https://russjcardiol.elpub.ru
doi:10.15829/1560-4071-2020-4037

ISSN 1560-4071 (print)
ISSN 2618-7620 (online)

Доказательные подходы к сравнению эффективности вмешательств в современной кардиологии: тенденции, ошибки и перспективы

Гиляревский С. Р.[1], Беленков Ю. Н.[2]

В статье рассматриваются ограничения доказательной информации, основанной на результатах обсервационных исследований. Подробно обсуждаются современные подходы к уменьшению систематических ошибок обсервационных исследований, в частности, с помощью ставшего популярным в последние годы метода propensity score matching, основанного на формировании групп с учетом предпочтительного назначения определенного типа терапии. Рассматриваются основные отличия рандомизированных и обсервационных исследований, а также приводятся аргументы против использования обсервационных исследований и усовершенствованных методов анализа для сравнения эффективности методов лечения в клинической практике. Обсуждается роль обсервационных исследований как источника доказательной информации. Указывается на обоснованность выполнения крупных проспективных обсервационных исследований для получения информации об эффектах применения лекарственных препаратов в клинической практике в период после выхода препарата на фармацевтический рынок, а также для получения данных об использовании препаратов у пациентов, характеристики которых отличаются от таковых у участников рандомизированных клинических исследований.

Ключевые слова: доказательная медицина, клинические исследования, фибрилляция предсердий, антикоагулянты.

Отношения и деятельность. Статья подготовлена при поддержке компании АО “Байер”. Номер одобрения: PP-XAR-RU-0493-1.

[1]Российская медицинская академия непрерывного профессионального образования (РМАНПО) Минздрава России, Москва; [2]ФГАОУ ВО Первый Московский государственный медицинский университет им. И. М. Сеченова (Сеченовский Университет) Министерства здравоохранения Российской Федерации, Москва, Россия.

Гиляревский С. Р.* — д.м.н., профессор кафедры клинической фармакологии и терапии, ORCID: 0000-0002-8505-1848, Беленков Ю. Н. — академик РАН, д.м.н., профессор, директор Клиники госпитальной терапии им. А. А. Остроумова, зав. кафедрой госпитальной терапии № 1 лечебного факультета, ORCID: 0000-0002-3014-6129.

*Автор, ответственный за переписку (Corresponding author):
sgilarevsky@rambler.ru

ДИ — доверительный интервал, ППОАК — прямые пероральные антикоагулянты, РКИ — рандомизированные клинические исследования, ФП — фибрилляция предсердий.

Рукопись получена 29.07.2020
Рецензия получена 06.08.2020
Принята к публикации 13.08.2020

CC BY 4.0

Для цитирования: Гиляревский С. Р., Беленков Ю. Н. Доказательные подходы к сравнению эффективности вмешательств в современной кардиологии: тенденции, ошибки и перспективы. *Российский кардиологический журнал*. 2020;25(8):4037. doi:10.15829/1560-4071-2020-4037

Evidence-based approaches to comparing the effectiveness of modern cardiology interventions: trends, bias and prospects

Gilyarevsky S. R.[1], Belenkov Yu. N.[2]

The article discusses the limitations of the evidence from observational studies. Modern approaches to reducing bias in observational studies are discussed in detail, in particular, propensity score matching, which has become popular in recent years. The main differences between randomized and observational studies are discussed. Arguments against the observational studies and improved methods of analysis to compare the treatments’ effectiveness in clinical practice are presented. The role of observational studies as a source of evidence is discussed. The article points out the validity of performing large-scale prospective observational studies to assess the effects of postmarketing drug use in clinical practice, as well as to obtain data on drug use in patients that differ from those in randomized clinical trials.

Key words: evidence-based medicine, clinical trials, atrial fibrillation, anticoagulants.

Relationships and Activities. The article was prepared with the support of JSC Bayer. Approval number: PP-XAR-RU-0493-1.

[1]Russian Medical Academy of Continuous Professional Education, Moscow; [2]I.M. Sechenov First Moscow State Medical University, Moscow, Russia.

Gilyarevsky S. R.* ORCID: 0000-0002-8505-1848, Belenkov Yu. N. ORCID: 0000-0002-3014-6129.

*Corresponding author:
sgilarevsky@rambler.ru

Received: 29.07.2020 **Revision Received:** 06.08.2020 **Accepted:** 13.08.2020

For citation: Gilyarevsky S. R., Belenkov Yu. N. Evidence-based approaches to comparing the effectiveness of modern cardiology interventions: trends, bias and prospects. *Russian Journal of Cardiology*. 2020;25(8):4037. (In Russ.) doi:10.15829/1560-4071-2020-4037

"Наука — это попытка привести хаотическое многообразие нашего чувственного опыта в соответствие с некоторой единой системой мышления"
Альберт Эйнштейн

Психологические предпосылки к использованию необоснованных данных в медицине

Психолог Даниэль Канеман, лауреат Нобелевской премии, выделяет две системы головного мозга [1]: Систему 1, ядро которой составляет ассоциативная память, и Систему 2, которая выполняет контролирующие функции. Система 1 постоянно интерпретирует происходящее в окружающем мире в любой заданный момент на основе той, часто ограниченной, информации, которая ей доступна, и в отсутствие анализа ее достоверности. Задача Системы 2 состоит в проверке моделей, "наспех" построенных Системой 1.

— Система 1 срабатывает автоматически и очень быстро, не требуя или почти не требуя усилий, но не давая ощущения намеренного контроля.

— Система 2 обеспечивает внимание, необходимое для сознательных умственных усилий, в т.ч. для выполнения сложных вычислений.

Риск искажения действительности может быть особенно высок в тех областях человеческой деятельности, в которых широко распространена так называемая интуитивная экспертиза. Не будет преувеличением мнение о том, что в работе врачей многих клинических специальностей такая экспертиза занимает существенное место.

Система 1 часто используется не только в клинической работе, но также при получении и интерпретации доказательной информации, а также при ее использовании для выработки эффективной, безопасной и, по возможности, научно обоснованной тактики.

Нередко врач получает такую информацию о преимуществах применения одного препарата по сравнению с другим, особенно препаратов, относящихся к одному классу: "Вы видите результаты анализа, который включал данные о ... 100 тыс., 200 тыс. и даже 500 тыс. пациентов, которые наблюдались в условиях "реальной клинической практики", а не рафинированных условиях клинического исследования. Результаты свидетельствуют о том, что препарат А намного безопаснее и/или эффективнее препарата В. И это прямое сравнение препарата А и препарата В". Система 1 получает достаточную информация для формирования мнения. Более того, нередко в последнее время в таком наблюдательном исследовании используют особые методы статистического анализа, которые позволяют сделать группы сравниваемых пациентов более сопоставимыми по определенным доступным для оценки характеристикам (подробнее об ограничениях таких методов см. далее). И тогда врач может услышать, что используемый метод "практически не отличается от рандомизации или даже превосходит его". То есть Система 1 врача получает следующий набор фраз, которые выстраиваются в "нужную" цепочку: "очень много пациентов — лечение в условиях реальной клинической практики — сходство групп по основным исходным характеристикам". Этого достаточно для Системы 1, чтобы у врача не осталось сомнений в том, что именно применение препарата А намного эффективнее и/или безопаснее препарат В. К сожалению, в таких случаях редко включается Система 2, необходимая для критического анализа и проверки обоснованности полученной информации и модели, "наспех" построенной Системой 1. Какие же вопросы должна была бы задать Система 2 в таких случаях, чтобы врач мог быть уверен в научной обоснованности указанных выводов? Вот минимум вопросов, на которые следует ответить: 1) каков источник и, соответственно, клиническая точность исходной информации о пациентах. Нередко такую информацию получают из баз данных страховых компаний, в которых в большинстве случаев медицинская информация не может быть такой же точной и подробной, как информация, извлеченная из медицинской документации; 2) насколько можно быть уверенным, что группы пациентов не различаются по тем характеристикам, которые не оценивали в ходе исследования, или по показателям, которые в принципе нельзя измерить (так называемые неучитываемые характеристики). 3) насколько изменение объема выборки при использовании специальных методов анализа обеспечивает статистическую мощность анализа; 4) ну и, наконец, главный вопрос, в какой степени указанные ранее факторы могут обусловливать систематические ошибки, т.к. такие ошибки затрудняют интерпретацию полученных результатов и формулирование более или менее обоснованных выводов. Если отсутствуют подтверждения высокого методологического уровня исследований, к сожалению, его результаты становятся "информационным мусором", а их использование — маркером научной недобросовестности эксперта.

Иногда интуитивно складывается мнение о том, что очень большая выборка будет сама по себе уменьшать систематическую ошибку наблюдательного исследования. Не оспаривая модное понятие и магию "больших данных" (big data), а также пользу анализа с использованием их для решения определенных задач, можно утверждать, что использование таких данных не позволяет получить обоснованные данные о сравнительной эффективности и/или безопасности методов лечения. Далее в статье мы постараемся объяснить, почему даже очень большой объем выборки в отсутствие соблюдения определенных методологических подходов не позволяет получить в клиниче-

ском исследовании научные основания для утверждения о преимуществах одного вмешательства по сравнению с другим.

Рандомизация как необходимое условие обоснованного сравнения методов лечения

Обязательное использование рандомизации при оценке эффектов лечения относится к базовым принципам доказательной медицины [2]. Нерандомизированные, т.е. обсервационные исследования, полезны для описательных целей, но не подходят для сравнения эффектов лечения в связи с невозможностью надежно устранить вмешивающиеся факторы (в англ. confounders), т.е. возможные источники систематических ошибок, которые обусловлены действием дополнительного неизмеренного фактора (вмешивающегося), если такой фактор связан с изучаемым воздействующим фактором (в отсутствие причинно-следственной связи с ним), но имеет причинно-следственную связь с изучаемым клиническим исходом [3]. О наличии вмешивающегося фактора говорят тогда, когда кажущаяся очевидной причинно-следственная связь между вмешательством (например, лечением) и клиническим исходом на самом деле, по крайней мере отчасти, обусловлена влиянием третьего (вмешивающегося) фактора [4]. Имеется несколько сложных методов анализа, которые используют для уменьшения различий между группами в обсервационных исследованиях по учтенным или известным характеристикам, но ни один из таких методов не позволяет учитывать неизмеренные характеристики [1]. Следует отметить, что в целом примерно для 90% клинических рекомендаций отсутствуют подтверждения, полученные в ходе выполнения крупных рандомизированных клинических исследований (РКИ) с высоким качеством [5]. Так что осознание ограниченной доказательной силы обсервационных исследований представляется очень важным.

Следует тем не менее отметить, что рандомизацию следует считать необходимым, но не достаточным условием получения обоснованной доказательной информации. В частности, недавние попытки сравнения эффективности и безопасности двух антитромботических препаратов в исследованиях, заведомо не имеющих достаточной статистической мощности и необходимого методологического качества, могут служить примером недобросовестной имитации использования принципов доказательной медицины. Одним из последних примеров может служить исследование ISAR-REACT 5, открытое РКИ, явно имеющее недостаточную статистическую мощность для сравнения эффективности и безопасность мощных ингибиторов $P2Y_{12}$, а также ряд очевидных методологических ограничений (открытый характер исследования и отсутствие использования плацебо при сравнение препарата для приема 2 раза/сут. с препаратом для приема 1 раз/сут., а также наблюдение за пациентами в основном по телефону, что ограничивало возможность оценки приверженности к терапии) [6]. В какой-то степени такие исследования могут стать еще большим источником формирования ложных суждений по сравнению с обсервационными исследованиями, т.к. Система 1 воспринимает слово “рандомизация” как критерий правильного подхода к выполнению исследования, а Система 2 не “просыпается” для проверки качества такого исследования.

Почему не так много РКИ с высоким качеством? Одна из причин такой ситуации, которую отмечают эксперты, состоит в особенности “экосистем” клинических исследований, которые практически не изменились с 80-х годов XX века [1]. В тот период влияние совокупности факторов обусловливало преимущественную разработку РКИ как инструмента, направленного на “победу” нового метода лечения в соревновании за одобрение регуляторных органов. Причем планирование каждого исследования, в зависимости от его целей, было строго индивидуальным. Отбор пациентов в такие исследования как правило не был систематическим, но скорее случайным, а пациенты, которых удавалось уговорить участвовать в исследовании, в последующем в ходе наблюдения довольно часто посещали исследовательский центр, что само по себе могло повлиять на полученные результаты [1].

Более того, учитывая высокую стоимость таких исследований, большинство РКИ с оценкой клинических исходов финансировалось фармацевтическими компаниями или производителями медицинских устройств, которые получали прямую выгоду от получения одобрения регуляторными органами соответствующего вмешательства. Очевидно, что имевшиеся условия выполнения РКИ создавали существенные финансовые барьеры на пути выполнения исследований, непосредственно не направленных на установления таких эффектов вмешательства, в которых были бы заинтересованы компании-производителей изучаемой медицинской продукции. Такие РКИ должны были быть посвящены сравнительной оценке новых и стандартных тактик лечения, вмешательств служб здравоохранения или оценке относительной эффективности уже одобренных к применению лекарственных препаратов или устройств. Кроме того, такие исследования могли быть посвящены оценке сочетанного использования разных тактик лечения и возможности изменения интенсивности терапии, ее продолжительности или доз лекарственных препаратов. Указанные ограничения становились причиной того, что многие важные исследования не удалось выполнить, а это, в свою очередь, оставляло существенные пробелы в доказательствах. Соответственно, отсутствие доказательств эффективности вмешательств, получен-

ных с РКИ с высоким методологическим качеством, препятствует созданию обоснованных клинических рекомендаций, что обусловливает большую вариабельность подходов к лечению пациентов, а также неоптимальность терапии, которая при этом нередко бывает дорогостоящей [7]. Ну и конечно в таких условиях всегда есть искушение использовать источники информации с невысокой доказательной силой, т.е. пытаться оценивать эффекты терапии с обсервационных исследований.

Осознание важности РКИ для выработки определенной тактики лечения стало основанием для более гибкого подхода к оценке результатов крупных РКИ, которые были выполнены без грубых методологических ошибок. Подходы к использованию доказательной информации для принятия клинического решения меняются. В частности, эксперты считают, что значение p не должно быть критерием для принятия клинического решения, а только указанием того, насколько результаты могли быть случайными [8]. Причем больше внимания при анализе результатов исследования и их пригодности для клинической практики предлагается уделять анализу выраженности эффекта и доверительным интервалам.

Ограниченная роль результатов обсервационных исследований для принятия клинического решения

Важный опыт использования доказательной информации при необходимости быстрого принятия решения о тактике лечения пациентов был получен в условиях быстро распространяющейся инфекции COVID-19. Например, на основании небольших обсервационных исследований быстро было принято решение об использовании гидроксихлорохина и макролидных антибиотиков при лечении широкого круга пациентов с такой инфекцией. К сожалению, такой опыт подтвердил высокую вероятность получения необоснованных данных в ходе выполнения обсервационных исследований, причем даже в случае проспективно собранных данных. В начале июня 2020г авторы двух статей с отчетом о результатах крупных обсервационных исследований, которые в течение мая 2020г были опубликованы в двух журналах с очень высоким импакт-фактором — New England Journal of Medicine и Lancet, отозвали свои публикации [9-12]. В первой статье были представлены результаты крупного обсервационного исследования, в котором анализировали данные о 8910 пациентах, госпитализированных по поводу COVID-19 в 169 стационаров Азии, Европы и Северной Америки в период с 20 декабря по 15 марта 2020г. В ходе выполнения исследования оценивали связь между наличием определенных сердечно-сосудистых заболеваний, лекарственной терапией таких заболеваний и смертностью во время пребывания в стационаре. Результаты исследования представлялись крайне важными, т.к. подтверждали мнение о безопасности и даже преимуществах применения ингибиторов ангиотензинпревращающего фермента и статинов у пациентов с COVID-19. Причем в соответствующем разделе статьи было указано, что “все авторы, ознакомившись с текстом статьи, поручаются за точность и полноту предоставляемых данных” [4]. Вторая статья, опубликованная в журнале Lancet, была посвящена оценке связи между приемом гидроксихлорохина или хлорохина в сочетании с макролидными антибиотиками или в их отсутствие и риском развития неблагоприятных исходов во время пребывания в стационаре пациентов с COVID-19. Данные для анализа были получены в многонациональном регистре, включавшем информацию о 96032 пациентах с подтвержденным лабораторно SARS-CoV-2. На основании выполненного анализа сделали вывод об отсутствии преимуществ приема гидроксихлорохина или хлорохина изолированно или в сочетании с макролидным антибиотиком по влиянию на смертность во время пребывания в стационаре по поводу COVID-19. Более того, было отмечено, что применение режима терапии, включавшего такие препараты, сопровождалось снижением выживаемости во время пребывания в стационаре и увеличением частоты развития желудочковых аритмий при лечении COVID-19. Полученные в этом исследовании данные стали основанием для приостановки набора пациентов в группу приема гидроксихлорохина в РКИ SOLIDARITY и ряда других РКИ по оценке эффективности и безопасности применения гидроксихлорохина при COVID-19.

Почему же были отозваны указанные статьи? Это было связано с тем, что на запрос экспертов о предоставлении подробной информации об участниках исследований, представители компании Surgisphere, которая отвечала за сбор данных, не смогли этого сделать. В основном это касалось данных, которые были получены в Африке и Австралии. Таким образом, обсервационные исследования не только в принципе не могли дать ответ на вопрос о причинно-следственной связи между применяемой терапией и клиническими исходами, но ошибки при его выполнении стали препятствием для какой-либо трактовки их результатов.

В целом полученный негативный опыт “наспех” выполненных исследований должен стать основанием как для усовершенствования подходов к рецензированию статей в медицинских журналах, так и осознанию ненадежности данных, полученных с помощью обсервационных исследований, основанных на анализе баз данных.

Следует отметить опубликование большого числа результатов ретроспективных анализов баз данных страховых компаний, включающих информацию о пациентах с фибрилляцией предсердий (ФП), в ходе выполнения которых была предпринята попытка сравнить частоту развития эмболий и/или кровотечений при применении антагонистов витамина К,

в основном, варфарина, и прямых пероральных антикоагулянтов (ППОАК). К неустранимым ограничениям любого анализа баз данных страховых заявок относят обсервационный характер исследований такого типа, а также невозможность установить вмешивающиеся факторы (confounding variables), включая такие, как сопутствующий прием доступных без рецепта ацетилсалициловой кислоты или нестероидных противовоспалительных средств, который может существенно влиять на риск развития кровотечений [13]. Несмотря на попытки сформировать группы с учетом возможных вмешивающихся факторов, в таких исследованиях практически невозможно исключить систематические ошибки, связанные с отбором, а частое отсутствие в базах данных лабораторных данных о функции почек не позволяет оценить правильность применяемой дозы ППОАК.

В качестве убедительного примера, иллюстрирующего ненадежность анализа баз данных страховых компаний для оценки риска развития кровотечений у пациентов, применяющих антитромботические препараты, можно привести результаты крупного обсервационного исследования. В ходе его выполнения сравнивали частоту развития неблагоприятных исходов у участников крупного проспективного обсервационного исследования TRANSLATE-ACS, которые были зарегистрированы не только в базе данных указанного исследования, но и базе данных страховых заявок [14]. В целом в анализ была включена информация о 12365 пациентах (71,9% мужчины; средний возраст 60±11,6 года), которые перенесли острый коронарный синдром и принимали антиагреганты. Кумулятивная частота развития инфаркта миокарда, инсульта и кровотечений в течение 1 года, которая основывалась на данных, содержащихся в страховых заявках, составляла 4,3, 0,9 и 5,0%, соответственно, а частота развития таких исходов по данным, подтвержденным врачами в ходе выполнения исследования TRANSLATE-ACS, достигала 4,7, 0,9 и 5,4%, соответственно. Таким образом, согласованность результатов анализа, выполненного с помощью базы данных страховых компаний и базы данных проспективного исследования, была лишь умеренной с коэффициентом k (для k статистики) для частоты развития инфаркта миокарда 0,76 (при 95% доверительном интервале (ДИ) 0,73-0,79), и для частоты развития инсульта 0,55 (95% ДИ 0,41-0,68). Причем для частоты развития умеренно выраженных или тяжелых кровотечений по шкале GUSTO согласованность была очевидно плохой с коэффициентом k=0,24 (при 95% ДИ 0,11-0,20).

Не менее убедительным примером ненадежности выводов ретроспективного анализа, основанного на базе данных страховых заявок, могут служить недавно опубликованные Fralick M, et al. результаты [15]. В целом анализ основывался на данных о почти 100 тыс. пациентов, которые впервые начали принимать апиксабан или ривароксабан по поводу ФП неклапанной природы. Пациенты, которым назначался апиксабан (n=59172) по сравнению с пациентами, которым был назначен ривароксабан (n=40706), были старше и имели более высокую вероятность наличия болезни почек или сердечно-сосудистых заболеваний и исходно получали больший объем лекарственной терапии по поводу сопутствующих заболеваний. После триминга данных с помощью метода PS matching были сформированы две группы по 39351 человек в каждой. Таким образом, из анализа были исключены данные почти о 20 тыс. пациентов, принимавших ривароксабан, и только о 1355 пациентах, принимавших апиксабан, таким образом, исходная выборка была уменьшена на 21%. Результаты анализа свидетельствовали о менее высоком риске развития ишемического инсульта или эмболий в сосуды большого круга кровообращения в группе апиксабана по сравнению с группой ривароксабана (отношение рисков 0,82 при 95% ДИ 0,68-0,98), а также менее высоком риске развития кровотечений из желудочно-кишечного тракта и внутричерепных кровоизлияний. Важно отметить, что сами авторы анализа подчеркнули, что неизмеренные вмешивающиеся факторы на 16% чаще отмечались в группе ривароксабана, что обусловливало завышение риска развития тромбоэмболических осложнений в этой группе на 16%. Кроме того, указанные факторы могли завышать риск развития тяжелых кровотечений в группе ривароксабана по сравнению с группой апиксабана на 240%. Авторы анализа отметили невозможность учета неизмеренных вмешивающихся факторов при формировании групп с помощью метода PS matching. В частности, отсутствовала информация о применении безрецептурных препаратов (например, ацетилсалициловой кислоты), а результаты лабораторных анализов невозможно было внести в шкалу PS из-за пропущенных данных. Кроме того, клинические исходы, учитываемые в анализе, основывались на информации о диагнозах в соответствии с кодами МКБ-9 и МКБ-10, а не на подтвержденных членами независимого комитета диагнозах.

Для чего нужны обсервационные исследования, в которых оценивают эффекты вмешательств?

Как указывалось ранее, существует согласованное мнение экспертов о том, что основным методом оценки эффектов лечебных вмешательств остаются РКИ [16]. Рандомизация считается единственным методом, гарантирующим одинаковое распределение пациентов в группу вмешательства и группу контроля как по известным, так и неизвестным характеристикам, и позволяет установить причинно-следственную связь между исследуемым вмешательством и риском развития определенного исхода. Однако высказыва-

ется мнение о том, что в некоторых случаях РКИ могут быть "необязательными, неприемлемыми, невозможными или несоответствующими" [17]. РКИ также критикуют за отсутствие внешней обоснованности (external validity), т.к. пациенты, включенные в РКИ, обычно моложе и имеют лучшее клиническое состояние по сравнению с большинством пациентов в клинической практике [18, 19].

Итак, для чего же нужны обсервационные исследования, в т.ч. и ретроспективные исследования, представляющие анализ крупных баз данных страховых компаний? По мнению многих экспертов, РКИ непригодны для выявления редких побочных эффектов применения препаратов и установления последствий их длительного использования. Такое мнение можно считать обоснованным. Во-первых, в связи с недостаточно большой выборкой РКИ, которая редко превышает 10 тыс. пациентов и часто <1 тыс. человек. Во-вторых, продолжительность исследований по оценке эффективности вмешательств редко превышает 5 лет и часто составляет около 1 года. Именно такие ограничения РКИ становятся основанием для выполнения обсервационных исследований для оценки эффектов применения терапии. Кроме того, целью таких исследований становится попытка выявления редких или отсроченных побочных эффектов. Обсервационные исследования выполняют также для наблюдения за действием препаратов у пациентов, которых обычно не включают в РКИ или включают редко. Еще одним основанием для выполнения таких исследований может быть описание характеристик пациентов, которым назначают препарат после его выхода на фармацевтический рынок, в частности, для ответа на вопрос, соблюдаются ли при применении препарата в клинической практике ограничения, которые были наложены регуляторными органами.

В таких случаях оптимальным подходом к получению наиболее объективной информации считается выполнение крупного проспективного обсервационного исследования, в котором возможно получение подробной информации о характеристиках пациентов и клинических исходах. В качестве примера такого исследования можно привести международное проспективное обсервационное исследование XANTUS по оценке эффектов приема ривароксабана у пациентов с ФП в условиях клинической практики [20]. В 311 исследовательских центрах Европы, Израиля и Канады в исследование было включено 6784 пациента. Важно отметить, что, в отличие от ретроспективных анализов баз данных страховых заявок, в исследовании XANTUS все изучаемые неблагоприятные клинические исходы, включая тяжелые кровотечения, инсульты и эмболии в сосуды большого круга кровообращения, подтверждались членами независимого комитета по подтверждению клинических исходов. Результаты такого исследования свидетельствовали о низкой частоте развития инсульта и тяжелых кровотечений при использовании ривароксабана в клинической практике у пациентов с ФП неклапанной природы.

Увеличивающаяся доступность медицинской информации административных баз данных страховых компаний или организаций, предоставляющих медицинские услуги, а также усовершенствования методов анализа для обработки очень больших баз данных, существенно облегчают выполнения обсервационных исследований. Такое распространение обсервационных исследований, которые с большой вероятностью искажают истинный эффект вмешательства, стало поводом для беспокойства экспертов [21]. Такая же обеспокоенность высказывается и по поводу метаанализа обсервационных исследований. Так, Milton Packer на страницах журнала Circulation [22] напоминает, что "метаанализ во всех случаях представляет собой обсервационное исследование, в основе которого лежит использование математического метода для объединения данных, что позволяет учитывать количественные характеристики, но не качество наблюдений, выполненных в разных исследованиях". Таким образом, выполнение метаанализа повышает точность количественных оценок, но не их истинность (вследствие методологических недостатков самих исследований, включенных в метаанализ).

Таким образом, учитывая указанные ограничения РКИ, нерандомизированные исследования могут дополнять результаты РКИ для оценки эффектов лечения. Но, с другой стороны, обсервационные исследования лишены внутренней обоснованности (internal validity), т.е. в связи с распределением пациентов в группы без рандомизации обусловливают вероятность систематических различий между группами как по известным, так и по неизвестным характеристикам пациентов. Причем различия по неустановленным или неизвестным характеристикам считаются еще более неблагоприятным фактором, повышающим вероятность систематических ошибок [10].

Современные подходы к уменьшению ограничений обсервационных исследований

Для уменьшения указанных ранее ограничений обсервационных исследований разработано несколько статистических методов для учета определенных различий между группами в ходе выполнения анализа. Стандартными процедурами для учета возможных известных вмешивающихся факторов считается применение многофакторных регрессионных моделей. Однако в последнее время все чаще для таких целей стали использовать анализ, в котором группы формируются с учетом шкал для оценки предпочтительного назначения терапии или вероятности преимуществ определенной тактики лечения, так называемый анализ propensity score (PS) [23].

Что такое propensity score — PS? Оценка по шкале PS отражает вероятность того, что пациенту назначена определенная тактика лечения, и такой показатель используется как интегральный показатель, отражающий разные известные характеристики пациента. В РКИ с распределением пациентов в группы в соотношении 1:1 вероятность назначения определенной тактики всегда точно составляет 0,5. В нерандомизированном исследовании такая вероятность для каждого пациента неизвестна и зависит от его характеристик. Использование шкалы PS облегчает формирование групп, сходных по исходным характеристикам, но только по известным или установленным характеристикам. В таких случаях использование логистических регрессионных моделей, в которых распределение пациентов в группы используется как зависимый показатель, а характеристика пациента до начала лечения как независимый показатель, помогает сформировать группы, не отличающиеся существенно по многих показателям. Используя известные (или установленные) параметры в такой модели PS значение по шкале PS может быть рассчитано для каждого пациента. Причем при выборе независимых переменных в модели PS необходим выбор тех характеристик, которые в первую очередь позволяют прогнозировать эффективность терапии (а не предпочтительное назначение определенной терапии), т.к. в таком случае ограничивается вариабельность эффекта лечения без дополнительного увеличения вероятности систематических ошибок [24]. Таким образом, результаты оценки с использованием PS могут варьировать в зависимости от выбранного подхода к учитываемым характеристикам пациентов. Еще раз следует подчеркнуть, что модель PS не может учитывать факторы, которые неизвестны или не измерялись, что существенно ограничивает обоснованность выводов, основанных на результатах такого анализа. Второй этап выполнения анализа PS включает выбор оцениваемой терапии.

В настоящее время имеется 4 метода применения PS [25]:

— ***PS matching*** — подбор по соответствию оценки по шкале PS;

— ***Inverse probability of treatment weighting (IPTW)*** — обратная вероятность применения определенного метода лечения;

— ***Stratification*** — стратификация;

— ***Regression adjustment for the PS*** — использование регрессионной модели с учетом оценки по PS.

Не вдаваясь в подробности каждого из типов анализа PS, постараемся коротко охарактеризовать каждый из них, отмечая их преимущества и ограничения.

PS matching: для каждого леченого пациента подбирается один нелеченый (или леченый другим препаратом) пациент (в случае распределения в группы в соотношении 1:1) или более одного такого пациента (в случае распределения в соотношении 1:n), у которого или которых имеется такая же или сходная оценка по PS, если различия по такой оценке не выходят за пределы заранее принятых границ [10]. Очевидно, что в таких случаях не для всех пациентов удается найти пару со сходной или близкой оценкой PS, что становится причиной исключения из анализа более или менее большого числа пациентов, ограничивая статистическую мощность анализа и отчасти увеличивая вероятность систематической ошибки, связанной с отбором.

IPTW — обратная вероятность применения определенного метода лечения: в случае формирования групп с помощью IPTW каждый пациент распределяется в группу реципрокно вероятности получения терапии в клинической практике, которая устанавливается с помощью оценки статистического веса. Соответственно, пациенту, у которого применяется терапия, присваивается взвешенное значение, равное 1/PS, а пациенту контрольной группы — 1/(1-PS). Несмотря на математическую обоснованность такого подхода взвешивания, его интерпретация, по мнению экспертов, может быть лишь интуитивной [26]. Таким образом, леченый пациент с низкой оценкой по PS (для вероятности применения определенного лечения) получает высокий вес, т.к. такой пациент имеет сходные характеристики с пациентом контрольной группы (имеющим низкую оценку по PS), таким образом повышается обоснованность сравнения между двумя группами. В ходе выполнения анализа для оценки эффекта лечения, данные о пациентах вводят с учетом такого статистического веса. Такой подход позволяет включать в анализ данные обо всех пациентах, сохраняя статистическую мощность анализа.

Stratification — стратификация представляет собой упрощенную форму метода PS matching. Базу данных разделяют на несколько подгрупп одинакового размера (например, разделение на квинтили) на основании рассчитанной PS. В каждой подгруппе эффект лечения оценивают с помощью стандартных методов, а общий эффект лечения рассчитывают с помощью метааналитических методов.

Наконец, *regression adjustment for the PS* — использование регрессионной модели с учетом оценки по PS — представляет собой анализ с помощью обычной регрессионной модели с использованием интересующего клинического исхода в качестве зависимой переменной, а эффекта лечения и оценки по шкале PS в качестве независимых переменных. Таким образом, влияние лечения на клинический исход оценивается с учетом PS.

Каждый из указанных методов имеет специфичные преимущества и недостатки, но метод *PS matching* обычно считается предпочтительным в большинстве случаев [27, 28]. Основным преимуществом метода *PS matching* считают возможность представления характеристик пациентов в группе лечения и группе контроля, как это обычно отражается в таблице в отче-

Таблица 1

Основные преимущества и ограничения определенных методов PS

Преимущества	Метод propensity score				
	Подбор по соответствию оценки по шкале PS (*PS matching*)	Стратификация	Обратная вероятность применения определенного метода лечения (IPTW)	Регрессионная модель с учетом оценки по PS	Стандартный регрессионный анализ
Упрощение сравнения леченых и нелеченых пациентов	Да	Частично	Частично	Нет	Нет
Возможность оценки сбалансированности групп по исходным характеристикам	Да	Да	Частично	Нет	Нет
Полное использование базы данных*	Нет	Да	Да	Да	Да
Сходство с РКИ**	Да	Частично	Частично	Нет	Нет
Устойчивость к резко выделяющимся значениям***	Да	Нет	Да	Да	Да
Меньшее число статистических допущений в модели	Да	Да	Частично	Нет	Нет

Примечание: * — меньшая вариабельность эффекта лечения, но увеличение опасности систематических ошибок, ** — формирование сходных групп и игнорирование предполагаемого результата, *** — пациенты с крайними значениями оценки по шкале PS.

Сокращения: PS — propensity score, IPTW — inverse probability of treatment weighting (обратная вероятность применения определенного метода лечения), РКИ — рандомизированное клиническое исследование.

тах о РКИ. Такая возможность позволяет ответить на важный вопрос о том, насколько различаются группы по известным характеристикам до формирования групп с использованием метода *PS matching*. Кроме того, подробная информация о характеристиках пациентов двух (или более) групп до применения *PS matching* необходима для того, чтобы понять в какой степени использование *PS matching* компенсирует имевшиеся исходно различия между группами.

Следует, тем не менее, отметить, что при формировании групп с использованием *PS matching* неизбежно исключение из анализа данных пациентов, для которых невозможно было найти соответствующую пару (так называемый тримминг данных). В этом заключается основной недостаток метода *PS matching*. Причем использование других методов PS подразумевает включение в анализ данных обо всех пациентах. Соответственно, исключение из анализа определенного числа пациентов из общей выборки при выполнении *PS matching* будет приводить к снижению статистической мощности анализа, но при этом сохраняется возможность оценить, данные о каких пациентах были преимущественно включены в анализ, а данные о каких пациентах, напротив, были скорее исключены из него. Однако не представляется возможным сделать убедительное заключение об эффектах терапии отдельно в каждой из таких подгрупп. Наконец, при выборе *PS matching* по сравнению с другими методами PS всегда возникает необходимость компромисса между возможностью снижения систематической ошибки за счет формирования групп, сходных по известным характеристикам, и вероятностью неточной оценки эффекта лечения вследствие неизбежного уменьшения выборки [17]. Метод *PS matching*, в первую очередь, должен использоваться в тех случаях, когда есть необходимость сформировать группы максимально сходные по характеристикам пациентов (за этот счет уменьшаются систематические ошибки). Однако учитывая неизбежное уменьшение числа пациентов, данные о которых включены в анализ, приходится допускать более выраженную вариабельность установленного эффекта лечения. В таблице 1 представлены основные преимущества и ограничения определенных методов PS.

Оценка качества модели PS может состоять только в проверке сбалансированности групп по исходным характеристикам пациентов. Никакие другие формальные тесты для проверки пригодности модели не подходят, т.к. непригодны для выявления неустановленных вмешивающихся факторов [10, 29].

Несмотря на усовершенствование методов PS, есть несколько неустранимых проблем, определяющих ненадежность полученных с помощью таких методов результатов. К ним относят обработку пропущенных значений, определение минимальных требований к выборке, программному обеспечению, а также влияние разных алгоритмов формирования пар. Указанными недостатками, вероятно, можно объяснить получение разных результатов исследований, включающих примерно сходную популяцию пациентов с ФП.

В качестве примера можно отметить неожиданные результаты ретроспективного анализа баз данных страховых компаний, который был выполнен в Германии с использованием метода PS [30]. В ходе выполнения анализа, включавшего данные о 37439 пациентах с ФП, были получены данные о том, что прием варфарина по сравнению с ППОАК сопровождается меньшей часто-

той развития таких неблагоприятных исходов, как смертность от любых причин, ишемический инсульт, инсульт (без уточнения его природы), транзиторная ишемическая атака, инфаркт миокарда, артериальная эмболия, геморрагический инсульт и тяжелое кровотечение (относительное увеличение риска (случаи на 100 пациентолет) для таких исходов при применении ППОАК достигало 1,22, 1,90, 2,04, 1,52, 1,26, 1,75, 0,94 и 1,92%, соответственно). Причем во всех случаях различия достигали уровня статистической значимости. Такие результаты явно не соответствуют данным, полученным в ходе выполнения крупных РКИ, по сравнительной оценке применения варфарина и ППОАК у пациентов с ФП. Следует отметить, что при подборе соответствующих пар с помощью метода PS из когорты пациентов, принимавших варфарин, и когорты пациентов, принимавших ППОАК, были исключены более 91 тыс. и 14 тыс. пациентов, соответственно. Очевидно, что такой "тримминг" данных определяет высокую вероятность отклонения результата от реального.

Такие же неожиданные, можно даже сказать неправдоподобные, данные недавно были получены в ходе выполнения обсервационного исследования с использованием метода *PS matching* [31]. Полученные результаты свидетельствовали о том, что применение тикагрелора в клинической практике не имеет преимуществ по сравнению с приемом клопидогрела по влиянию на риск развития тяжелых осложнений коронарной болезни сердца в течение года после стентирования коронарной артерии по поводу острого коронарного синдрома. Причем исходная выборка была почти в 2 раза меньше (n=11185) по сравнению с выборкой РКИ PLATO по сравнительной оценке применения тикагрелора и клопидогрела (n=18624), а после "триминга" исходных данных для формирования групп число пациентов в каждой группе уменьшилось до 3711 человек. Очевидно, что результаты такого исследования не могут считаться надежными по указанным ранее причинам.

Следует отметить, что источник финансирования подобных исследований также может влиять на полученные результаты и выводы, т.к. изменение условий анализа при наличии определенного конфликта интересов может влиять на полученные данные.

Например, результаты анализа, выполненного Jonasson C, et al. [32] с использованием метода PS, в который были включены данные о более чем 52 тыс. пациентов с ФП, свидетельствовали о менее высоком риске развития кровотечений при применении апиксабана и дабигатрана по сравнению с ривароксабаном. В разделе "финансирование исследования" представлена информация о финансовой поддержке проекта со стороны фармацевтических компаний, которые производят или организуют продажу апиксабана: The study received additional support through the Bristol-Myers Squibb/Pfizer-sponsored European Investigator Initiated research Program (ERISTA) (Grant number 2016-ELI-0407). Очевидно, что такой явный конфликт интересов становится основанием для сомнений в обоснованности полученных результатов.

Практические подходы к интерпретации данных наблюдательных исследований

Развитие технологий для анализа информации привело к колоссальному росту публикаций результатов исследований, выполненных с использованием медицинских баз данных. Как следствие в ежедневной работе врачу все чаще приходится сталкиваться с анализом научной информации для принятия клинических решений. Как выработать для себя упрощенную схему для оценки исследований с различным дизайном, в частности, наблюдательных и рандомизированных исследований? Учитывая целый ряд методологических ограничений, описанных ранее в статье, результаты наблюдательных исследований имеют скорее гипотетический характер и должны интерпретироваться с осторожностью. Наблюдательные исследования могут быть рассмотрены клиницистами для получения информации о редких клинических ситуациях, которые не были подробно изучены в крупных исследованиях (например, больные со специфической сопутствующей патологией). В тех же областях медицины, где имеются и данные высококачественных РКИ, и доступны результаты наблюдательных исследований (например, тактика антикоагулянтной терапии у пациентов с неклапанной ФП) следует ориентироваться именно на РКИ, как на "золотой стандарт" доказательной медицины [33]. Только специально запланированные РКИ могут быть базой для сравнения относительной эффективности и безопасности нескольких исследуемых вмешательств, поскольку только в них рандомизация способна в значительной степени "уравнять" различающиеся характеристики пациентов в двух группах сравнения, способные повлиять на результат. Так, эксперты Европейского общества кардиологов отметили, что рандомизированные клинические исследования должны занимать центральное место в контексте предоставления высококачественной доказательной базы для оценки эффективности и безопасности терапевтических вмешательств [34]. Наконец, важно помнить, что, в отличие от РКИ, результаты наблюдательных исследований, в т.ч. крупных анализов "больших данных", сравнивающих несколько препаратов между собой, не могут опровергать результаты РКИ, а только подтверждать или дополнять их.

Заключение

Таким образом, единственно надежным и научно обоснованным подходом к оценке эффективности и безопасности лечебных вмешательств может считаться только выполнение РКИ с достаточной стати-

стической мощностью в отсутствие явных методологических недостатков. Обсервационные исследования играют важную роль как источник доказательной информации для выявления редких побочных эффектов терапии или выявления эффектов терапии в ходе длительного наблюдения или у пациентов с характеристиками, которые существенно отличаются от таковых у участников РКИ. Однако обсервационные исследования даже при использовании современных методов, позволяющих уменьшить различия между группами по известным исходным характеристикам, нельзя использовать для сравнения эффективности и безопасности применения лекарственных препаратов. Это обусловлено тем, что в обсервационном исследовании никогда нельзя обеспечить отсутствие различий между группами по тем характеристикам, которые либо не измерялись, либо которые в принципе нельзя учесть. А раз так, в ходе выполнения любого обсервационного исследования нельзя устранить систематическую ошибку, обусловленную различиями по исходным характеристикам пациентов в группах. В то же время в достаточно крупном РКИ группы не различаются как по известным, так и по неустановленным или трудно учитываемым характеристикам, т.е. исключается главный источник систематических ошибок при сравнении эффектов вмешательств. При знакомстве с любым источником доказательной информацию необходимо критически оценивать ее качество, т.е. используя терминологию Даниэля Канемана, включать Систему 2, которая должна помочь уточнить первое впечатление о результатах любого исследования, которое формирует “доверчивая” Система 1.

Отношения и деятельность. Статья подготовлена при поддержке компании АО “Байер”. Номер одобрения: PP-XAR-RU-0493-1.

Литература/References

1. Kahneman D. Thinking, fast and slow. M:AST. 2014. 656 p. (In Russ.) Канеман Д. Думай медленно, решай быстро. пер. с англ.; Москва: АСТ, 2014. 656 с. ISBN: 978-5-17-080053-7.
2. Fanaroff AC, Califf RM, Lopes RD. New Approaches to Conducting Randomized Controlled Trials. J Am Coll Cardiol. 2020;75:556-9. doi:10.1016/j.jacc.2019.11.043.
3. Murray KW, Duggan A. Understanding Confounding in Research Pediatr Rev. 2010;31(3):124-6. doi:10.1542/pir.31-3-124.
4. Meuli L, Dick F. Understanding Confounding in Observational Studies. Eur J Vasc Endovasc Surg. 2018;55(5):737. doi:10.1016/j.ejvs.2018.02.028.
5. Fanaroff AC, Califf RM, Windecker S, et al. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. JAMA. 2019;321:1069-80. doi:10.1001/jama.2019.1122.
6. Schüpke S, Neumann FJ, Menichelli M, et al; ISAR-REACT 5 Trial Investigators. Ticagrelor or Prasugrel in Patients with Acute Coronary Syndromes. N Engl J Med. 2019;381:1524-34. doi:10.1056/NEJMoa1908973.
7. Nanna MG, Navar AM, Wang TY, et al. Practice-level variation in statin use and low-density lipoprotein cholesterol control in the United States: results from the Patient and Provider Assessment of Lipid Management (PALM) registry. Am Heart J. 2019;214:113-24. doi:10.1016/j.ahj.2019.05.009.
8. Harrington D, D'Agostino RBSr, Gatsonis C, et al. New Guidelines for Statistical Reporting in the Journal. N Engl J Med. 2019;381:285-6. doi:10.1056/NEJMe1906559.
9. Mehra MR, Desai SS, Kuy S, et al. Cardiovascular disease, drug therapy, and mortality in Covid-19. N Engl J Med. published online May 1. doi:10.1056/NEJMoa2007621.
10. Mehra MR, Desai SS, Ruschitzka F, Patel AN. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020; published online May 22. doi:10.1016/S0140-6736(20)31180-6.
11. Mehra MR, Desai SS, Kuy S, et al. Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. doi:10.1056/NEJMoa2007621 [ahead of print, 2020 Jun 4] [retraction of: N Engl J Med. 2020 May 1]. N Engl J Med . 2020;382(26):2582. doi:10.1056/NEJMc2021225.
12. Mehra MR, Ruschitzka F, Patel AN. Retraction-Hydroxychloroquine or Chloroquine With or Without a Macrolide for Treatment of COVID-19: A Multinational Registry Analysis. Retraction of Publication Lancet. 2020 Jun 5;S0140-6736(20)31324-6. doi:10.1016/S0140-6736(20)31324-6. Ahead of print.
13. Hellenbart EL, Faulkenberg KD, Finks SW. Evaluation of bleeding in patients receiving direct oral anticoagulants. Vasc Health Risk Manag. 2017;13:325-42. doi:10.2147/VHRM.S121661.
14. Guimarães PO, Krishnamoorthy A, Kaltenbach LA, et al. Accuracy of Medical Claims for Identifying Cardiovascular and Bleeding Events After Myocardial Infarction: A Secondary Analysis of the TRANSLATE-ACS Study. JAMA Cardiol. 2017;2(7):750-7. doi:10.1001/jamacardio.2017.1460.
15. Fralick M, Colacci M, Schneeweiss S, et al. Effectiveness and Safety of Apixaban Compared With Rivaroxaban for Patients With Atrial Fibrillation in Routine Practice: A Cohort Study. Ann Intern Med. 2020;172(7):463-73. doi:10.7326/M19-2522.
16. Kuss O, Blettner M, Börgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl Int. 2016;113(35-36):597-603. doi:10.3238/arztebl.2016.0597.
17. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312:1215-8. doi:10.1136/bmj.312.7040.1215.
18. McKee M, Britton A, Black N, et al. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. BMJ. 1999;319:312-5. doi:10.1136/bmj.319.7205.312.
19. Rothwell PM. External validity of randomised controlled trials: „to whom do the results of this trial apply?“. Lancet. 2005;365:82-93. doi:10.1016/S0140-6736(04)17670-8.
20. Camm AJ, Amarenco P, Haas S, et al; XANTUS Investigators. XANTUS: a real-world, prospective, observational study of patients treated with rivaroxaban for stroke prevention in atrial fibrillation. Eur Heart J. 2016;37(14):1145-53. doi:10.1093/eurheartj/ehv466.
21. Gueyffier F, Cucherat M. The limitations of observation studies for decision making regarding drugs efficacy and safety. Therapie. 2019;74:181-5. doi:10.1016/j.therap.2018.11.001.
22. Packer M. Are Meta-Analyses a Form of Medical Fake News? Thoughts About How They Should Contribute to Medical Science and Practice. Circulation. 2017;136:2097-9. doi:10.1161/CIRCULATIONAHA.117.030209.
23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41-55. doi:10.1093/biomet/70.1.41.
24. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol. 2006;163:1149-56. doi:10.1093/aje/kwj149.
25. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Med Decis Making. 2009;29:661-77. doi:10.1177/0272989X09341755.
26. Stuart EA, Marcus SM, Horvitz-Lennon MV, et al. Using non-experimental data to estimate treatment effects. Psychiatr Ann. 2009;39:719-28. doi:10.3928/00485713-20090625-07
27. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg. 2007;134:1128-35. doi:10.1016/j.jtcvs.2007.07.021.
28. Morgan SL, Harding DJ. Matching estimators of causal effects: prospects and pitfalls in theory and practice. Sociological Methods Res. 2006;35:3-60.
29. Weitzen S, Lapane KL, Toledano AY, et al. Weaknesses of goodness of fit tests for evaluating propensity score models: the case of the omitted confounder. Pharmacoepidemiol Drug Saf. 2005;14:227-38. doi:10.1002/pds.986.
30. Mueller S, Groth A, Spitzer SG, et al. Real-world Effectiveness and Safety of Oral Anticoagulation Strategies in Atrial Fibrillation: A Cohort Study Based on a German Claims Dataset. Pragmat Obs Res. 2018;9:1-10. doi:10.2147/POR.S156521.
31. Turgeon RD, Koshman SL, Youngson E, et al. Association of Ticagrelor vs Clopidogrel With Major Adverse Coronary Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention. JAMA Intern Med. 2020;180:420-8. doi:10.1001/jamainternmed.2019.6447.
32. Jonasson C, Ghanima W, Söderdahl F, Halvorsen S. Comparison of Dabigatran, Rivaroxaban, and Apixaban for Effectiveness and Safety in Atrial Fibrillation: A Nationwide Cohort Study. Eur Heart J Cardiovasc Pharmacother. 2020;6:75-85. doi:10.1093/ehjcvp/pvz086.
33. Healey JS. Stroke Prevention in Atrial Fibrillation: What Is Real World, and What Do Real-World Data Reveal? J Am Coll Cardiol. 2018;72(2):154-5. doi:10.1016/j.jacc.2018.04.057.
34. Bowman L, Baras A, Bombien R, et al. Understanding the use of observational and randomized data in cardiovascular medicine. Eur Heart J. 2020;41(27):2571-8. doi:10.1093/eurheartj/ehaa020.