

Интерпретация результатов клинического исследования: стратегия и тактика построения доказательной базы

Г. П. Тихова

Карельский научный центр РАН, 185910, Петрозаводск

Interpretation of study findings: Strategy and tactics of evidence base development

G. P. Tikhova

FSBI "Karelian Research Center of Russian Academy of Science", 185910, Petrozavodsk

В статье представлены рекомендации и алгоритм анализа результатов клинического исследования, полученных в ходе статистической обработки данных, с целью подтверждения статистической достоверности эффекта и выявления возможных ошибок смещения и конфаундеров. *Ключевые слова:* клиническое исследование, статистическая достоверность гипотезы, ошибка смещения, вмешивающийся фактор.

Для цитирования: Регионарная анестезия и лечение острой боли. 2015; 9 (3): 62–69

The paper proposes recommendations and algorithm for analyzing clinical study results obtained in statistical processing of data in order to confirm statistical significance of the effect and detect possible bias and confounders. *Keywords:* clinical study, statistical significance, bias, confounder.

Citation: Regionarnaya anesteziya i lechenie ostroy boli. 2015; 9 (3): 62–69 (In Russ.)

Интерпретация полученных результатов статистической обработки данных – один из самых важных, волнующих и загадочных этапов клинического исследования. Здесь нет места рутине, которая обычно присутствует во время сбора или обработки данных, это время инсайтов и догадок, рождения нового знания или подтверждения идей, сложившихся из практического опыта и ожидавших объективной проверки. Это процесс сложный и творческий, требующий с одной стороны логической строгости, широкой эрудиции и свободного владения профессией, а с другой – достаточной смелости отойти от общепринятой стандартной позиции с тем, чтобы увидеть что-то новое и значимое. Для успешного проведения этого этапа исследования не существует рецептов, есть только советы новичкам: бывалый исследователь, как правило, уже имеет свои приемы и секреты, отточенные личным опытом. Возможно, в этом кроется одна из причин того, что в курсах по теории и практике клинических исследований интерпретации результатов уделяется значительно меньше внимания, чем разработке дизайна исследования или методам статистического анализа данных. Это тот случай, когда

доля внимания не адекватна степени важности и сложности процесса.

Мы поговорим сегодня именно о задачах и проблемах интерпретации результатов клинического исследования. Хотя выше было отмечено, что готовых рецептов для решения задач этого этапа нет, все же существует некая схема, почти алгоритм, следуя которому можно глубоко и всесторонне проанализировать результаты и получить сильные аргументы в поддержку своих идей и предположений. В этой статье мы рассмотрим исследования, цель которых обнаружить и/или подтвердить наличие некоторого эффекта или взаимосвязи между фактором воздействия и исходом. Что мы видим в результате любого такого исследования, независимо от его дизайна? Это некое число (коэффициент корреляции, относительный риск и т.п.), которое выражает силу или степень статистической ассоциативной связи между фактором воздействия (или риска в обсервационных исследованиях) и исходом (синдромом, выраженностью заболевания и т.п.), который является предметом нашего исследовательского интереса. Это полученное число пока не более чем формально рассчитанная метрика ассоциативной связи между интересующим нас фактором и исходом, которые, по нашим предположениям, должны быть как-то связаны. Однако что мы

Для корреспонденции:

Тихова Галина Петровна, e-mail: tikhovag@gmail.com

Correspondence to:

Galina Tikhova, e-mail: tikhovag@gmail.com

реально хотим узнать, проводя свое исследование? Сама метрика, число, нас мало интересует. На самом деле нам необходимо решить, является ли полученная ассоциативная связь причинно-следственной, т.е. существует ли объективно, в природе, такая связь между фактором воздействия и исходом. Если да, то любое изменение этого фактора должно вызывать изменение риска развития данного исхода. Общая цель анализа полученных результатов заключается в выявлении таких «настоящих» связей факторов с исследуемым событием (исходом) и отсеивании ассоциаций, получившихся случайно по причине вариабельности данных, неправильного дизайна или некорректного применения методов статистической обработки измерений.

Чтобы найти крупинки золота, надо тщательно промыть золотоносную породу через несколько фильтров. Примерно так же надо пропустить полученные результаты через три основных аналитических фильтра, чтобы получить достоверные зависимости и связи между изучаемыми факторами и событиями. Эти три фильтра составляет следующий каскад вопросов и тестов:

1. Является ли полученная ассоциация между фактором и событием валидной (достоверной) статистической ассоциативной связью или она обусловлена случайностью?

Для ответа на этот вопрос необходимо оценить вероятность

- случайной ошибки;
- ошибки смещения;
- вмешательства конфаундеров (сторонних или неучтенных воздействий).

2. Является ли полученная ассоциация объяснимой?

Для ответа на этот вопрос используется набор критериев (*positive criteria*), которые позволяют вынести вердикт относительно объяснимой возможности (правдоподобия) такой связи между фактором и исходом, оценивается:

- сила полученной ассоциации;
- согласованность с другими исследованиями;
- биологическое правдоподобие (убедительность);
- дозозависимый эффект воздействия (при исследовании препаратов или процедур, которые можно дозировать).

3. Можно ли расширить область приложения полученного результата за пределы целевой популяции (генерализуемость)? Это свойство называется иначе «внешняя валидность» (*external validity*) клинического исследования. Решение этого вопроса находится целиком в компетенции эксперта, проводящего анализ результатов исследования,

но мы обсудим ниже основные принципы этого анализа.

Вся эта трехступенчатая схема анализа полученных результатов в конечном итоге позволяет ответить на один важнейший вопрос: можем ли мы уверенно утверждать, что именно изученный нами фактор воздействия обуславливает изменение исследуемого исхода? Нет ли какого-то альтернативного объяснения нашим исследовательским находкам?

Итак, рассмотрим каждый из трех основных вопросов подробнее.

1. Валидность статистической ассоциативной связи

Случайная ошибка

В первую очередь необходимо обратить внимание на объем выборки. Чем он меньше, тем больше вероятность быть введенным в заблуждение просто в силу вероятностной природы измерений. Этот вопрос обсуждался в наших статьях неоднократно [1]. Поскольку в исследованиях мы можем лишь оценить силу связи, а не получить точное ее значение (ведь большая часть популяции остается за пределами наших измерений), то как любая оценка, сделанная в условиях неполной информации (некоторой степени неопределенности), наш результат будет иметь случайную ошибку. Случайной эта ошибка называется не потому, что мы ее не ожидаем [2], как раз наоборот, мы о ней отлично осведомлены. Она называется случайной, потому что обусловлена стохастической (вероятностной) природой исследуемых нами данных. Такая ошибка зависит от вариабельности показателей, связь между которыми исследуется, а также от объема выборки [1]. Все очень просто: чем больше вариабельность, тем больший объем выборки требуется для достижения приемлемой величины случайной ошибки. Но это все в теории, а что делать, когда мы уже имеем рассчитанную по нашим данным величину эффекта с его конкретной случайной ошибкой? Нам надо сделать три последовательных шага:

- Оценить величину эффекта.

Если вы получили маленькое значение коэффициента корреляции (или любого другого статистического параметра, отражающего силу связи) – это повод задуматься, насколько реальна и существенна эта связь, даже если она отвечает всем формальным требованиям объективности.

- Проверить гипотезу о его статистической значимости (достоверном отличии от нулевого эффекта).

В результате этой проверки вы получите значение ошибки p , которое позволит вам оценить вероятность ложного принятия решения о наличии

связи (или эффекта фактора воздействия), тогда как в реальности такой связи нет, т. е. фактор никак не влияет на исход. Обычно пороговым уровнем для принятия такого решения является $p = 0,05$ (уровень значимости), но это просто дань традиции, каждый исследователь может устанавливать свой уровень значимости, более жесткий или более мягкий.

– Определить точность полученной оценки исследуемого эффекта.

Точность оценки определяет интервал, который с заданной вероятностью включает реальное значение исследуемого эффекта, это не что иное как доверительный интервал, чаще всего 95%, но, как и уровень значимости, его вероятность можно варьировать в соответствии с условиями конкретного исследования.

Внимание к этим трем позициям при анализе полученных коэффициентов связи позволит исключить сомнительные эффекты и оставить только те, в защиту которых вы сможете привести весь арсенал статистических аргументов.

Каждое исследование в обязательном порядке имеет так называемую первичную или основную цель (primary goal). Это главный вопрос, на который исследование должно дать ответ. Под него выстраивается дизайн, определяются конечные точки и подбираются подходящие методы статистического анализа данных. Однако было бы очень расточительно потратить массу времени и сил, чтобы получить одну четырехпольную табличку и этим ограничиться. Как правило, в исследование включаются вторичные цели (secondary goals), которые рассматриваются, исходя из возможностей заданного дизайна. Обычно это более глубокое изучение связи (эффекта) между фактором и исходом внутри подгрупп, созданных при стратификации всей выборки по каким-то признакам, например, разбиение ее по полу или возрасту пациентов, степени выраженности сопутствующей патологии и т. п. Исследование изменения эффекта в зависимости от пола, возраста или других признаков называется анализом подгрупп [3] или анализом модификации эффекта (связи). Действительно, достаточно часто найденное на общей выборке значение эффекта трансформируется при его вычислении внутри подгрупп той же выборки, сформированных по половому, возрастному, географическому или другим признакам. Однако, проводя такой анализ, необходимо иметь в виду одну тонкость. Если изучение модификации эффекта в подгруппах было запланировано еще на стадии разработки дизайна, причем не просто принято решение о таком анализе, а сформулированы конкретные гипотезы о том, при каких признаках эффект, возможно,

будет изменяться, то такая ситуация называется проверкой гипотез. Если идея проанализировать модификацию эффекта в подгруппах возникла после того, как данные были собраны и обработаны в общей выборке, вы попадаете в ситуацию формулирования гипотез. Дело не только в различии названий, сами ситуации кардинальным образом разнятся и заставляют по-разному воспринимать результаты анализа подгрупп. Дело в том, что если вы до получения данных имеете предположение, что связь между фактором и исходом может быть разной, например у женщин и мужчин, у молодых и пожилых пациентов, вы готовите ваши данные, имея в виду проверку этих гипотез. Как правило, таких предварительных соображений не так много, даже если категориальных переменных в вашем исследовании большое количество. Таким образом, число проверок дополнительных гипотез будет невелико и достигнутая статистическая достоверность на традиционном уровне значимости, равном 0,05, будет объективным аргументом в пользу вашего предположения, высказанного до начала сбора данных. Однако обычной является другая ситуация, когда происходит своего рода сканирование модификации изучаемой связи по всем возможным подгруппам пациентов с целью обнаружить статистически достоверное различие в каком-нибудь варианте стратификации, причем без формулирования каких бы то ни было вразумительных предположений о поведении эффекта и к тому же после того как данные уже получены. Такой процесс в англоязычной литературе называется fishing expedition – поход на рыбалку. На самом деле термин подобран очень метко, он ясно отражает возможности и доверие к результатам подобного исследования. Подобно рыбалке, при сканировании эффекта по всем возможным подгруппам, мы не знаем заранее, что ожидаем, и можем получить как крупную рыбу, так и пустую консервную банку, но проблема в том, что в отличие от рыбалки, мы, к сожалению, не можем достоверно идентифицировать, что же мы выловили на самом деле, даже если получим статистическую достоверность различия. Почему? Все очень просто, достаточно вспомнить, что означает наш порог (уровень значимости), равный 0,05 [4]. Напомним: он означает, что из 100 сравнений, которые показали достоверность различия между подгруппами, 5 будут ошибочными. Теория не позволяет определить, какие именно 5 полученных достоверных различий вводят нас в заблуждение. Условие принятия гипотезы о межгрупповом различии, установленное на уровне 0,05, будет ограждать нас от неправильных решений, только если мы будем проверять заранее сформулированные предположения и их

количество будет невелико. Обычно таких заранее сформулированных гипотез немного, гораздо меньше сотни и даже едва ли наберется с десятков. Однако при сканировании эффекта по всем возможным переменным и признакам, включенным в исследование, количество таких сравнений становится запредельным, поэтому у нас есть большая вероятность получить смесь из действительно существующих различий и тех, которые получились случайно, так сложилась наша выборка, и в другом исследовании, на другой выборке, эти различия не будут статистически достоверны, а возможно, и вовсе будут отсутствовать. Это не значит, что такое сканирование эффекта по всем подгруппам бесполезно. Как раз наоборот, оно имеет большое значение, особенно в пилотных обсервационных исследованиях, т. к. позволяет сделать новые предположения, *сформулировать гипотезы*, которые затем можно *проверить в эксперименте или наблюдении*, сконструированном специально под эти цели. Речь идет лишь о том, что к таким результатам надо относиться с большой осторожностью, даже если вы получили статистическую достоверность, поскольку в ситуации «похода на рыбалку» статистические методы не гарантируют тот уровень значимости для ваших сравнений, который вы установили. Если коротко резюмировать все вышесказанное об анализе эффекта в подгруппах, интерпретация результатов проверки гипотез в ситуациях

«тестирования гипотез» и «формулирования гипотез» должна быть совершенно различной.

Итак, первый этап интерпретации результатов завершен, вы убедились, что полученный коэффициент связи статистически достоверен, т. е. объяснить наличие эффекта случайностью нельзя. Однако утверждать, что связь действительно существует, еще рано. Необходимо провести анализ возможных ошибок смещения и исключить наличие неучтенных конфаундеров, сторонних вмешивающихся факторов.

2. Ошибка смещения

Ошибка смещения систематически занижает или завышает значение коэффициента связи (эффекта), в некоторых ситуациях она может породить связь там, где ее нет, или наоборот, скрывать ее существование. Смещение оценки эффекта относительно его реального значения дает любой источник систематической ошибки при проведении клинического исследования (рис. 1). Ошибка смещения коварна и достаточно часто фатальна. В отличие от случайной ошибки, которая всегда присутствует при статистической оценке какого-либо параметра или коэффициента связи, и для ее расчета и снижения имеются специально разработанные методы, ошибка смещения искажает действительность, вводит в заблуждение, и если она допущена, то не существует методов не только ее исправить, но даже хотя бы оценить.

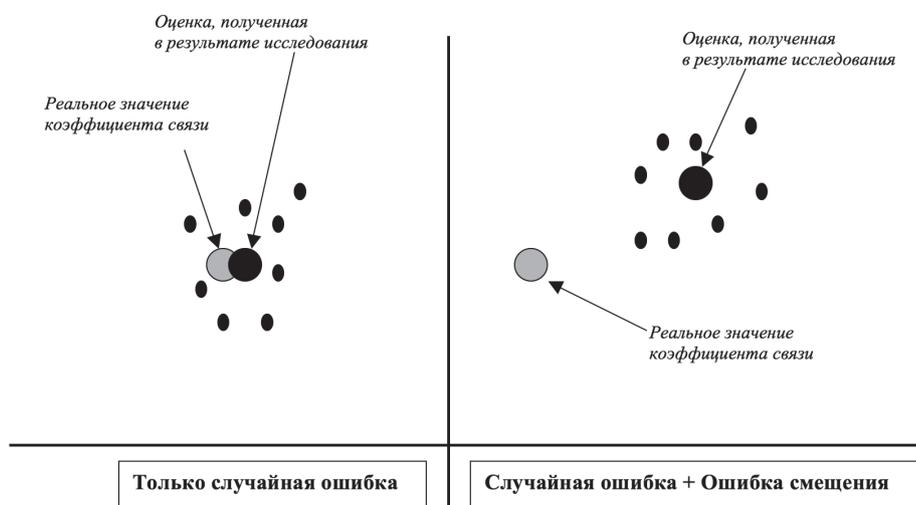


Рис. 1. Последствия случайной ошибки и ошибки смещения, допущенной при оценке реального значения эффекта. Серым кругом обозначено реальное значение эффекта, черными маленькими эллипсами – набор данных, черным кругом – оценка реального значения эффекта, рассчитанная из полученных данных. Случайная ошибка присутствует в любом исследовании, т. к. оно имеет дело с переменными вероятностной природы. При отсутствии ошибки смещения все измерения ложатся равномерно со всех сторон от реального значения, поэтому оценка будет тем ближе к нему, чем больше выборка. Величину случайной ошибки всегда можно оценить и достоверно рассчитать доверительным интервалом, который с заданной вероятностью будет включать реальное значение эффекта. Если что-то в исследовании сделано неправильно, то помимо случайной ошибки возникает ошибка смещения, системный сдвиг всех измерений относительно реального значения в одну сторону. Оценка, рассчитанная при наличии ошибки смещения, будет занижать или завышать значение эффекта, а доверительный интервал не будет покрывать его реальное значение

Случайная ошибка никогда не нарушает реальной картины и максимум что может сделать, это размыть ее так, что детали будут плохо видны, ошибка смещения подобна подрывной деятельности диверсанта: неизвестно, где, когда и какой урон она нанесет, но последствия могут быть очень серьезными и что самое неприятное, исправить или хотя бы снизить значение этой ошибки, как правило, невозможно. Все это должно настраивать на то, чтобы внимательно инспектировать свое и другие исследования на возможность наличия такой ошибки. На самом деле, ошибкам смещения и конфаундерам надо посвящать отдельное более детальное сообщение, но мы отметим здесь основные позиции, на которые надо всегда обращать внимание, чтобы минимизировать вероятность этой серьезной ошибки.

Источником ошибки смещения могут стать просчеты:

- на этапе разработки дизайна и включения пациентов в исследование (selection bias);
- на этапе наблюдения, регистрации и сбора данных (observation bias);
- на этапе первичной обработки данных (performance bias);
- и даже на стадии интерпретации результатов в терминах области исследования. Если исследователь является горячим приверженцем одной из проверяемых гипотез, он будет невольно стараться выделить из полученных результатов максимум выводов в пользу своего предпочтения [4].

Источники ошибок смещения при включении пациентов в исследование

Чтобы избежать или хотя бы свести к минимуму вероятность ошибок смещения на этапе создания выборки пациентов, необходимо всегда иметь в виду одну важнейшую идею: выборка должна быть репрезентативной, т. е. представлять свойства исследуемой популяции в полном объеме и не искажать их баланса, характерного для этой популяции. Если в популяции количество мужчин и женщин соотносится как 1:1, то и в выборке этот паритет должен приблизительно сохраняться. Если молодых людей в данной популяции в 2 раза больше, чем пожилых, то в выборке должно быть примерно в 2 раза больше молодых пациентов. Этот простой вопрос на самом деле очень сложен и критичен. Для каждого исследования он должен быть максимально тщательно проработан желательно вместе со специалистом по обработке данных, потому что ошибки, допущенные на этом этапе, исправить нельзя никакими последующими ухищрениями.

Рандомизация является наилучшим способом избежать систематического сдвига данных относительно реального значения при наборе выборки для исследования, хотя и она не гарантирует от ошибок

смещения на все 100%. Однако рандомизация возможна только в клинических испытаниях, где исследователь имеет активную роль, назначая тот или иной метод лечения пациенту и регистрируя у него результат этого лечения – исход. В наблюдательных исследованиях, где мы имеем лишь пассивную роль наблюдателя за влиянием фактора на пациентов, рандомизация невозможна. Для таких клинических исследований разработаны различные приемы формирования выборки, позволяющие снизить риск систематической ошибки. Например, если нет возможности без очевидных искажений сохранить свойства популяции при включении пациентов в выборку для исследования, то тогда стараются набрать две сравниваемые выборки таким образом, чтобы они были статистически однородны, т.е. чтобы между ними не было статистически значимого различия по важнейшим характеристикам пациентов, способным повлиять на результат исследуемого фактора воздействия. Важным правилом при формировании выборки для наблюдательных исследований является устранение по максимуму участия любого члена исследовательской группы в принятии решения о включении в выборку очередного пациента. Это не всегда реализуемо в полной мере, но стремиться к этому надо. Дело в том, что доктор, заинтересованный в продвижении какой-то идеи, невольно будет стремиться набирать «интересных» пациентов, у которых, например, проведено больше исследований или более выражена патология, на которую воздействует исследуемый метод лечения. «Скучные» пациенты будут отсеяны как не представляющие исследовательского интереса. Этот феномен далек от подтасовки фактов, он происходит помимо воли исследователя, как бы он не пытался себя контролировать, так устроен человек. Этот факт изучен и отмечен во многих работах по методологии научных исследований, поэтому предпочтительнее всего привлечь для набора пациентов в исследование доктора, не имеющего к нему никакого отношения.

Источники ошибок смещения на этапе наблюдения, регистрации и сбора первичной обработки данных

Основным источником ошибок смещения на этапе сбора данных является нарушение протокола исследования. Строгое соблюдение протокола обеспечивает одинаковое выполнение методики воздействия, времени наблюдения, полноты регистрации результатов и измерений. Если сказать коротко: малейшее отклонение от протокола приводит к тому, что пациенты в разных группах наблюдаются по-разному, что систематически смещает их данные. Слово «по-разному» является ключевым при инспекции результатов исследования на наличие ошибки смещения, так как именно

различные условия исполнения протокола для одной и второй группы приводят к системному сдвигу в данных. Второй ловушкой на данном этапе может стать ангажированность наблюдателя или пациента, особенно если конечными точками являются результаты опросников или балльные оценки различных шкал. Доказано, что больные с более выраженной патологией внимательнее относятся к фиксации у себя разных ощущений, стремятся более подробно описать врачу свое состояние, чем те, у кого то же заболевание имеет более легкую форму. Аналогично – и доктор может под влиянием собственных профессиональных убеждений невольно исказить получаемую от пациента информацию. Для исключения подобных системных ошибок используется прием маскирования от наблюдателя информации о том, кому из пациентов какой метод лечения (или фактор воздействия/риска) применялся. В англоязычной литературе такие исследования характеризуются как слепые (blind). Маскированию может подлежать не только наблюдатель, регистрирующий данные пациентов, но и сами пациенты (конечно, не во всех случаях это реализуемо), а также специалист по статистической обработке данных

и даже тот, кто интерпретирует результаты исследования. В зависимости от количества маскированных этапов, исследования могут быть просто слепыми, а также двойными и тройными слепыми. На самом деле, чем больше разделения разных этапов между специалистами, тем меньше выводы исследования подвержены опасности системного сдвига, но это увеличивает стоимость и продолжительность научной работы, кроме того, не каждый этап можно замаскировать. При решении этого вопроса необходимо руководствоваться здравым смыслом, чтобы не перегнуть палку.

3. Конфаундеры (сторонние вмешивающиеся факторы)

Любую полученную статистически достоверную связь между фактором воздействия и исходом необходимо внимательно и критически проанализировать с тем, чтобы исключить вмешательство неучтенного (а порой и неизвестного) стороннего (третьего) игрока – конфаундера. На рисунке 2 схематично показано, каким образом конфаундер скрыто взаимодействует с обеими исследуемыми признаками и «загрязняет» их связь своим вмешательством.

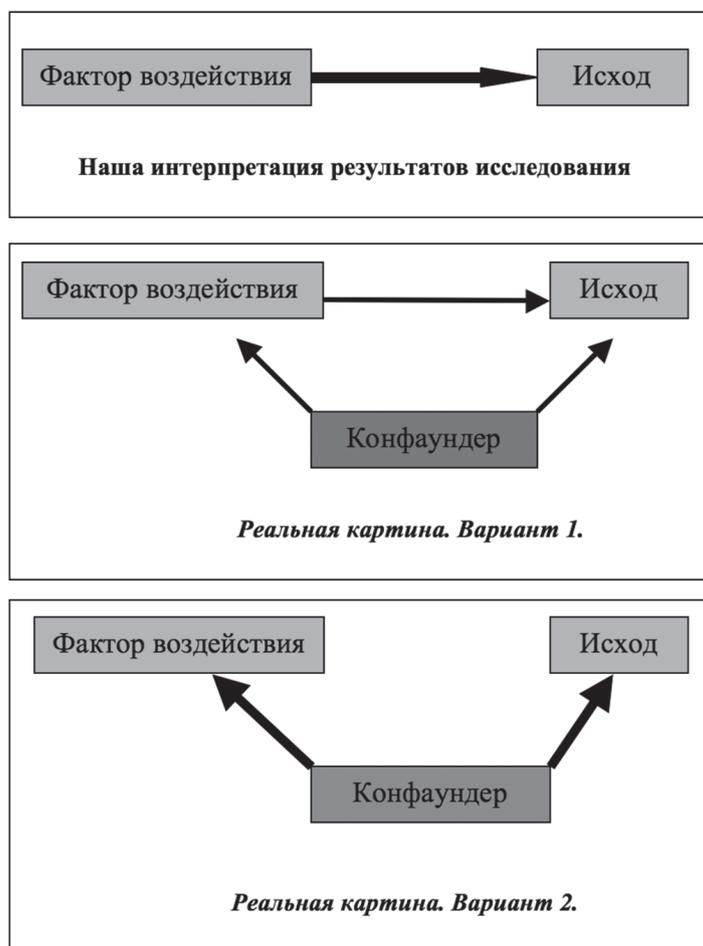


Рис. 2. Два варианта вмешательства стороннего неучтенного процесса (конфаундера) в связь между фактором воздействия и исходом

Влияние конфаундера на обе переменные может быть таким значительным, что он может частично или полностью отвечать за ассоциацию, которую мы наблюдаем между фактором и исходом в нашем исследовании. Он не менее коварен, чем ошибка смещения, поскольку также приводит к ложным заключениям о связи фактора и исхода, искажая реальную картину.

Влияние вмешивающихся факторов можно снизить до приемлемого уровня или совсем исключить – на этапе дизайна и сбора данных методом ограничения выборок или подбора пар соответствия (эти методы требуют отдельного детального изложения);

– на этапе анализа данных методом стратификации выборки по признаку, существенно влияющему на исход, или включением в план статистической обработки полученных данных многофакторного анализа.

Надо отметить, что стратификация при обработке данных хоть и позволяет максимально убрать влияние конфаундера, однако может породить другую проблему – недостаточный объем выборки в одной из подгрупп. Если признак распределен в выборке неравномерно, например мужчин значительно больше, чем женщин, то при стратификации по полу, может оказаться, что подгруппа женщин слишком мала, чтобы делать статистически достоверные заключения. В том случае, когда исследователь заранее предполагает, что пол (или любой другой признак) может повлиять на результаты исследования, лучше этим озаботиться на этапе дизайна и включения пациентов в исследование. Необходимо спланировать набор пациентов таким образом, чтобы по данному признаку выборка была сбалансированной. Для этого применяются перечисленные методы ограничения (restriction) и подбора пар соответствия (matching). При проведении клинического испытания самым надежным методом уберечься от конфаундеров является все та же рандомизация. Именно потому, что она сохраняет свойства популяции в полученной выборке (при достаточном ее объеме, разумеется!), она же сохраняет и влияние всех игроков исследуемого процесса в неискаженном балансе. Дополнительный бонус заключается в том, что это касается как известных нам сторонних вмешательств, так и тех, о которых мы даже не подозреваем. Однако доля клинических испытаний (экспериментов с активной ролью исследователя) у нас пока, к сожалению, не велика, поэтому рассмотренные выше приемы анализа полученных результатов должны стать рутинной процедурой для каждого исследователя.

Наличие объективной связи между фактором и исходом – это всего лишь наше суждение (вывод), а не факт, и как всякое суждение оно должно

быть подкреплено сторонними доказательствами и аргументами, выходящими за рамки нашего исследования. Не существует исчерпывающих гарантированных тестов для «оценки» или «проверки» наличия в природе причинно-следственной связи между нашим фактором и исходом. Эту роскошь мы имеем только на этапе проверки статистической ассоциации между исследуемыми признаками. Заключение о существовании в реальности связи между фактором и исходом должно базироваться на всей доступной информации, полученной не только в ходе исследования, но и из других научных и практических источников. В Школе Медицины Гарвардского университета для этой цели был разработан набор позитивных критериев (positive criteria), который позволяет систематизировать все нестатистические аргументы в пользу объективного существования связи (эффекта) между фактором воздействия и исходом:

– Сила связи (числовое значение полученного статистического коэффициента).

Сильная связь минимизирует возможность вмешательства в нее неучтенных и неизвестных сторонних факторов (конфаундеров).

– Полнота доказательства и согласие с другими исследованиями. Учет всех возможных нюансов в проведении исследования и интерпретации его результатов, а также согласованность его выводов с другими исследованиями, проведенными в отличных условиях, на различных популяциях и т. д., позволяет утверждать, что выводы исследования отражают реальную картину.

– Биологическая убедительность (правдоподобие). Если есть возможность объяснить находки исследования разумным биологическим механизмом, это укрепит доказательную базу ваших выводов.

– Наличие дозозависимого эффекта. Если уровень выраженности исхода изменяется при изменении дозы воздействия, то это также свидетельствует в пользу объективного наличия связи между ними.

Последним этапом интерпретации результатов исследования является определение пределов в популяции, до которых мы можем распространить наши заключения [5]. Строго говоря, наши выводы доказательно верны только для той популяции, которую мы обозначили как предмет исследования и описали ее в критериях включения и исключения в разделе «Материалы и методы». Однако профессиональное сообщество не может позволить себе проводить десятки одинаковых исследований на разных популяциях в разных условиях, чтобы получить гарантированные выводы. Определение внешней валидности, т.е. на каких пациентов до каких допустимых пределов можно распространить выводы и рекомендации исследования, находится

полностью в компетенции автора работы. Только специалист и профессионал может успешно провести такую экспертизу, опираясь на свои знания и практический опыт.

Все описанные этапы и рекомендации актуальны не только для проведения собственного исследования, они также полезны для критической оценки и всестороннего анализа новой информации, получаемой при чтении научных статей и монографий.

Литература

1. Тихова Г. П. Методология планирования клинического исследования. Вопрос N1: Как определить необходимый объем выборки? *Регионарная анестезия и лечение острой боли*. 2014; 8(3): 57–63.
2. Тихова Г. П. Значение и интерпретация ошибки среднего в клиническом исследовании и эксперименте. *Регионарная анестезия и лечение острой боли*. 2013; 7(3): 50–53.
3. Lagakos S. W. The Challenge of Subgroup Analyses — Reporting without Distorting *N Engl J Med*. 2006; 354:1667–1669
4. Тихова Г. П. Расчет и интерпретация относительного риска и других статистических параметров, полученных из четырехпольной таблицы частот. *Регионарная анестезия и лечение острой боли*. 2012; 6(3): 69–75

5. Rothwell P. M. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *Lancet*. 2005; 365(9453): 82–93.
6. Pannucci C. J., Wilkins E. G. Identifying and Avoiding Bias in Research. *Plast Reconstr Surg*. 2010; 126(2): 619–625.

References

1. Tikhova G. P. Planning clinical research. Question #1: How to calculate enough sample volume? *Regionarnaya anesteziya i lechenie ostroy boli*. 2014; 8(3): 57–63.
2. Tikhova G. P. Importance and uninterpretation of standard error of mean in clinical stude and trial. *Regionarnaya anesteziya i lechenie ostroy boli*. 2013; 7(3): 50–53. (In Russian).
3. Lagakos S. W. The Challenge of Subgroup Analyses – Reporting without Distorting *N Engl J Med*. 2006; 354:1667–1669.
4. Tikhova G. P. Fourfold frequency table – Occam’s Razor in the world of statistics. Part 1. Calculating relative risk and other parameters from fourfold frequency table. *Regionarnaya anesteziya i lechenie ostroy boli*. 2012; 6(3): 69–75. (In Russian).
5. Rothwell P. M. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *Lancet*. 2005; 365(9453): 82–93.
6. Pannucci C. J., Wilkins E. G. Identifying and Avoiding Bias in Research. *Plast Reconstr Surg*. 2010; 126(2): 619–625.