Словарь терминов и сокращений, используемых в доказательной медицине

Абсолютный риск, АР (absolute risk, AR): вероятность того, что у конкретного лица определенный клинический исход возникнет в течение определенного периода времени. Значения этого показателя находятся в диапазоне от 0 до 1, их также можно выразить в процентах. Этот термин используется в отношении как неблагоприятных клинических исходов (например, развитие инфаркта миокарда), так и благоприятных (например, выздоровление).

Анализ данных в зависимости от полученного вмешательства (on treatment analysis): анализ, в который включаются данные только об участниках, наблюдавшихся до конца исследования и получивших определенное вмешательство.

Анализ данных в зависимости от назначенного вмешательства (анализ, проводимый исходя из допущения, что все больные получили предписанное вмешательство; intention to treat analysis, ITT-analysis): анализ, в который включаются данные обо всех участниках, рандомизированно отнесенных к той или иной группе в начале исследования, независимо от того, какое вмешательство применялось в действительности.

Анализ данных в подгруппах (subgroup analysis): в ходе клинических испытаний или мета-анализов может проводиться анализ данных лишь об определенной подгруппе участников, у которых размер эффекта может отличаться от среднего значения выявленного эффекта.

Анализ данных с учетом сведений о выбывших больных (last observation carried forward, LOCF): метод включения в анализ данных последнего обследования или измерения, которое по времени не соответствует окончанию исследования.

Анализ чувствительности (sensitivity analysis): анализ, в ходе которого оценивается, насколько результаты мета-анализа чувствительны к различным ограничениям включенных в него данных. Примером таких ограничений может служить включение либо только крупных испытаний, либо только испытаний высокого методологического качества, либо только недавних исследований. Совпадение полученных данных можно считать убедительным доказательством эффективности изучаемого вмешательства и обобщаемости результатов мета-анализа.

Блоковая рандомизация (block randomisation): способ рандомизации, обеспечивающий требуемое число участников в каждой группе.

Величина p: степень вероятности того, что выявленное различие в эффективности разных вмешательств носит случайный характер. Если такая вероятность меньше 1 на 20 (т.е. $p<0,05$), результат обычно расценивается как статистически значимый.

Взаимодействие (interaction): изменение силы воздействия одного переменного фактора на клинический исход, которое происходит под влиянием другого переменного фактора. В ходе РКИ проведение теста на взаимодействие позволяет выявить различие в размере эффекта между подгруппами. См. также *Анализ данных в подгруппах.*

Взвешенная разность средних, ВРС (weighted mean difference, WMD): показатель, описывающий размер эффекта вмешательства в тех случаях, когда анализируются непрерывные (например, изменение роста или суммарного показателя по шкале выраженности симптомов), а не дихотомические данные (например, наступление смерти или развитие инфаркта миокарда). Средние различия параметров, на основании которых судят о клинических исходах в сравниваемых группах, оцениваются с учетом разного размера выборки и разного методологического качества отдельных исследований. ВРС — абсолютный показатель, поэтому он выражается в тех же единицах, что и исходный показатель клинического исхода.

Гетерогенность (неоднородность; heterogeneity): в применении к мета-анализу означает различия между исследованиями. Эти различия могут быть обусловлены использованием разных методов статистической обработки данных (статистическая гетерогенность) или различиями в основных характеристиках участников, применявшихся вмешательствах, клинических исходах или критериях их оценки (клиническая гетерогенность). Любой из этих видов гетерогенности может сделать невозможным обобщение данных в мета-анализе.

Гнездное (вложенное) исследование случай—контроль (nested case-control study): исследование, при котором вероятность включения в группу «контроля» зависит от времени, на протяжении которого изучаемый фактор воздействовал на данного человека. Гнездное исследование случай—контроль является разновидностью метода подбора контрольной группы к случаям, при этом контрольная группа каждого случая является независимой (от других обследуемых) выборкой из исходной популяции.

Гомогенность (однородность; homogeneity): понятие, противоположное гетерогенности.

Доверительный интервал, ДИ (confidence interval, CI): мера точности оценки показателя, отражающая диапазон, в котором могут находиться его реальные значения. Диапазон 95% ДИ будет включать в себя 95% результатов, полученных в исследованиях с аналогичной структурой, одинаковыми размерами выборки и характеристиками участников. Это положение близко по смыслу к оценке размера эффекта, но неравнозначно ей; нельзя утверждать, что истинный размер эффекта (который точно никогда не известен) с 95% вероятностью будет находиться в пределах 95% ДИ. Если при оценке относительного риска (ОР; см. ниже) или отношения шансов (ОШ; см. ниже) значения 95% ДИ больше 1, это означает отсутствие эффекта или даже вред вмешательства. С практической точки зрения, преимущество использования ДИ вместо величины p заключается в том, что этот показатель позволяет представить диапазон возможных эффектов вмешательства.

Индекс потенциального вреда, ИПВ (number needed to harm, NNH): один из показателей, позволяющих оценить недостатки изучаемого вмешательства. Он равен среднему числу больных в определенной группе, которые должны получить данное вмешательство, чтобы у одного дополнительного больного развился неблагоприятный исход. ИПВ=1/ПАР (см. ниже). Этот показатель обычно округляют в сторону уменьшения.

Исследование случай—контроль (case control study): исследование, структура которого предусматривает сравнение двух групп участников с развившимся и неразвившимся клиническим исходом (обычно неблагоприятным) с целью выявления различий во влиянии определенных факторов (обычно факторов риска) на развитие этого клинического исхода. Такая структура исследования наиболее пригодна при попытках определить причину возникновения редко встречающихся заболеваний, например некоторых разновидностей рака.

Исследование только случаев (case only study): исследование случаев без контрольной группы. Используется в генетической эпидемиологии и имеет несколько вариантов: перекрестную схему

исследования и схему исследования серии случаев с самостоятельным контролем.

Исходный риск (baseline risk): риск развития изучаемого клинического исхода в отсутствие лечения; обычно оценивается в контрольной группе.

Кластерная рандомизация (claster randomisation): способ рандомизации, предусматривающий назначение одинакового вмешательства в определенной группе участников. Примером кластерной рандомизации может служить отнесение в одну группу жителей целой деревни, пациентов определенной больницы или учащихся определенной школы. Если в дальнейшем будет проводиться анализ данных, полученных у каждого из участников в отдельности, а не у всей группы в целом, возможно появление систематической ошибки.

Клиническая значимость (clinical significance): важность результата с клинической точки зрения. Здесь слово «значимость» соответствует общепринятому пониманию слова «важность» (в отличие от статистической значимости, см. ниже); если в английском тексте термин «significance» приводится без дополнительного уточнения, то имеется в виду статистическая значимость.

Клинический исход (outcome): изучаемое в ходе исследования событие, которое можно оценить в виде дихотомических (например, наступление смерти, развитие инфаркта миокарда или снижение показателя выраженности боли на определенное количество баллов) или непрерывных (например, изменение роста или суммарного показателя по шкале выраженности симптомов) данных.

Когортное исследование (cohort study): неэкспериментальное исследование, структура которого позволяет проследить за группой (когортой) участников и выявить различия в частоте развития у них определенных клинических исходов. Включение в когорту участников с различным уровнем воздействия какого-то отдельного фактора риска (например, курение) может оказаться полезным при попытке оценить это воздействие в качестве возможной причины развития определенного клинического исхода (например, рак легких). Результаты проспективных когортных исследований (в ходе которых сначала формируется когорта, а затем в течение определенного времени осуществляется наблюдение за участниками) более достоверны, чем результаты ретроспективных когортных исследований.

Контрольная группа (control group, controls): одна из групп сравнения при проведении рандомизированного контролируемого испытания (РКИ; см. ниже); в этой группе участники получают плацебо или стандартное лечение либо не получают никакого лечения.

Косвенные критерии оценки (surrogate outcomes, intermediary outcomes, surrogate end points, surrogate outcomes measures; суррогатные исходы, промежуточные исходы): критерии оценки, которые не имеют прямого практического значения, но, как считается, отражают значимые исходы. Например, величина АД не имеет непосредственного значения для больного, но часто используется как мера исхода в клинических испытаниях, потому что повышенное АД — фактор риска развития инсульта и сердечных приступов. Косвенными исходами могут быть физиологические или биохимические маркеры, которые можно сравнительно быстро и легко измерить и которые используются для прогнозирования важных клинических исходов. Косвенные исходы часто используются в тех случаях, когда для выявления клинических исходов требуется длительное наблюдение.

Мета-анализ (meta-analysis): метод статистического анализа, в ходе которого объединяются результаты нескольких исследований, а итоговая оценка представлена в виде одного взвешенного показателя (при этом бóльший вес обычно присваивают более крупным исследованиям или исследованиям более высокого методологического качества).

Метод подстановок (multiple imputation): разновидность многофакторного анализа, при котором включаются ранее не учитывавшиеся («отброшенные») данные.

Модель постоянных эффектов (fixed effects model): при использовании этой модели в ходе мета-анализа исходят из допущения (зачастую необоснованного), что различия между исследованиями объясняются только случайными отклонениями от постоянного эффекта.

Модель случайных эффектов (random effects model): при использовании этой модели исходят из допущения о существовании различий в размере эффекта, выявляемого в разных исследованиях; эти различия учитываются как дополнительный источник вариабельности данных. В таком случае доверительные интервалы несколько более широкие, чем при использовании модели постоянных эффектов (см. выше). Предполагается также, что распределение размера эффекта носит случайный характер; центральную точку этого распределения определяют при расчете обобщенного размера эффекта.

Обоснованность (validity): термин, отражающий уровень доказательности полученных данных или методологическое качество исследования. О внутренней обоснованности (internal validity), или достоверности, данных можно говорить в тех случаях, когда исследование спланировано и проведено таким образом, что его результаты свободны от систематических ошибок и позволяют точно оценить изучавшийся эффект. Внешняя обоснованность (external validity), или обобщаемость, отражает степень применимости результатов данного исследования в обычной клинической практике.

Одномоментное исследование (cross sectional study): исследование, структура которого предусматривает однократную оценку воздействия отдельного фактора и/или наличия какого-то заболевания в определенной популяции. Такие исследования позволяют оценивать распространенность тех или иных заболеваний в популяции.

Относительный риск, ОР (relative risk, RR): отношение вероятностей развития определенного исхода в группах сравнения. При ОР>1 вероятность развития этого исхода в основной группе выше, чем в контрольной, а при ОР<1 — ниже. При низкой частоте развития исхода этот показатель близок по значению к отношению шансов (ОШ; см. ниже).

Отношение риска (hazard ratio, HR): показатель, который в широком смысле эквивалентен относительному риску (ОР; см. выше), но применяется в случаях, когда величина риска меняется со временем. Этот показатель включает в себя информацию, собираемую неоднократно за период наблюдения. Чаще всего используется при оценке изменений выживаемости. Отношение риска, равное 0,5, означает, что в группе вмешательства риск смерти в 2 раза ниже, чем в контрольной.

Отношение шансов, ОШ (odds ratio, OR): один из показателей эффективности изучаемого вмешательства; рассчитывается как ОШ развития определенного клинического исхода в основной группе к шансам его развития в контрольной группе. Чем ближе значение ОШ к 1, тем меньше различий в эффективности вмешательств, применявшихся в основной и контрольной группах. Следует отметить, что этот показатель может применяться при расчете шансов развития как неблагоприятных (смерть, развитие инвалидности), так и благоприятных (выживание) клинических исходов. При низкой частоте развития изучаемого клинического исхода ОШ близок по значению к относительному риску (ОР; см. выше), а при возрастании частоты развития исхода значения этих показателей начинают различаться.

Перекрестное рандомизированное испытание (crossover randomised trial): испытание, в ходе которого участникам первой группы рандомизированно назначают сначала одно вмешательство (с после-

дующей оценкой клинических исходов), а затем другое (также с последующей оценкой клинических исходов). Во второй группе эти вмешательства применяются в обратном порядке. Иногда до начала исследования, а также между его первой и второй фазами предусматриваются периоды «вымывания» (когда участники не получают никаких вмешательств), чтобы свести к минимуму вероятность лекарственных взаимодействий. Интерпретация результатов перекрестных РКИ может быть очень сложной.

Плацебо (placebo): биологически неактивное вещество, которое получают участники в контрольной группе. В идеале такой препарат полностью соответствует изучаемому лекарственному средству по внешнему виду, вкусовым или иным свойствам, но не влияет на течение или исходы заболевания. При испытаниях немедикаментозных вмешательств под плацебо понимают имитацию вмешательства (см. выше).

Повышение абсолютного риска, ПАР (absolute risk increase, ARI): абсолютная разница между риском развития определенного клинического исхода в основной и контрольной группах. Используется в случаях, когда риск развития этого исхода в основной группе выше, чем в контрольной; рассчитывается как разность между абсолютным риском (АР; см. выше) в основной группе и АР в контрольной группе. По величине ПАР нельзя судить об изменении риска в одной группе относительно другой; для этого следует использовать показатель относительного риска (ОР; см. выше).

Повышение относительного риска, ПОР (relative risk increase, RRI): относительное повышение риска развития определенного исхода в основной группе по сравнению с контрольной; ПОР=ОР – 1.

Поисковое исследование (exploratory analysis): при такой организации исследования нет проверяемой гипотезы, и соответственно его результаты не подвергают сложной статистической обработке, а лишь рассматривают каждый признак — нет ли различия, достигающего порога статистической значимости. Результаты такого исследования не являются доказательством различия больных и здоровых по этому признаку, а лишь указывают на возможные различия, несмотря на достижение «статистической значимости».

Применимость (applicability): возможность применения результатов клинического испытания у отдельных больных. Прямое применение данных о причинной зависимости, полученных в ходе рандомизированного испытания, возможно только при полном соблюдении всех его условий. В реальной практике результаты испытаний следует применять с учетом конкретных условий и индивидуальных характеристик больного, определяющих возможные клинические исходы.

Псевдорандомизированное испытание (quasi randomised trial): испытание, в котором способ отнесения участников к той или иной группе нельзя считать действительно рандомизированным; примером может служить распределение участников по группам в зависимости от даты рождения, дня недели, номера истории болезни, месяца, года или порядкового номера при включении в испытание.

Размер эффекта, или стандартизованная разность средних, СРС (effect size, standardized mean differences): в медицинской литературе термин «размер эффекта» используется при самых разных способах оценки эффектов лечения. Под этим термином обычно подразумевается стандартизованная разность средних — статистический показатель, который позволяет объединять непрерывные переменные (такие как выраженность боли или рост), оцениваемые с помощью разных шкал; рассчитывается как разность двух средних, деленная на величину стандартного отклонения внутри группы (см. ниже).

Рандомизированное контролируемое испытание, РКИ (randomised controlled trial, RCT): испытание, в котором участников в случайном порядке (рандомизированно) распределяют в две группы и более — по меньшей мере одну основную (где применяется изучаемое вмешательство) и контрольную (где применяется плацебо или другое вмешательство). Такая структура исследования позволяет сравнить преимущества и недостатки применяемых вмешательств.

Распространенность (prevalence): доля лиц с определенным признаком или заболеванием в определенной популяции в определенный момент времени.

Регрессионный анализ (regression analysis): процесс построения математической модели, наилучшим образом описывающей зависимую переменную в виде функции независимых переменных. В разных ситуациях используются различные регрессионные модели (чаще всего линейные, логистические и модели пропорционального риска).

Серии случаев (case series): исследования, в ходе которых анализируются данные о больных с каким-либо заболеванием (группа сравнения отсутствует).

Систематическая ошибка (bias): систематическое отклонение результатов от истинных значений, связанное с особенностями проведения исследования.

Систематическая ошибка, связанная с отбором участников исследования (selection bias): явное или скрытое несоответствие исходных характеристик участников исследования вследствие неправильного отбора и отнесения их к той или иной группе, приводящее к различиям в прогнозе между группами сравнения.

Систематическая ошибка, связанная с предпочтительной публикацией положительных результатов исследования (publication bias): вероятность опубликования материалов исследования зависит от полученных результатов. Известно, что при публикации предпочтение нередко отдается исследованиям с положительными и статистически значимыми результатами; в итоге при анализе опубликованных данных оценка эффективности изучаемого вмешательства может оказаться завышенной.

Систематический обзор (systematic review): обзор, в котором четко сформулирован изучаемый вопрос, подробно описаны методы поиска, отбора, оценки и обобщения результатов различных исследований, соответствующих изучаемому вопросу. Систематический обзор может включать в себя мета-анализ (см. выше), но его применение необязательно. Термин «систематический обзор», применяемый без особого уточнения, обычно подразумевает систематический обзор РКИ.

Сокрытие вмешательства, или применение слепого метода (blinding/blinded): о полном сокрытии вмешательства можно говорить лишь в том случае, когда никто из участников испытания (больные; лица, в той или иной мере обеспечивающие применение вмешательства; лица, проводящие обследование, регистрацию данных или их анализ) не знает, в какую группу был распределен тот или иной больной вплоть до получения окончательных результатов.

Сокрытие порядка отнесения больных к той или иной группе (allocation concealment): способ предотвращения систематической ошибки, связанной с отбором больных; при этом лица, осуществляющие формирование групп, не знают о том, в какой именно группе окажется тот или иной больной. При использовании такого сокрытия исследователи не могут повлиять (случайно или осознанно) на распределение больных.

Смертность (mortality): отношение числа умерших к общей численности определенной популяции за определенный период времени.

Снижение абсолютного риска, САР (absolute risk reduction, ARR): абсолютная разница между риском развития определенного клинического исхода в основной и контрольной группах. Показатель

используется в случаях, когда риск развития этого исхода в контрольной группе выше, чем в основной; рассчитывается вычитанием абсолютного риска (АР; см. выше) в основной группе из АР в контрольной группе. По величине САР нельзя судить об изменении риска в одной группе относительно другой; для этого следует использовать показатель относительного риска (ОР; см. выше).

Снижение относительного риска, СОР (relative risk reduction, RRR): относительное снижение риска развития определенного исхода в основной группе по сравнению с контрольной; СОР=1 – ОР.

Снижение шансов (odds reduction): показатель, дополняющий отношение шансов (ОШ; см. выше). Рассчитывается как 1 – ОШ; при низкой частоте развития клинического исхода этот показатель близок по значению к снижению относительного риска (СОР; см. выше).

Стандартизованная разность средних, СРС (standardised mean difference, SMD): применяется для оценки размера эффекта (см. выше) при анализе непрерывных (например, изменение роста, массы тела или суммарного показателя выраженности симптомов), а не дихотомических (например, наступление смерти, развитие инфаркта миокарда) данных. Если в разных исследованиях используются различные способы измерения (разные шкалы), в ходе систематического обзора нельзя провести прямое сравнение или объединение результатов этих исследований. В таких случаях проводится коррекция выявленных различий с учетом применявшейся шкалы оценки. СРС рассчитывается как отношение и поэтому не имеет единицы измерения.

Стандартизованное отношение заболеваемости, СОЗ (standardized incidence ratio, SIR): отношение наблюдаемого числа новых случаев заболевания к предполагаемому. Используется как стандартный показатель для сравнения частоты развития заболевания в изучаемой когорте и предполагаемой частоты его развития в популяции.

Статистическая значимость (statistical significance): одна из характеристик полученного результата; термин «различие статистически значимо» означает низкую вероятность того, что выявленное между группами различие носит случайный характер. Часто сообщается, что «статистическая значимость полученного результата меньше 5% ($p<0,05$)» — это означает, что случайно такое (или большее) различие может возникнуть лишь в 1 из 20 подобных случаев и реже. Если в английском тексте термин «significance» приводится без дополнительного уточнения, то имеется в виду статистическая значимость.

Статистическая мощность (power): одна из характеристик исследования. Достаточная статистическая мощность позволяет выявить клинически важные различия (например, в эффективности сравниваемых вмешательств), если таковые существуют в действительности. Статистическая мощность исследования повышается при использовании большей выборки или более точных способов оценки клинического исхода.

Статистически значимый результат (statistically significant): принято считать, что статистическая значимость этого результата меньше 5% (см. выше). О получении статистически значимого результата можно говорить и в тех случаях, когда диапазон 95% ДИ не включает значения, соответствующие отсутствию эффекта.

Статистически незначимый результат (not significant/non-significant): этот термин обычно означает более высокую вероятность того, что выявленное между группами различие носит случайный характер, т.е. аналогичное (или большее) различие может возникнуть случайно более чем в 1 из 20 подобных случаев (более чем в 5% случаев); истинное различие между группами может при этом отсутствовать. Однако статистически незначимый результат еще не означает отсутствия эффекта; в этих случаях следует говорить лишь о том, что в данном исследовании не были получены убедительные доказательства наличия эффекта вмешательства. Получение такого результата может быть обусловлено несколькими причинами: тем, что статистическая мощность (чувствительность) исследования оказалась недостаточной для выявления существующего эффекта, действительным отсутствием эффекта или существенным влиянием элемента случайности.

Факторный протокол (factorial design): применяется для сравнения двух и более вмешательств с контрольным в рамках одного испытания; при этом используются разные способы множественной рандомизации.

Число больных, которых необходимо лечить, ЧБНЛ (number needed to treat, NNT): один из показателей эффективности лечения. Это среднее число больных, которых необходимо лечить определенным методом в течение определенного времени, чтобы достичь определенного благоприятного исхода или предотвратить один неблагоприятный исход. ЧБНЛ=1/САР. Этот показатель обычно округляют в сторону увеличения. Следует отметить, что абсолютные показатели позволяют оценить усилия, необходимые для достижения положительного эффекта, однако на эти показатели существенное влияние оказывают как само лечение, так и величина исходного риска (см. выше) у отдельных больных. Когда в мета-анализ включены данные об участниках разных испытаний с разным исходным риском, то определить единое значение ЧБНЛ для всех участников невозможно; в этом случае (если исследования негетерогенны) более пригодны такие относительные показатели, как отношение шансов или относительный риск (ОШ или ОР; см. выше). Значение ЧБНЛ, которое используется при изложении результатов мета-анализа, обычно представляет собой комбинированный показатель, рассчитываемый на основе суммарного отношения шансов (ОШ) и средней величины исходного риска (для всех контрольных групп).

Шансы (odds): форма выражения вероятности развития определенного клинического исхода; рассчитывается как отношение числа больных с данным исходом к числу больных без данного исхода.