

Современные подходы к интерпретации результатов крупных рандомизированных клинических исследований

Гиляревский С.Р.*

Российская медицинская академия непрерывного профессионального образования, Москва, Россия

В статье обсуждаются современные подходы к интерпретации результатов рандомизированных исследований, которые позволяют уменьшить искажения информации при их представлении широкому кругу врачей. Приводятся данные о роли учета множественных сравнений при планировании и анализе крупных рандомизированных клинических исследований. Подробно рассматривается обоснованность и подходы к использованию иерархического подхода к оценке статистической значимости показателей («конечных точек») в тех случаях, когда в соответствии с заранее принятым протоколом исследования предполагается оценивать несколько показателей. Рассматриваются подходы к интерпретации значения p , а также 95% доверительных интервалов. Особое внимание в статье уделяется интерпретации компонентов основного комбинированного показателя, так как при интерпретации именно таких данных возможны искажения представления результатов исследования и манипуляции мнением врачей. Особенно это касается данных о смертности и смертности от осложнений сердечно-сосудистых заболеваний. Обсуждаются современные подходы к анализу результатов исследования с помощью методов, отличных от привычного анализа Каплана-Мейера, в частности, метода, основанного на расчете показателя «win ratio». В качестве примеров приводятся данные, полученные в современных клинических исследованиях.

Ключевые слова: рандомизированные клинические исследования, комбинированные показатели, учет множественных сравнений, иерархический подход.

Для цитирования: Гиляревский С.Р. Современные подходы к интерпретации результатов крупных рандомизированных клинических исследований. *Рациональная Фармакотерапия в Кардиологии* 2022;18(3):311-319. DOI:10.20996/1819-6446-2022-05-01.

Modern Approaches to the Interpretation of the Results of Large Randomized Clinical Trials

Gilyarevsky S.R.*

Russian Medical Academy of Continuous Professional Education, Moscow, Russia

Modern approaches to interpreting the results of randomized trials, which can reduce the distortion of information when they are presented to a wide range of doctors, are discussed in the article. Data on the role of taking into account multiple comparisons in the planning and analysis of large randomized clinical trials are given. The validity and approaches to using a hierarchical approach to assessing the statistical significance of indicators ("endpoints") in cases where, in accordance with a pre-adopted study protocol, several indicators are supposed to be evaluated are considered in detail. Approaches to interpreting the p value as well as 95% confidence intervals are considered. Particular attention is paid to the interpretation of the components of the main combined indicator, since when interpreting just such data, the presentation of the results of the study and the manipulation of doctors' opinions may be distorted. This is especially true for data on mortality and mortality from complications of cardiovascular diseases. Modern approaches to the analysis of research results using methods other than the usual Kaplan-Meier analysis, in particular, the method based on the calculation of the "win ratio" are discussed. Data obtained in modern clinical studies are given as examples.

Keywords: randomized clinical trials, combined scores, multiple comparisons, hierarchical approach.

For citation: Gilyarevsky S.R. Antiarrhythmic and Hemodynamic Effects of Omega-3 Polyunsaturated Fatty Acids on Cardiovascular Diseases. *Rational Pharmacotherapy in Cardiology* 2022;18(3):311-319. DOI:10.20996/1819-6446-2022-05-01.

*Corresponding Author (автор, ответственный за переписку): sgilarevsky@rambler.ru

Введение

Для внедрения в клиническую практику новых методов лечения требуется подтверждение их эффективности и безопасности по сравнению со стандартной терапией или плацебо. Такое подтверждение возможно только в ходе выполнения правильно спланированных и имеющих достаточную статистическую мощность рандомизированных клинических исследований (РКИ). Для выполнения таких исследований требуются большие финансовые затраты и использование значительных трудовых ресурсов, а также большой период на-

блюдения за участниками исследований. Все эти факторы ограничивают внедрение новых методов лечения в клиническую практику. Упрощение протоколов исследований, которые нередко используют для преодоления указанных проблем, почти всегда приводит к снижению качества исследования, соответственно, и к сложности (или даже невозможности) трактовки его результатов. В такой ситуации особенно важным представляется использование принятых экспертами правил анализа результатов крупных РКИ, соблюдение которых позволяет свести к минимуму возможность неточной интерпретации важной доказательной информации. В этой статье будут обсуждаться правила оценки результатов таких исследований, несоблюдение

Received/Поступила: 25.02.2022

Accepted/Принята в печать: 06.04.2022

которых может привести к их искаженному представлению, а также может становиться источником манипуляции мнением врачей.

В частности, предполагается обсудить роль множественных сравнений при анализе результатов крупных РКИ, а также важность учета таких сравнений при интерпретации комбинированных показателей (или так называемых «конечных точек») частоты развития определенных клинических исходов. Отдельно в статье рассматриваются современные принципы оценки комбинированных показателей и их компонентов. Кроме того, в статье будут представлены современные подходы к оценке не только первых развившихся клинических исходов, которая обычно используется при использовании метода Каплана-Мейера, но и последующих неблагоприятных клинических исходов, т.е. подходы, которые позволяют точнее представить частоту развития наиболее значимых клинических исходов.

Проблема множественных сравнений и подходы к ее решению

Одна из главных задач при проведении современных крупных РКИ состоит в создании условий, которые уменьшают вероятность так называемой ошибки I типа, т.е. позволяют избежать получения ложноположительных результатов исследования. Если в исследовании эффект лечения оценивают по нескольким клиническим исходам или показателям, вероятность ошибки I типа возрастает. Это обусловлено тем, что при анализе большого числа показателей увеличивается вероятность того, что случайно будут получены статистически значимые различия между группами по какому-либо показателю [1].

Один из подходов к уменьшению вероятности такого случайного результата состоит в использовании поправок, направленных на увеличение порогового уровня статистической значимости таким образом, чтобы суммарное значение p не превышало общепринятого уровня 0,05. Известным примером таких поправок могут служить поправки Бонферрони. Например, если необходимо оценить 10 показателей, то для каждого из таких показателей статистически значимыми различиями между группами будут различия при $p < 0,005$, а не $p < 0,05$, т.е. будет повышен порог статистической значимости. Таким образом, используя поправки Бонферрони для множественных сравнений, при определении порога статистической значимости для каждого из оцениваемых показателей значение $p = 0,05$ следует разделить на число оцениваемых показателей. Для указанного примера значение p для каждого из таких показателей будет составлять $0,05 / 10 = 0,005$.

Следует отметить, что недостатком такого подхода может быть увеличение вероятности ошибки II типа, т.е. вероятности получения ложноотрицательных ре-

зультатов, и, следовательно, увеличение вероятности невыявления действительно существующих преимуществ вмешательства. Для преодоления таких недостатков используют иерархический подход к оценке нескольких показателей. В чем состоят основные принципы реализации такого иерархического подхода?

Прежде всего, все оцениваемые показатели или клинические исходы заранее располагаются в протоколе исследования в иерархическом порядке. Первый показатель, т.е. расположенный выше всех в иерархии, оценивается как обычный показатель, и статистически значимыми считаются различия между группами при $p < 0,05$. Если по этому первому показателю не достигаются статистически значимые различия между группами, дальнейшая процедура оценки по другим показателям прекращается, и вмешательство в ходе выполнения данного исследования считается неэффективным.

Если по первому показателю выявлены статистически значимые различия, то переходят к ответу на вопрос о наличии хотя бы еще одного преимущества применения изучаемого вмешательства. Поскольку другие клинические исходы или показатели заранее были расположены в иерархическом порядке, исключается «расстановка сетей» для выявления дополнительных преимуществ. Проверка гипотезы о преимуществах вмешательства по второму оцениваемому показателю также выполняется с использованием общепринятого подхода к оценке статистической значимости, при котором различия между группами считаются статистически значимыми при $p < 0,05$.

Если результаты такой проверки свидетельствуют о статистически значимых различиях между группами и по второму показателю, приступают к проверке преимуществ по показателю, занимающему третье место в иерархии. Такой алгоритм повторяют для всех остальных показателей, которые предполагалось оценивать в заранее определенном иерархическом порядке до тех пор, пока не будут проверены все показатели, или не будет установлено отсутствие статистически значимых различий между группами по одному из показателей, оцениваемых с помощью иерархического подхода. Как только такое отсутствие по определенному показателю будет установлено, дальнейшая оценка показателей прекращается. В таких случаях даже если при анализе расположенных ниже в иерархии показателей будут установлены статистически значимые различия ($p < 0,05$), учитывать их нельзя, так как в противном случае это свидетельствовало бы о «расстановки сетей» для выявления положительных эффектов, которые при таком подходе могут быть просто случайными.

К преимуществам иерархического подхода относят возможность установления в ходе выполнения одного исследования преимуществ изучаемого вмешательства по нескольким показателям или клиническим исходам

в отсутствие увеличения риска допущения ошибки I типа, т.е. риска преувеличения истинных результатов исследования. Взаимосвязь анализа различных показателей с использованием иерархического подхода позволяет достаточно надежно предотвращать возможность получения ложноположительных результатов при оценке дополнительных преимуществ вмешательства. Именно поэтому такой метод стал широко применяться после его одобрения в 2002 г. представителями европейских регуляторных органов при выполнении множественных статистических тестов как рекомендуемый к использованию в РКИ для учета множественных сравнений [2].

Наконец, иерархический метод может использоваться в различных ситуациях, где необходим учет множественных сравнений, в том числе, при анализе комбинированных показателей, при выполнении анализа в подгруппах и т.д. Например, первый тест в иерархии может состоять в проверке гипотезы о не менее высокой эффективности (non-inferiority) изучаемого вмешательства по сравнению со стандартным при оценке с помощью комбинированного показателя, а второй тест – состоять в анализе того же или другого показателя, но для проверки преимуществ изучаемого вмешательства (superiority).

Процедура иерархического теста в настоящее время считается единственным приемлемым методом, который позволяет установить дополнительные преимущества вмешательства. Его использование дает возможность получить свидетельства нескольких преимуществ изучаемого в одном исследовании вмешательства.

Методологические подходы к использованию иерархического теста

Как уже указывалось, в ходе выполнения анализа гипотезы проверяют в определенном порядке одну за другой в соответствии с иерархией, указанной в протоколе исследования. Соответственно, как указывалось ранее, не для всех показателей, при сравнении которых p будет менее порогового уровня 0,05, различия будут считаться статистически значимыми, так как преимущество будет считаться доказанным только для тех случаев, когда p будет менее 0,05, но были получены статистически значимые различия по показателю, расположенному выше в иерархии. Выбор показателей, которые будут анализироваться с помощью иерархического подхода и установление порядка показателей в иерархии представляют собой самостоятельные задачи. Остановимся вкратце на том, как решаются такие задачи.

Как выбирают показатели, анализируемые с помощью иерархического подхода?

Выбор клинических исходов или показателей, которые предполагается анализировать с помощью иерар-

хического подхода, представляет собой несложную задачу, так как положение показателя или исхода определяется преимуществами вмешательства, которые хотят установить в первую очередь. Как и в других случаях выбора критериев оценки оцениваемые с помощью иерархического подхода показатели выбирают в зависимости от задаваемого клинического вопроса, и таким образом они отражают цели лечения пациентов с определенным заболеванием и/или характеристиками. Показатели безопасности также могут оцениваться с использованием иерархического подхода в тех случаях, когда хотят оценить дополнительное влияние вмешательства на риск развития нежелательных явлений, связанных с его применением. Пример алгоритма выполнения анализа показателей с использованием иерархического подхода представлен на рис. 1.

Как располагают показатели в иерархии?

Подходы к выбору первого в иерархии показателя такие же, как в целом при выборе основного показателя, т.е. тщательно отбирают клинические исходы, которые имеют наибольшее значение в данной клинической ситуации и в наибольшей степени соответствуют целям лечения данного заболевания, но в то же время будут достаточно общими для того, чтобы на основании результатов можно было сделать обобщающий вывод. При расчете объема выборки в первую очередь учитывают частоту развития исходов, включенных в такой показатель, и предполагаемую выраженность влияния изучаемого вмешательства на такие исходы.

Для остальных показателей или исходов, оцениваемых с помощью иерархического подхода, их предполагаемая частота должна быть основным критерием, так как считается, что именно такие исходы представляют клинический интерес. Рекомендуют рассчитывать выборку отдельно для каждого показателя, оцениваемого в иерархическом порядке, с целью оценки необходимого числа участников исследования с учетом предполагаемых результатов, а также и для того, чтобы иметь возможность на основании полученных расчетов своевременно изменить порядок показателей/исходов в иерархии. При построении иерархии следует учитывать возможное отсутствие статистической мощности исследования для оценки показателей, включенных в иерархию. Поэтому необходимо найти компромисс для того, чтобы упорядочить эти показатели, но при этом сохраняя по возможности значимость их для пациентов.

Например, более приемлемым, с точки зрения исследователя, в качестве основного показателя будет использование комбинированного показателя смертности и частоты развития осложнений заболевания, так как частота развития исходов в таком случае наибольшая. Следовательно, исходы, частота которых мень-

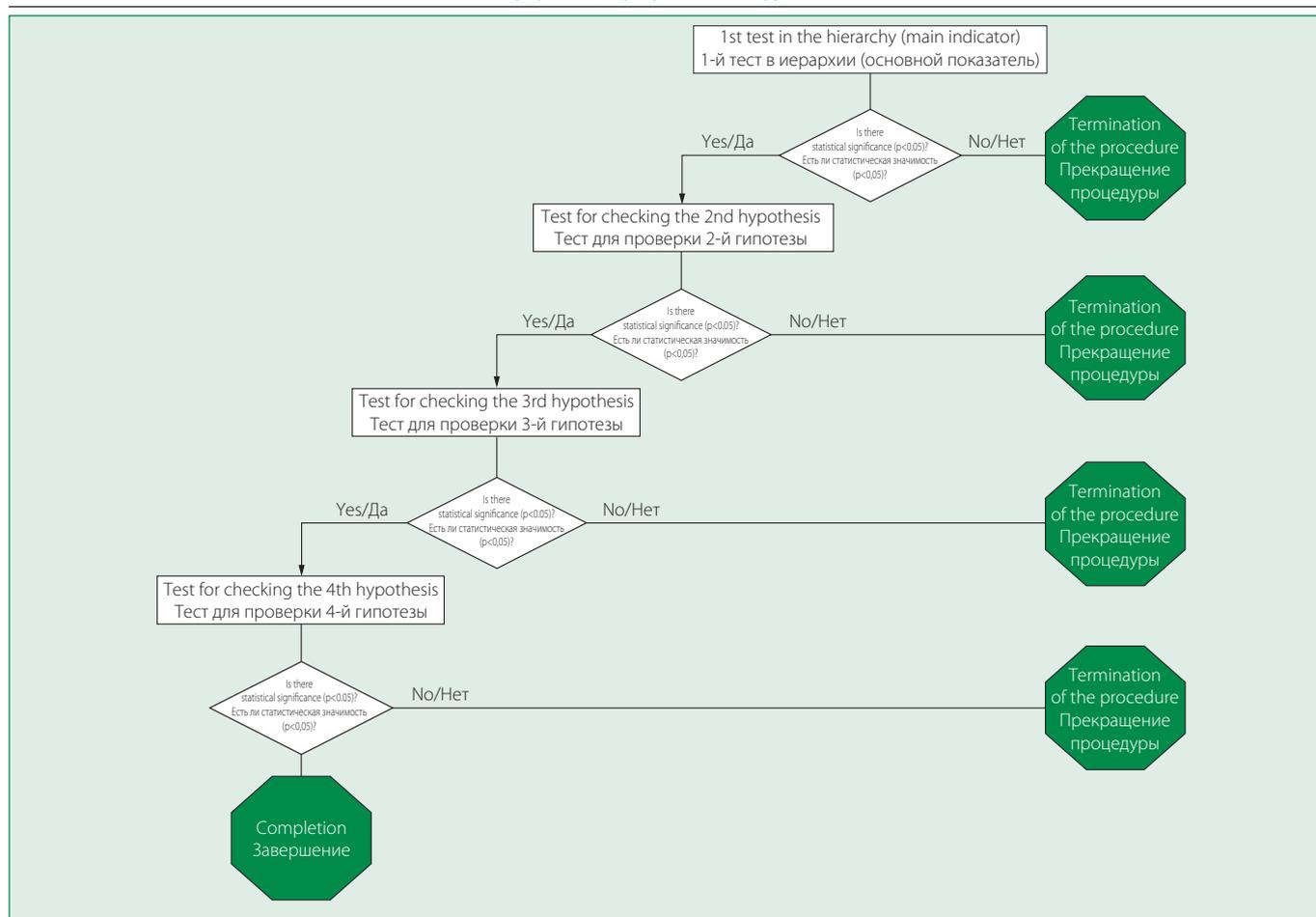


Figure 1. Example of algorithm for analyses using hierarchy approach

Рисунок 1. Пример алгоритма выполнения анализа показателей с использованием иерархического подхода

ше (но клиническая значимость выше) должны располагаться в иерархии в более низком положении, так как для их оценки снижается статистическая мощность и, следовательно, точность установленного эффекта.

Однако считается, что порядок клинических исходов в иерархии может быть не так важен, так как цель анализа в целом состоит в выявлении клинических преимуществ вмешательства по влиянию на эффективность и, возможно, безопасность. Главная цель использования иерархического подхода состоит в исключении неограниченного поиска хотя бы каких-нибудь положительных эффектов вмешательства, так как установление положительного эффекта в таком случае может быть случайным. В табл. 1 представлены основные задачи и принципы использования иерархического подхода к оценке нескольких клинических исходов в РКИ.

Анализ результатов исследования EMPA-REG OUTCOME и использование этических принципов их представления

Исследование EMPA-REG OUTCOME было выполнено для оценки эффективности и безопасности при-

менения эмпаглифлозина по 10 и 25 мг 1 р/сут по сравнению с плацебо у пациентов с сердечной недостаточностью и сниженной фракцией выброса (СНсНФВ) левого желудочка [3]. Основным комбинированный показатель включал смертность от осложнений сердечно-сосудистых заболеваний (ССЗ), частоту развития несмертельного инфаркта миокарда или несмертельного инсульта, который анализировали в целом для пациентов, принимающих эмпаглифлозин (независимо от его дозы) или плацебо. Главный дополнительный комбинированный показатель включал исходы, учитывающиеся в основном показателе, и частоту госпитализаций по поводу нестабильной стенокардии.

Основная цель исследования состояла в проверке гипотезы о том, что применение эмпаглифлозина (в целом в дозе 10 или 25 мг/сут) будет не менее эффективно по влиянию на основной показатель по сравнению с плацебо при допустимой верхней границе доверительного интервала (ДИ) отношения шансов 1,3. При сравнении суммарных данных о пациентах, принимавших эмпаглифлозин по 10 или 25 мг/сут, и пациентов, принимавших плацебо, использовали 4-х этапную так-

Table 1. Key objectives and principles for using a hierarchical approach to assess multiple clinical outcomes in randomized clinical trials

Таблица 1. Основные задачи и принципы использования иерархического подхода к оценке нескольких клинических исходов в РКИ

Основные задачи и принципы

- Использование процедуры иерархического тестирования направлено на выявление дополнительных эффектов вмешательства.
- Представление результатов оценки показателей, которые в иерархии расположены ниже показателя, для которого не была установлена статистическая значимость, может быть только описательным, как и для показателей, которые не оценивали с использованием иерархического тестирования.
- Процедура иерархического тестирования в целом приемлема с точки зрения статистики, и ее применение позволяет учитывать возможный риск общей погрешности, но использование такого метода не исключает возможность или необходимость обсуждения наличия систематических ошибок в исследовании, а также клиническое значение выбранных показателей или клинической значимости влияния вмешательства на определенные показатели.
- Не все оцениваемые показатели должны быть включены в иерархию, но на результаты, полученные без использования иерархического подхода, нельзя ссылаться, они должны считаться лишь описательными.
- На такие правила следует обращать особое внимание при подготовке материалов и представлении результатов исследований, особенно, производителями медицинской продукции.

тику иерархического тестирования в следующем порядке: (1) проверка гипотезы о не менее высокой эффективности применения эмпаглифлозина по сравнению с плацебо по влиянию на основной показатель (non-inferiority); (2) проверка гипотезы о не менее высокой эффективности применения эмпаглифлозина по сравнению с плацебо по влиянию на главный дополнительный показатель (non-inferiority); (3) проверка гипотезы о более высокой эффективности применения эмпаглифлозина по сравнению с плацебо по влиянию на основной показатель (superiority); (4) проверка гипотезы о более высокой эффективности применения эмпаглифлозина по сравнению с плацебо по влиянию на главный дополнительный показатель (superiority).

Следует отметить, что статистическая значимость различий по основному показателю по данным анализа, выполненного для проверки гипотезы о преимуществах приема эмпаглифлозина по сравнению с плацебо, достигала уровня статистической значимости, которая была очень близка к пограничной ($p=0,04$), а по данным такого анализа для главного дополнительного показателя различия между группами не достигали уровня статистической значимости ($p=0,08$).

Представление всех других показателей, оцениваемых в исследовании EMPA-REGOUTCOME, включая общую смертность и смертность от осложнений ССЗ, может быть только описательным, так как анализ таких показателей без использования иерархического подхода не позволяет исключить случайный характер полученных результатов.

Как подходить к анализу комбинированных показателей?

Использование комбинированных показателей для оценки результатов РКИ важно для достижения в ходе выполнения исследования числа клинических исходов, которое необходимо для обеспечения статистической мощности исследования [4].

Сочетание двух или более клинических исходов в одном показателе увеличивает частоту анализируемых исходов и повышает вероятность выявления различий между группами по основному или дополнительному показателю. Такое объединение нескольких клинических исходов в одном показателе приведет к большей частоте развившихся исходов и повысит статистическую точность исследования, а также будет обуславливать возможность включения в исследование меньшего числа пациентов и, соответственно, снижение затрат на его выполнение, а также позволит завершить исследование в более ограниченные сроки [4,5].

Тем не менее, интерпретация комбинированных показателей в некоторых случаях становится сложной задачей, так как отсутствуют единые стандартизованные подходы к интерпретации таких показателей. В то же время считается, что трактовка таких показателей, как если бы они были одним клиническим исходом, представляется не совсем правильным подходом к интерпретации результатов исследований, в которых использовали комбинированные показатели [6]. В настоящее время имеются лишь ограниченные рекомендации по интерпретации результатов, полученных с помощью комбинированных показателей [7].

Одна из трудностей интерпретации комбинированных показателей состоит в необходимости интерпретации комбинированного показателя, включающего клинические исходы, которые могут не иметь большого клинического значения. Сочетание в одном показателе нескольких клинических исходов или показателей, имеющих большую вариабельность клинической значимости (как для врачей, так пациентов), вызывает опасения при попытке использования комбинированного показателя как основы для принятия клинического решения. Сходные опасения возникают в тех случаях, когда имеются существенные различия по частоте развития большинства или хотя бы наиболее важных клинических исходов, а также вариабельность компонентов комбинированного показателя.

Подходы к интерпретации комбинированных показателей

Окончательный вопрос, который врачи должны задавать при оценке исследований, в которых использовали комбинированный показатель, таков: позволит ли использование комбинированного показателя учитывать его при принятии клинического решения.

Для ответа на этот вопрос следует в свою очередь ответить на несколько принципиальных вопросов, касающихся комбинированного показателя [8]:

- Имеют все компоненты комбинированного показателя одинаковую значимость для пациентов?
- Имеется ли сходная частота развития более значимых и менее значимых клинических исходов?
- Можно ли быть уверенным, что для компонентов комбинированного показателя достигнуто сходное снижение относительного риска?
- Имеется ли сходная биологическая основа компонентов комбинированного показателя, при которой можно было бы предполагать сходное снижение относительного риска?
- Сходно ли снижение относительного риска компонентов комбинированного показателя и достаточно ли узки соответствующие ДИ?

При положительном ответе на такие вопросы в целом, можно использовать комбинированный показатель для принятия клинического решения. При отрицательных ответах при принятии такого решения, скорее, надо полагаться на анализ ДИ отдельных компонентов комбинированного показателя.

Одинаково ли значимы компоненты комбинированного показателя для пациентов?

В тех случаях, когда компоненты комбинированного показателя имеют примерно одинаковую клиническую значимость, не так важно, за счет какого компонента комбинированный показатель снижается в первую очередь, так как если различие между группами по такому показателю в целом достигает уровня статистической значимости, можно предполагать, что важный для пациента компонент сыграл существенную роль.

Однако, чем больше компоненты комбинированного показателя различаются по клинической значимости, тем больше сомнений будет по поводу пользы учета комбинированного показателя. Объединение в одном показателе наиболее клинически важного показателя (т.е. смертности) с таким относительно незначимым, или менее значимым показателем, как степень интенсификации терапии или потребность в госпитализациях, может создавать трудности в интерпретации результатов, полученных с помощью комбинированного показателя. Одним из примеров не совсем удачного набора компонентов комбинированного показателя можно считать исследование PREAMI (Perindopril and Re-

modelling in the Elderly with Acute Myocardial Infarction) [9], цель которого состояла в оценке эффективности приема периндоприла по сравнению с плацебо у пациентов, перенесших инфаркт миокарда. Основным комбинированный показатель включал общую смертность, частоту госпитализаций по поводу хронической сердечной недостаточности (ХСН), а также показатель ремоделирования левого желудочка, который диагностировали при увеличении конечного диастолического объема левого желудочка на 8% или более. В целом основной комбинированный показатель снижался статистически значимо за счет приема периндоприла, но такое снижение было достигнуто только за счет влияния на косвенный показатель ремоделирования левого желудочка в отсутствие статистически значимого или существенного влияния на смертность или частоту госпитализаций [10].

Если более клинически значимый клинический исход развивается намного реже, чем менее значимый, информативность комбинированного показателя существенно снижается. Таким образом очевидно, что чем больше различие по частоте развития наиболее значимых и наименее значимых для пациента клинических исходов, тем больше сомнений в пользе комбинированного показателя для принятия клинического решения.

Как при оценке основного показателя и его компонентов интерпретировать отношение риска и доверительные интервалы?

В ходе выполнения исследования DAPA-HF (Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure) [11], в которое были включены пациенты с СНснФВ, были получены данные о том, что в группе приема дапаглифлозина и группе плацебо частота развития неблагоприятных исходов, включенных в основной комбинированный показатель, достигла 16,3 и 21,2% соответственно, снижение абсолютного риска составило 4,9%.

Можно ли полностью доверять оценкам эффектов вмешательства по сравнению с контролем (с плацебо или препаратом сравнения), которые основаны на анализе частоты развития неблагоприятных исходов в процентах? Любая такая оценка по определению таит в себе неточность в связи с всегда ограниченным числом пациентов, включенных в исследование, т.к. чем меньше исследование, тем меньше точность полученных результатов.

Такую статистическую неопределенность результатов можно преодолеть с помощью расчета 95% ДИ [12]. Как указывалось ранее, в исследовании DAPA-HF неблагоприятные исходы, включенные в основной показатель, в группе дапаглифлозина и группе плацебо

развились у 16,3% и 21,2% пациентов соответственно (отношение риска 0,74 при 95% ДИ от 0,65 до 0,85; $p < 0,001$). О чем говорят такие ДИ? Они отражают вариабельность полученного эффекта и указывают, в каких границах будет реально находиться такой эффект. При многократном повторении эксперимента (например, 100 раз или более) с таким же протоколом полученные различия по ДИ будут очень небольшими. То есть, ДИ отражает диапазон истинного снижения риска развития неблагоприятных исходов. При расчете 95% ДИ, тем не менее, следует помнить, что имеется 2,5% вероятность того, что часть истинных результатов будет располагаться слева от этого интервала, и 2,5% вероятность того, что они будут располагаться справа от него. В приведенном примере 95% ДИ от 0,65 до 0,85 указывает, что в таких границах истинные границы эффекта находится в диапазоне отношения рисков, то есть, если бы протокол, использованный в исследовании DAPA-HF, много раз был повторен в других исследованиях, границы эффекта находились бы в таком диапазоне.

Как представлять результаты анализа компонентов комбинированного показателя исследования DAPA-HF этично (то есть таким образом, чтобы не исказить смысл полученных результатов)? Результаты анализа свидетельствуют о снижении каждого из компонентов основного показателя, что указывает на сходное снижение риска развития неблагоприятных исходов, включенных в такой показатель. Однако нельзя ответить на вопрос о статистически значимом снижении таких исходов, несмотря на то, что верхняя граница 95% ДИ для каждого из них была меньше 1,0. Почему в таких случаях никогда нельзя говорить о статистически значимом снижении риска таких исходов, в том числе, о снижении риска смерти осложнений ССЗ? Ответ был дан ранее, но напомним, что каждый из компонентов основного показателя не анализировали с использованием иерархического подхода, и, следовательно, какие бы результаты ни были получены при анализе таких исходов, они могут быть случайными. Таким образом, при представлении результатов исследования DAPA-HF неэтично говорить, что «прием дапаглифлозина снижает риск смерти от осложнений ССЗ». Такое представление результатов исследования вводит в заблуждение.

О чем свидетельствует в данном случае значение $p < 0,001$? Первый этап анализа результатов крупного РКИ состоит в проверке нулевой гипотезы. В данном случае нулевая гипотеза состояла в том, что применение дапаглифлозина у пациентов с СНснФВ не будет отличаться по эффективности от приема плацебо. В ходе выполнения исследования такая нулевая гипотеза была отвергнута, так как прием дапаглифлозина оказался более эффективным по сравнению с плацебо.

Значение p позволяет ответить только на вопрос о том, насколько вероятно было бы получить те результаты, которые были получены, если бы нулевая гипотеза была верна. В данном случае такая вероятность составляет $< 0,1\%$, то есть, можно отвергнуть нулевую гипотезу, так как ее правильность при полученных результатах крайне низка, а значит, низка вероятности случайности полученных результатов.

Однако в тех случаях, когда значение p превышает принятый уровень статистической значимости (обычно 0,05), то есть, нулевая гипотеза не отвергнута, нельзя считать, что нулевая гипотеза верна [13]. Более того, при достаточно большом объеме выборки в большинстве случаев можно отвергнуть нулевую гипотезу, так что значение p будет говорить не только о правдивости гипотезы, но и о выборке [14, 15].

Следует подчеркнуть, что значение p не позволяет ответить на вопрос о выраженности эффекта или о подтверждении основной гипотезы исследования. Кроме того, при повторении протокола исследования можно получить разные значения p . Использование ДИ для оценки эффективности вмешательств позволяет преодолеть такие ограничения, так как позволяет ответить на вопрос как о выраженности эффекта, так и о результатах, которое были бы получены при многократном воспроизведении исследования с таким же протоколом.

Преимущества оценки частоты повторных госпитализаций и методологические подходы к ее выполнению

Пациенты с ХСН нередко госпитализируются повторно, что в целом отражает менее благоприятное течение заболевания. Однако в большинстве современных исследований повторные неблагоприятные исходы, в частности, госпитализации, не учитываются [16]. Оценка частоты повторных госпитализаций в течение многих лет в большинстве крупных РКИ не проводилась. В исследовании CHARM-Preserved только 53% от общего числа госпитализаций по поводу декомпенсации сердечной недостаточности (ГДСН) и 57% от общего числа смертельных исходов были включены в основной показатель, который анализировался с использованием стандартного подхода, т.е., с оценкой продолжительности периода до развития первого неблагоприятного исхода, включенного в такой показатель. Очевидно, что учет повторно развивающихся исходов в целом лучше отражает клинически значимые исходы как с точки зрения врача, так и пациента, а также позволяет точнее оценить эффект вмешательства в течение всего периода наблюдения [17]. Следует также отметить, что при анализе повторных исходов существенно увеличивается статистическая мощность исследования, что позволяет, по крайней

мере, теоретически, почти в 2 раза уменьшить размер выборки по сравнению со стандартным подходом к анализу [18]. Так, в РКИ EMPHASIS-HF по оценке эффектов эплеренона по сравнению с плацебо у пациентов с ХСН анализ основного показателя включал оценку продолжительности периода до развития первого неблагоприятного исхода, включая период до первой ГДСН, но также в ходе выполнения анализа учитывались вторая и последующие ГДСН. Следует отметить, что в группе плацебо в исследовании EMPHASIS-HF повторные госпитализации составляли около 42% от общего числа ГДСН [19].

К сожалению, анализ всех ГДСН имеет ряд ограничений, включая вариабельность показаний к госпитализации и продолжительности пребывания пациента в стационаре, а также кластеризацию исходов в небольшой группе пациентов и влияние вмешивающихся факторов, связанных с различиями по смертности у пациентов, которые были госпитализированы, и пациентов, которые не госпитализировались [19].

Для преодоления недостатков анализа, основанного на оценке продолжительности периода до развития первого неблагоприятного исхода, был предложен метод, который называется «win ratio» («отношение числа победителей») [16,17,19,20]. При использовании такого метода пациенты двух групп формируются в пары в зависимости от риска развития неблагоприятного исхода. В РКИ по оценке эффективности вмешательств анализируют компоненты основного показателя, т.е. обычно число умерших от осложнений ССЗ и число ГДСН. В каждой паре пациентов выбирают «победителей» и «проигравших», если в данной паре был пациент, у которого первым развившимся исходом была смерть от осложнений ССЗ. Период, в течение которого устанавливают «победителя» и «проигравшего» в каждой паре, ограничивается периодом, в течение которого имеются данные для каждого из двух пациентов. Так, если один из пациентов умер в тот момент, когда данные для второго пациента отсутствуют, такой случай смерти уже не учитывается, но можно выполнить сравнение в такой паре по частоте госпитализаций или других показателям, и при различии по этому показателю в период, когда были данные об обоих пациентах, можно установить «победителя или проигравшего».

Если в анализируемой паре не было смертельных исходов, то выбирают «победителей» и «проигравших», если в данной паре был пациент, у которого первым развившимся исходом была ГДСН. Если в анализируемой паре не было исходов, включенных в основной показатель, для такой пары признается «ничья». «Win ratio» рассчитывают, как отношение общего числа «победителей» и общего числа «проигравших». Для такого отношения также легко рассчитать 95% ДИ и значение

p. В случае неприемлемости формирования пар, оценить показатель «win ratio» можно с помощью альтернативного подхода с использованием попарного сравнения без формирования групп со сходным риском. Считается, что в целом такой подход в большей степени важен для оценки числа смертельных исходов. Однако использование метода «win ratio» не исключает и оценку продолжительности периода до развития исхода.

В ходе выполнения международного рандомизированного двойного слепого плацебо-контролируемого исследования EMPULSE оценивали клинические преимущества и безопасность приема эмпаглифлозина по сравнению с плацебо у пациентов, госпитализированных по поводу острой декомпенсации сердечной недостаточности, а для анализа полученных данных использовали метод «win ratio». Пациентов включали в исследование независимо от ФВ ЛЖ и наличия или отсутствия сахарного диабета. Рандомизацию во всех случаях выполняли в период пребывания пациентов в стационаре после стабилизации их состояния. Результаты исследования оценивали с помощью суммарного показателя «win ratio», который рассчитывали с помощью попарного сравнения результатов с использованием стратификации с учетом впервые развившейся острой декомпенсации сердечной недостаточности или декомпенсации ХСН [21]. Полученные в исследовании данные свидетельствуют о клинических преимуществах приема эмпаглифлозина у госпитализированных пациентов с сохраненной и сниженной ФВ ЛЖ с впервые развившейся острой декомпенсацией сердечной недостаточности или с острой декомпенсацией ХСН по сравнению с плацебо независимо от наличия сахарного диабета 2-го типа. Таким образом, была подтверждена гипотеза об эффективности приема эмпаглифлозина при раннем его назначении на этапе стационарного лечения.

Заключение

В клинических исследованиях обычно используют комбинированные показатели, которые включают два компонента или более. Подразумевается, что использование комбинированных показателей повышает статистическую мощность исследования, позволяет уменьшить объем выборки, а также сократить сроки выполнения исследования и снизить затраты на его проведение. Однако польза от использования комбинированного показателя должна быть тщательно сопоставлена со сложностью интерпретации результатов, полученных с помощью такого показателя. Более того, чем больше различия компонентов комбинированного показателя по клинической значимости, частоте развития и полученным результатам для каждого из них, тем меньше информативность комбинированного по-

казателя и, следовательно, польза для принятия клинического решения. Представляется перспективным использование метода «win ratio» для оценки результатов РКИ, так как он обеспечивает более полную оценку частоты развития неблагоприятных клинических исходов, а также позволяет учитывать их клиническую

значимость даже при относительно небольшом числе пациентов, включенных в исследование.

Отношения и Деятельность. Нет.
Relationships and Activities. None.

References / Литература

1. Laporte S, Divin  M, Girault D. Clinical research and methodology: What usage and what hierarchical order for secondary endpoints? *Therapie*. 2016;71(1):27-41. DOI:10.1016/j.therap.2016.01.002.
2. Committee for Medicinal Products for Human Use (CHMP). Points to consider on multiplicity issues in clinical trials – CPMP/EWP/908/99; 2002 [cited 2022 Jan 10]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-multiplicity-issues-clinical-trials_en.pdf.
3. Zinman B, Wanner C, Lachin JM, et al.; EMPA-REG OUTCOME Investigators. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015;373(22):2117-28. DOI:10.1056/NEJMoa1504720.
4. Goldberg R, Gore JM, Barton B, Gurwitz J. Individual and composite study endpoints: separating the wheat from the chaff. *Am J Med*. 2014;127(5):379-84. DOI:10.1016/j.amjmed.2014.01.011.
5. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289(19):2554-9. DOI:10.1001/jama.289.19.2554.
6. McCoy CE. Understanding the Use of Composite Endpoints in Clinical Trials. *West J Emerg Med*. 2018;19(4):631-634. DOI:10.5811/westjem.2018.4.38383.
7. Armstrong PW, Westerhout CM. Composite End Points in Clinical Research: A Time for Reappraisal. *Circulation*. 2017;135(23):2299-307. DOI:10.1161/CIRCULATIONAHA.117.026229.
8. Guyatt G, Rennie D, Meade MO, Cook DJ eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. New York, NY: McGraw-Hill; 2015.
9. PREAMI: Perindopril and Remodelling in Elderly with Acute Myocardial Infarction: study rationale and design. *Cardiovasc Drugs Ther*. 2000;14(6):671-9. DOI:10.1023/a:1007879201470.
10. Ferrari R. Perindopril and Remodeling in Elderly with Acute Myocardial Infarction Investigators. Effects of angiotensin-converting enzyme inhibition with perindopril on left ventricular remodeling and clinical outcome: results of the randomized Perindopril and Remodeling in Elderly with Acute Myocardial Infarction (PREAMI) Study. *Arch Intern Med*. 2006;166(6):659-66. DOI:10.1001/archinte.166.6.659. PMID: 16567606.
11. McMurray JJV, Solomon SD, Inzucchi SE, et al.; DAPA-HF Trial Committees and Investigators. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2019;381(21):1995-2008. DOI:10.1056/NEJMoa1911303.
12. Bulpitt CJ. Confidence intervals. *Lancet*. 1987;1(8531):494-7. DOI:10.1016/S0140-6736(87)92100-3.
13. Cohen J. The Earth is round ($p < 0.05$). *Am Psychol*. 1994;49(12):997-1003. DOI:10.1037/0003-066X.49.12.997.
14. Cohen J. Things I have learned (so far). *Am Psychol*. 1990;45(12):1304. DOI:10.1037/0003-066X.45.12.1304.
15. Tukey JW. The philosophy of multiple comparisons. *Statist Sci*. 1991;6(1):100-16. DOI:10.1214/ss/1177011945.
16. Anker SD, McMurray JJ. Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials? *Eur Heart J*. 2012;33(22):2764-5. DOI:10.1093/eurheartj/ehs277.
17. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176-82. DOI:10.1093/eurheartj/ehs352.
18. Rogers JK, Pocock SJ, McMurray JJ, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail*. 2014;16(1):33-40. DOI:10.1002/ejhf.29.
19. Rogers JK, McMurray JJ, Pocock SJ, et al. Eplerenone in patients with systolic heart failure and mild symptoms: analysis of repeat hospitalizations. *Circulation*. 2012;126(19):2317-23. DOI:10.1161/CIRCULATIONAHA.112.110536.
20. Borer JS, B hm M, Ford I, et al.; SHIFT Investigators. Effect of ivabradine on recurrent hospitalization for worsening heart failure in patients with chronic systolic heart failure: the SHIFT Study. *Eur Heart J*. 2012;33(22):2813-20. DOI:10.1093/eurheartj/ehs259.
21. EMPULSE: Empagliflozin vs. Placebo in Improving Outcomes in Acute HF Patients [cited 2022 Jan 10]. Available from: <https://www.acc.org/latest-in-cardiology/articles/2021/11/08/12/02/sun-430pm-empulse-aha-2021>.

About the Authors / Сведения об Авторах:

Гиляревский Сергей Руджерович [Sergey R. Gilyarevsky]

ORCID 0000-0002-8505-1848