. doi: [10.1007/s40264-012-0002-x](https://doi.org/10.1007/s40264-012-0002-x).
 15. Gruber S, Chakravarty A, Heckbert SR, et al. Design and analysis choices for safety surveillance evaluations need to be tuned to the specifics of the hypothesized drug-outcome association. *Pharmacoepidemiol Drug Saf.* 2016 Sep;25(9):973-81. doi: [10.1002/pds.4065](https://doi.org/10.1002/pds.4065).
 16. Madigan D, Ryan PB, Schuemie M. Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. *Ther Adv Drug Saf.* 2013 Apr;4(2):53-62. doi: [10.1177/2042098613477445](https://doi.org/10.1177/2042098613477445).
 17. Madigan D, Stang P, Berlin J, et al. A systematic statistical approach to evaluating evidence from observational studies. *Annu Rev Stat Appl.* 2014;1:11-39. doi: [10.1146/annurev-statistics-022513-115645](https://doi.org/10.1146/annurev-statistics-022513-115645).
 18. Jeong E, Park N, Choi Y, Park RW, Yoon D. Machine learning model combining features from algorithms with different analytical methodologies to detect laboratory-event-related adverse drug reaction signals. *PLoS One.* 2018 Nov 21;13(11):e0207749. doi: [10.1371/journal.pone.0207749](https://doi.org/10.1371/journal.pone.0207749). Erratum in: *PLoS One.* 2019 Apr 9;14(4):e0215344.
 19. Fan DF, Yu YC, Ding XS, et al. Exploring the drug-induced anemia signals in children using electronic medical records. *Expert Opin Drug Saf.* 2019 Oct;18(10):993-999. doi: [10.1080/14740338.2019.1645832](https://doi.org/10.1080/14740338.2019.1645832).
 20. Lee S, Choi J, Kim HS, et al. Standard-based comprehensive detection of adverse drug reaction signals from nursing statements and laboratory results in electronic health records. *J Am Med Inform Assoc.* 2017 Jul 1;24(4):697-708. doi: [10.1093/jamia/ocw168](https://doi.org/10.1093/jamia/ocw168).
 21. Wei R, Jia LL, Yu YC, et al. Pediatric drug safety signal detection of non-chemotherapy drug-induced neutropenia and agranulocytosis using electronic healthcare records. *Expert Opin Drug Saf.* 2019 May;18(5):435-441. doi: [10.1080/14740338.2019.1604682](https://doi.org/10.1080/14740338.2019.1604682).
 22. Mansour A, Ying H, Dews P, et al. Fuzzy rule-based approach for detecting adverse drug reaction signal pairs. Paper presented at: 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2013). *The authors — Published by Atlantis Press.* 2013;384-391. doi: [10.2991/eusflat.2013.60](https://doi.org/10.2991/eusflat.2013.60).
 23. Park MY, Yoon D, Lee K, et al. A novel algorithm for detection of adverse drug reaction signals using a hospital electronic medical record database. *Pharmacoepidemiol Drug Saf.* 2011 Jun;20(6):598-607. doi: [10.1002/pds.2139](https://doi.org/10.1002/pds.2139).
 24. Tham MY, Ye Q, Ang PS, et al. Application and optimisation of the Comparison on Extreme Laboratory Tests (CERT) algorithm for detection of adverse drug reactions: Transferability across national boundaries. *Pharmacoepidemiol Drug Saf.* 2018 Jan;27(1):87-94. doi: [10.1002/pds.4340](https://doi.org/10.1002/pds.4340).
 25. Lai EC, Hsieh CY, Kao Yang YH, Lin SJ. Detecting potential adverse reactions of sulpiride in schizophrenic patients by prescription sequence symmetry analysis. *PLoS One.* 2014 Feb 27;9(2):e89795. doi: [10.1371/journal.pone.0089795](https://doi.org/10.1371/journal.pone.0089795).
 26. Zhan C, Roughead E, Liu L, et al. A data-driven method to detect adverse drug events from prescription data. *J Biomed Inform.* 2018 Sep;85:10-20. doi: [10.1016/j.jbi.2018.07.013](https://doi.org/10.1016/j.jbi.2018.07.013).
 27. Zhan C, Roughead E, Liu L, et al. Detecting potential signals of adverse drug events from prescription data. *Artif Intell Med.* 2020 Apr;104:101839. doi: [10.1016/j.artmed.2020.101839](https://doi.org/10.1016/j.artmed.2020.101839).
 28. Pratt N, Chan EW, Choi NK, et al. Prescription sequence symmetry analysis: assessing risk, temporality, and consistency for adverse drug reactions across datasets in five countries. *Pharmacoepidemiol Drug Saf.* 2015 Aug;24(8):858-64. doi: [10.1002/pds.3780](https://doi.org/10.1002/pds.3780).
 29. Hoang T, Liu J, Roughead E, et al. Supervised signal detection for adverse drug reactions in medication dispensing data. *Comput Methods Programs Biomed.* 2018 Jul;161:25-38. doi: [10.1016/j.cmpb.2018.03.021](https://doi.org/10.1016/j.cmpb.2018.03.021).
 30. Arnaud M, Bégaud B, Thurin N, et al. Methods for safety signal detection in healthcare databases: a literature review. *Expert Opin Drug Saf.* 2017 Jun;16(6):721-732. doi: [10.1080/14740338.2017.1325463](https://doi.org/10.1080/14740338.2017.1325463).
 31. Karimi S, Wang C, Metke-Jimenez A, et al. Text and data mining techniques in adverse drug reaction detection. *ACM Comput Surv.* 2015;47(4):1-39. doi: [10.1145/2719920](https://doi.org/10.1145/2719920).
 32. Kaguelidou F, Durrieu G, Clavenna A. Pharmacoepidemiological research for the development and evaluation of drugs in pediatrics. *Therapie.* 2019 Apr;74(2):315-324. doi: [10.1016/j.therap.2018.09.077](https://doi.org/10.1016/j.therap.2018.09.077).
 33. Jones JK. The role of data mining technology in the identification of signals of possible adverse drug reactions: value and limitations. *Curr Ther Res Clin Exp.* 2001;62(9):664-672. doi: [10.1016/S0011-393X%2801%2980072-2](https://doi.org/10.1016/S0011-393X%2801%2980072-2).
 34. Nelson JC, Ulloa-Pérez E, Bobb JF, Maro JC. Leveraging the entire cohort in drug safety monitor-

- ing: part 1 methods for sequential surveillance that use regression adjustment or weighting to control confounding in a multisite, rare event, distributed data setting. *J Clin Epidemiol*. 2019 Aug;112:77-86. doi: [10.1016/j.jclinepi.2019.04.012](https://doi.org/10.1016/j.jclinepi.2019.04.012).
35. Lai EC, Pratt N, Hsieh CY, et al. Sequence symmetry analysis in pharmacovigilance and pharmacoepidemiologic studies. *Eur J Epidemiol*. 2017 Jul; 32(7):567-582. doi: [10.1007/s10654-017-0281-8](https://doi.org/10.1007/s10654-017-0281-8).
 36. Coloma PM, Trifirò G, Patadia V, Sturkenboom M. Postmarketing safety surveillance: where does signal detection using electronic healthcare records fit into the big picture? *Drug Saf*. 2013 Mar;36(3):183-97. doi: [10.1007/s40264-013-0018-x](https://doi.org/10.1007/s40264-013-0018-x).
 37. Wisniewski AF, Bate A, Bousquet C, et al. Good Signal Detection Practices: Evidence from IMI PROTECT. *Drug Saf*. 2016 Jun;39(6):469-90. doi: [10.1007/s40264-016-0405-1](https://doi.org/10.1007/s40264-016-0405-1).
 38. Prieto-Merino D, Quartey G, Wang J, Kim J. Why a Bayesian approach to safety analysis in pharmacovigilance is important. *Pharm Stat*. 2011 Nov-Dec;10(6):554-9. doi: [10.1002/pst.524](https://doi.org/10.1002/pst.524).
 39. Suling M, Pigeot I. Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics*. 2012 Dec 13;4(4):607-40. doi: [10.3390/pharmaceutics4040607](https://doi.org/10.3390/pharmaceutics4040607).
 40. Coloma PM, Trifirò G, Schuemie MJ, et al. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? *Pharmacoepidemiol Drug Saf*. 2012 Jun;21(6):611-21. doi: [10.1002/pds.3197](https://doi.org/10.1002/pds.3197).
 41. Gault N, Castañeda-Sanabria J, De Rycke Y, et al. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. *BMC Med Res Methodol*. 2017 Feb 8;17(1):25. doi: [10.1186/s12874-016-0278-0](https://doi.org/10.1186/s12874-016-0278-0).
 42. Lian Duan L, Khoshneshin M, Street WN, Liu M. Adverse drug effect detection. *IEEE J Biomed Health Inform*. 2013 Mar;17(2):305-11. doi: [10.1109/TITB.2012.2227272](https://doi.org/10.1109/TITB.2012.2227272).
 43. Sauzet O, Carvajal A, Escudero A, et al. Illustration of the weibull shape parameter signal detection tool using electronic healthcare record data. *Drug Saf*. 2013 Oct;36(10):995-1006. doi: [10.1007/s40264-013-0061-7](https://doi.org/10.1007/s40264-013-0061-7).
 44. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res*. 2013 Feb;22(1):39-56. doi: [10.1177/0962280211403602](https://doi.org/10.1177/0962280211403602).
 45. Schuemie MJ, Coloma PM, Straatman H, et al. Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. *Med Care*. 2012 Oct;50(10):890-7. doi: [10.1097/MLR.0b013e31825f63bf](https://doi.org/10.1097/MLR.0b013e31825f63bf).
 46. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol-DrugSaf*. 2011 Mar;20(3):292-9. doi: [10.1002/pds.2051](https://doi.org/10.1002/pds.2051).
 47. Suchard MA, Zorych I, Simpson SE, et al. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013 Oct;36 Suppl 1:S83-93. doi: [10.1007/s40264-013-0100-4](https://doi.org/10.1007/s40264-013-0100-4).
 48. Zhou X, Douglas IJ, Shen R, Bate A. Signal Detection for Recently Approved Products: Adapting and Evaluating Self-Controlled Case Series Method Using a US Claims and UK Electronic Medical Records Database. *Drug Saf*. 2018 May;41(5):523-536. doi: [10.1007/s40264-017-0626-y](https://doi.org/10.1007/s40264-017-0626-y).
 49. Morel M, Bacry E, Gaïffas S, et al. ConvSCCS: convolutional self-controlled case series model for lagged adverse event detection. *Biostatistics*. 2020 Oct 1;21(4):758-774. doi: [10.1093/biostatistics/kxz003](https://doi.org/10.1093/biostatistics/kxz003).
 50. Schuemie MJ, Trifirò G, Coloma PM, et al. Detecting adverse drug reactions following long-term exposure in longitudinal observational data: The exposure-adjusted self-controlled case series. *Stat Methods Med Res*. 2016 Dec;25(6):2577-2592. doi: [10.1177/0962280214527531](https://doi.org/10.1177/0962280214527531).
 51. Simpson SE. A positive event dependence model for self-controlled case series with applications in postmarketing surveillance. *Biometrics*. 2013 Mar;69(1):128-36. doi: [10.1111/j.1541-0420.2012.01795.x](https://doi.org/10.1111/j.1541-0420.2012.01795.x).
 52. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf*. 2010 Aug;19(8):858-68. doi: [10.1002/pds.1926](https://doi.org/10.1002/pds.1926).
 53. Ryan PB, Schuemie MJ, Gruber S, et al. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013 Oct;36 Suppl 1:S59-72. doi: [10.1007/s40264-013-0099-6](https://doi.org/10.1007/s40264-013-0099-6).
 54. Madigan D, Schuemie MJ, Ryan PB. Empirical performance of the case-control method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013 Oct;36 Suppl 1:S73-82. doi: [10.1007/s40264-013-0105-z](https://doi.org/10.1007/s40264-013-0105-z).
 55. Grosso A, Douglas I, MacAllister R, et al. Use of the self-controlled case series method in drug safety assessment. *Expert Opin Drug Saf*. 2011 May;10(3):337-40. doi: [10.1517/14740338.2011.562187](https://doi.org/10.1517/14740338.2011.562187).
 56. Murphy S, Castro V, Colecchi J, et al. Partners HealthCare OMOP Study Report; 2011.
 57. Takeuchi Y, Shinozaki T, Matsuyama Y. A comparison of estimators from self-controlled case series, case-crossover design, and sequence symme-

- try analysis for pharmacoepidemiological studies. *BMC Med Res Methodol*. 2018 Jan 8;18(1):4. doi: [10.1186/s12874-017-0457-7](https://doi.org/10.1186/s12874-017-0457-7).
58. Thurin NH, Lassalle R, Schuemie M, et al. Empirical assessment of case-based methods for drug safety alert identification in the French National Healthcare System database (SNDS): Methodology of the ALCAPONE project. *Pharmacoepidemiol Drug Saf*. 2020 Sep;29(9):993-1000. doi: [10.1002/pds.4983](https://doi.org/10.1002/pds.4983).
 59. Norén G, Hopstadius J, Bate A, et al. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Disc*. 2010;20:361-387. doi: [10.1007/s10618-009-0152-3](https://doi.org/10.1007/s10618-009-0152-3).
 60. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res*. 2013 Feb;22(1):57-69. doi: [10.1177/0962280211403604](https://doi.org/10.1177/0962280211403604).
 61. Jin H, Chen J, He H, et al. Signaling potential adverse drug reactions from administrative health databases. *IEEE Trans Knowl Data Eng*. 2010;22(6):839-853. doi: [10.1109/TKDE.2009.212](https://doi.org/10.1109/TKDE.2009.212).
 62. Reps J, Garibaldi J, Aickelin U, et al. Comparison of algorithms that detect drug side effects using electronic healthcare databases. *Soft Comput*. 2013;17(12):2381-2397. doi: [10.2139/ssrn.2823255](https://doi.org/10.2139/ssrn.2823255).
 63. Jin H, Chen J, Kelman C, et al. Mining unexpected associations for signalling potential adverse drug reactions from administrative health databases. *Lecture Notes in Computer Science*. 2006;3918:867-876. doi: [10.1007/11731139_101](https://doi.org/10.1007/11731139_101).
 64. Reps J, Garibaldi J, Aickelin U, et al. Comparing data-mining algorithms developed for longitudinal observational databases. Paper presented at: 2012 12th UK Workshop on Computational Intelligence, UKCI 2012; 2012, Edinburgh, UK. doi: [10.1109/UKCI.2012.6335771](https://doi.org/10.1109/UKCI.2012.6335771).
 65. Ji Y, Ying H, Dews P, et al. An exclusive causal-leverage measure for detecting adverse drug reactions from electronic medical records. Paper presented at: Annual Conference of the North American Fuzzy Information Processing Society—NAFIPS; 2011, El Paso. doi: [10.1109/NAFIPS.2011.5751957](https://doi.org/10.1109/NAFIPS.2011.5751957).
 66. Ji Y, Ying H, Dews P, et al. A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance. *IEEE Trans Inf Technol Biomed*. 2011 May;15(3):428-37. doi: [10.1109/TITB.2011.2131669](https://doi.org/10.1109/TITB.2011.2131669).
 67. Ji Y, Ying H, Tran J, et al. A method for mining infrequent causal associations and its application in finding adverse drug reaction signal pairs. *IEEE Trans Knowl Data Eng*. 2013;25(4):721-733. doi: [10.1109/TKDE.2012.28](https://doi.org/10.1109/TKDE.2012.28).
 68. Hallas J, Wang SV, Gagne JJ, et al. Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. *Eur J Epidemiol*. 2018 Jun;33(6):545-555. doi: [10.1007/s10654-018-0386-8](https://doi.org/10.1007/s10654-018-0386-8).
 69. Arnaud M, Bégaud B, Thiessard F, et al. An Automated System Combining Safety Signal Detection and Prioritization from Healthcare Databases: A Pilot Study. *Drug Saf*. 2018 Apr;41(4):377-387. doi: [10.1007/s40264-017-0618-y](https://doi.org/10.1007/s40264-017-0618-y).
 70. Hellfritzsche M, Rasmussen L, Hallas J, Pottegård A. Using the Symmetry Analysis Design to Screen for Adverse Effects of Non-vitamin K Antagonist Oral Anticoagulants. *Drug Saf*. 2018 Jul;41(7):685-695. doi: [10.1007/s40264-018-0650-6](https://doi.org/10.1007/s40264-018-0650-6).
 71. Wahab IA, Pratt NL, Ellett LK, Roughead EE. Sequence Symmetry Analysis as a Signal Detection Tool for Potential Heart Failure Adverse Events in an Administrative Claims Database. *Drug Saf*. 2016 Apr;39(4):347-54. doi: [10.1007/s40264-015-0391-8](https://doi.org/10.1007/s40264-015-0391-8).
 72. Zhou X, Bao W, Gaffney M, et al. Assessing performance of sequential analysis methods for active drug safety surveillance using observational data. *J Biopharm Stat*. 2018;28(4):668-681. doi: [10.1080/10543406.2017.1372776](https://doi.org/10.1080/10543406.2017.1372776).
 73. Kulldorff M, Davis R, Kolczak M, et al. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Seq Anal*. 2011;30:58-78. doi: [10.1080/07474946.2011.539924](https://doi.org/10.1080/07474946.2011.539924).
 74. Brown JS, Kulldorff M, Chan KA, Davis RL, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007 Dec;16(12):1275-84. doi: [10.1002/pds.1509](https://doi.org/10.1002/pds.1509).
 75. Brown JS, Kulldorff M, Petronis KR, et al. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf*. 2009 Mar;18(3):226-34. doi: [10.1002/pds.1706](https://doi.org/10.1002/pds.1706).
 76. Cook AJ, Tiwari RC, Wellman RD, et al. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:72-81. doi: [10.1002/pds.2320](https://doi.org/10.1002/pds.2320).
 77. Li L. A conditional sequential sampling procedure for drug safety surveillance. *Stat Med*. 2009 Nov 10;28(25):3124-38. doi: [10.1002/sim.3689](https://doi.org/10.1002/sim.3689).
 78. Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. *Biomet-*

- rics. 2003 Jun;59(2):323-31. doi: 10.1111/1541-0420.00039.
79. Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf.* 2013 May;22(5):517-23. doi: 10.1002/pds.3423.
80. HuybrechtsKF, KulldorffM, Hernández-DíazS, et al. Active Surveillance of the Safety of Medications Used During Pregnancy. *Am J Epidemiol.* 2021 Jun 1;190(6):1159-1168. doi: 10.1093/aje/kwaa288.
81. Schachterle SE, Hurley S, Liu Q, et al. An Implementation and Visualization of the Tree-Based Scan Statistic for Safety Event Monitoring in Longitudinal Electronic Health Data. *Drug Saf.* 2019 Jun;42(6):727-741. doi: 10.1007/s40264-018-00784-0.
82. Brown JS, Petronis KR, Bate A, et al. Drug Adverse Event Detection in Health Plan Data Using the Gamma Poisson Shrinker and Comparison to the Tree-based Scan Statistic. *Pharmaceutics.* 2013 Mar 14;5(1):179-200. doi: [10.3390/pharmaceutics5010179](https://doi.org/10.3390/pharmaceutics5010179).
83. Reps JM, Garibaldi JM, Aickelin U, et al. A supervised adverse drug reaction signalling framework imitating Bradford Hill's causality considerations. *J Biomed Inform.* 2015 Aug;56:356-68. doi: [10.1016/j.jbi.2015.06.011](https://doi.org/10.1016/j.jbi.2015.06.011).
84. Whalen E, Hauben M, Bate A. Time Series Disturbance Detection for Hypothesis-Free Signal Detection in Longitudinal Observational Databases. *Drug Saf.* 2018 Jun;41(6):565-577. doi: [10.1007/s40264-018-0640-8](https://doi.org/10.1007/s40264-018-0640-8).
85. Karlsson I, Zhao J. Dimensionality reduction with random indexing: an application on adverse drug event detection using electronic health records. Paper presented at: Proceedings of IEEE Symposium on Computer-Based Medical Systems; 2014; New-York, 304–307. doi:[10.1109/CBMS.2014.22](https://doi.org/10.1109/CBMS.2014.22).
86. Reps JM, Garibaldi JM, Aickelin U, et al. Signalling paediatric side effects using an ensemble of simple study designs. *Drug Saf.* 2014 Mar;37(3):163-70. doi: [10.1007/s40264-014-0137-z](https://doi.org/10.1007/s40264-014-0137-z).
87. Bagattini F, Karlsson I, Rebane J, Papapetrou P. A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records. *BMC Med Inform Decis Mak.* 2019 Jan 10;19(1):7. doi: 10.1186/s12911-018-0717-4.
88. Demailly R, Escolano S, Haramburu F, Tubert-Bitter P, Ahmed I. Identifying Drugs Inducing Prematurity by Mining Claims Data with High-Dimensional Confounder Score Strategies. *Drug Saf.* 2020 Jun;43(6):549-559. doi: [10.1007/s40264-020-00916-5](https://doi.org/10.1007/s40264-020-00916-5).
89. Bampa M, Papapetrou P. Mining adverse drug events using multiple feature hierarchies and patient history windows. Paper presented at: IEEE International Conference on Data Mining Workshops, ICDMW, volume 2019; 2019; Beijing, China, 925–932. doi:[10.1109/ICDMW.2019.00135](https://doi.org/10.1109/ICDMW.2019.00135).
90. Zhao J, Henriksson A, Kvist M, et al. Handling Temporality of Clinical Events for Drug Safety Surveillance. *AMIA Annu Symp Proc.* 2015 Nov5;2015:1371-80.
91. DuMouchel W, Ryan PB, Schuemie MJ, Madigan D. Evaluation of disproportionality safety signaling applied to healthcare databases. *Drug Saf.* 2013 Oct;36 Suppl1:S123-32. doi: [10.1007/s40264-013-0106-y](https://doi.org/10.1007/s40264-013-0106-y).
92. Norén G, Hopstadius J, Bate A, Edwards IR. Safety surveillance of longitudinal databases: results on real-world data. *Pharmacoepidemiol Drug Saf.* 2012;21(6):673-5. doi:[10.1002/pds.3258](https://doi.org/10.1002/pds.3258). [myRWD](https://doi.org/10.1002/pds.3258)