

ДОКАЗАТЕЛЬНОСТЬ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ РАЗЛИЧНЫХ ДИЗАЙНОВ


Е. О. Борисова , Ю. Н. Еремина, О. В. Гульбекова

Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва, Россия

В данной статье различные дизайны клинических исследований, которые используются для получения новых научных знаний в области клинической медицины, рассматриваются с позиций их доказательной ценности в изучении причинно-следственных взаимоотношений между воздействующим фактором и результатом его потенциального влияния на здоровье людей. Обсуждаются основные различия между наблюдательными и экспериментальными исследованиями, а также их ограничения, обусловленные особенностями дизайнов клинических исследований. В заключение делается вывод о том, что к оценке достоверности результатов клинических исследований надо подходить с пониманием тех ограничений, которые характерны для различных дизайнов. Точность клинических исследований зависит от влияния многих факторов, которые способны приводить к искажению получаемых результатов по сравнению с их истинными значениями. При этом отмечается, что наблюдательные исследования подвержены систематическим и случайным ошибкам в большей степени, чем экспериментальные. Это объясняется тем, что в силу особенностей дизайна наблюдательные исследования не могут контролировать ошибки, связанные с возможной несопоставимостью групп сравнения. Они способны выявлять наличие статистической связи между явлениями, но доказать, что связь носит причинно-следственный характер, могут только рандомизированные клинические исследования. Точность рандомизированного клинического исследования может быть повышена с помощью систематических обзоров с метаанализом.

Ключевые слова: дизайны клинических исследований, наблюдательные контролируемые исследования, экспериментальные клинические исследования, систематический обзор, метаанализ, систематические ошибки, доказательность выводов, уровни достоверности доказательств, уровни убедительности рекомендаций

Вклад авторов: Е. О. Борисова — анализ научной литературы, написание текста, подготовка рукописи для публикации; Ю. Н. Еремина — анализ научной литературы, подготовка рукописи для публикации; О. В. Гульбекова — редактирование текста.

 **Для корреспонденции:** Елена Олеговна Борисова
ул. Островитянова, д. 1, г. Москва, 117997, Россия; eoborissova@mail.ru

Статья поступила: 23.07.2022 **Статья принята к печати:** 21.08.2022 **Опубликована онлайн:** 30.09.2022

DOI: 10.24075/medet.2022.054

LEVELS OF EVIDENCE AND STUDY DESIGNS


Borisova EO , Eremina JuN, Gulbekova OV

Pirogov Russian National Research Medical University, Moscow, Russia

In this article, various designs of clinical trials used to obtain new scientific knowledge in the field of clinical medicine are considered from the position of their evidential value in studying the cause-and-effect relationship between the influencing factor and result of its potential effect on human health. Basic differences between observational and experimental trials, their limitations due to peculiarities of design of clinical trials are being discussed. A conclusion is made that validity of results of clinical trials should be assessed taking into account the limitations that are typical of various designs. Accuracy of clinical trials depends on many factors that can distort the obtained results as compared with true values. It is noted that observational trials are subject to systematic and accidental errors to a greater extent than experimental ones. It occurs because design characteristics do not allow observational trials to control the mistakes associated with possible incompatibility of comparison groups. They can detect a statistical relation between the phenomena, but only randomized clinical trials can prove that there is a causal relationship. Accuracy of a randomized clinical trial can be increased using systematic reviews and meta-analysis.

Keywords: designs of clinical trials, observational controlled trials, experimental clinical trials, systematic review, meta-analysis, systematic errors, substantiation of conclusions, evidence reliability levels, recommendation strength levels

Author conclusions: Borisova EO — analysis of scientific literature, writing a text, preparation of a manuscript for publication; Eremina JuN — analysis of scientific literature, preparation of a manuscript for publication; Gulbekova OV — text editing.

 **Correspondence should be addressed:** Elena O. Borisova
ul. Ostrovityanova, 1, Moscow, 117997, Russia; eoborissova@mail.ru

Received: 23.07.2022 **Accepted:** 21.08.2022 **Published online:** 30.09.2022

DOI: 10.24075/medet.2022.054

Получение новых научных знаний в области современной клинической медицины во многом базируется на результатах клинических эпидемиологических исследований. Именно они позволяют выявлять факторы, способствующие возникновению и прогрессированию заболеваний, оценивать количественный вклад этих факторов в развитие и дальнейшее течение заболеваний, стратифицировать популяцию по степени риска и определять прогноз, мониторировать уровень факторов риска и оценивать эффективность профилактических программ, планировать клинические исследования

(КИ), формулировать и проверять гипотезы. Во многом благодаря эпидемиологическим исследованиям было показано значение дислипидемии, артериальной гипертензии, курения и сахарного диабета в развитии среди населения атеросклероза и связанных с ним заболеваний, были выполнены КИ и разработаны рекомендации по лечению и профилактике этих заболеваний, как на популяционном, так и индивидуальном уровнях [1].

В клинической эпидемиологии используются несколько типов КИ, различающихся по своей структуре и направленных на поиск ответов на

определенные клинические вопросы, касающиеся оценки распространенности патологических состояний, поиска и изучения причин или факторов риска развития заболеваний, оценки частоты, относительного риска и прогноза заболеваемости. Важными клиническими вопросами являются оценка эффективности профилактических, диагностических и лечебных медицинских вмешательств.

Каждая из этих задач может быть решена с помощью КИ, с определенной логической структурой, включающей методы отбора людей в исследование, формирования групп сравнения, сбора информации, способы ее анализа и интерпретации. Форма проведения исследования, созданная для поиска ответов на поставленные клинические вопросы, называется дизайном исследования. В дизайне исследования заложена та степень точности, с которой полученный в данном исследовании результат может отражать реальные связи между событиями.

В настоящей статье внимание уделяется тем факторам, ограничивающим степень достоверности исследования, которые связаны с особенностями его дизайна, рассматривается структура и степень достоверности различных дизайнов в сравнительном плане, а также описываются рекомендации по определению уровней достоверности доказательств и уровней убедительности рекомендаций.

ТОЧНОСТЬ ИССЛЕДОВАНИЯ

Достоверность исследования определяется его точностью, которая складывается из степени, с которой результаты данного исследования можно применить к другим группам (внешняя валидность или обобщаемость), степени, в которой исследование может исключить альтернативное объяснение полученных результатов (внутренняя валидность), и степени точности оценки измеряемых параметров (надежность) [2].

Внешняя валидность выборки определяется степенью ее репрезентативности по отношению к популяции, из которой она была выделена [3]. Научные КИ проводятся не на всей популяции людей, которые страдают изучаемой патологией или на которых действует предполагаемый фактор риска, а на части этой популяции (выборке). Если участники выборки по своим характеристикам полностью соответствуют характеристикам популяции, то есть репрезентативны популяции, полученные результаты можно распространить (обобщить, генерализировать) на всех людей этой популяции. Однако выборка может быть репрезентативна только если она формировалась методом случайного отбора. Метод случайного отбора подразумевает выявление всех больных, страдающих данной патологией, а затем случайное, т. е. равновероятное, включение в выборку представителей всех типов больных из общей популяции. В медицинских исследованиях это практически невозможно. Таким образом, больные в КИ могут отличаться от всех больных, страдающих изучаемым заболеванием, не только по возрасту, полу и национальности, но и по социальному статусу и материальному достатку, по отношению к своему здоровью, по месту проживания, тяжести состояния и многим другим характеристикам, что делает выборку не случайной и не вполне репрезентативной. В таких случаях говорят о недостаточной внешней валидности выборки.

Выводы, сделанные в исследованиях на неслучайных выборках, могут быть перенесены на популяцию в целом с определенной долей ошибки (смещением). Такая ошибка

возникает при формировании выборки и называется систематической ошибкой выборки.

Под **систематической ошибкой** в статистике понимают непреднамеренное, но регулярное, неслучайное, однонаправленное отклонение рассчитанных показателей от их действительных значений [4].

Чем меньше репрезентативность выборки, тем менее точным является исследование, тем больше шансов того, что на результаты такого исследования действовали и другие факторы (ошибки), искажающие выводы. Репрезентативность выборки можно повысить увеличением ее численности, поэтому доверие к исследованиям с большим числом участников, как правило, выше.

Внутренняя валидность определяется тем, насколько хорошо дизайн исследования может исключить альтернативные объяснения своих выводов. Различия в результатах сравниваемых групп могут быть не только следствием изучаемого фактора. Существуют и другие, альтернативные объяснения. Нельзя исключить влияния на результат других факторов, которых исследователь не планировал изучать, не учел или о которых он просто не знал, но которые также способны воздействовать на исход. Влияние этих факторов в случае, если они распределены неравномерно в группах сравнения и контроля, будет смещать истинные результаты вмешательства и приводить исследователя к неточным, ошибочным выводам. Такие факторы вызывают однонаправленное смещение (искажение) результатов исследования и называются систематическими ошибками отбора. К ошибкам отбора относятся все факторы, которые приводят к несопоставимости группы исследования и группы контроля.

На результаты КИ могут оказывать влияние и другие систематические ошибки, такие как ошибки при сборе информации, ошибки памяти, ошибки при выбывании участников из исследования, ошибки при оценке и анализе результатов и некоторые другие [5, 6]. Все систематические ошибки могут создать видимость различий, когда в действительности их нет, или наоборот, скрыть различия, которые на самом деле существуют. Вероятность систематической ошибки существует при любых наблюдениях и может возникнуть на любом этапе исследования. Объем выборки не влияет на величину систематической ошибки.

Для того, чтобы с уверенностью говорить, что наблюдаемый результат является следствием изучаемого фактора, а не систематических ошибок, необходимо исключить или снизить их значимость. Это достигается на этапе формирования выборки через повышение ее репрезентативности или на этапе формирования групп сравнения посредством рандомизации, а также могут быть частично учтены на этапе анализа результатов исследования. Основным методом, позволяющим минимизировать влияние большинства систематических ошибок, является рандомизация — случайное распределение больных по группам сравнения. При этом системно действующие ошибки тоже равномерно распределяются по группам сравнения, и, при достаточной численности групп, перестают оказывать свое смещающее действие.

НАДЕЖНОСТЬ ИССЛЕДОВАНИЯ

Еще одним альтернативным объяснением различий результатов в сравниваемых группах может быть случайная ошибка. **Случайная ошибка** — это отклонение результата отдельного наблюдения (или измерения) от

его истинного значения, возникающее при оформлении учетных документов, измерении или регистрации данных, обусловленное исключительно случайным стечением обстоятельств. Случайная ошибка с одинаковой вероятностью может привести к завышению или занижению оценки результатов исследования. Любые наблюдения подвержены действию случайности, и полностью исключить случайные ошибки нельзя, но их можно минимизировать, используя более точные методы оценки параметров исследования, например, стандартизированные, или увеличивая число больных в исследовании.

Случайная ошибка может быть оценена и учтена на этапе анализа результатов с помощью статистического анализа, который позволяет ответить на вопрос, с какой вероятностью результаты исследования могли быть получены случайно. В медицинских исследованиях допустимым уровнем вероятности получения случайного результата является значение p меньше 0,05 [7].

Выраженная случайная ошибка обычно наблюдается в небольших выборках, сделанных из популяции с высокой степенью неоднородности характеристик (например, включаются жители и городские, и сельские, и мужчины, и женщины, и с вредными привычками, и без них в широком возрастном диапазоне). Чем выше неоднородность выборки, тем больше вероятность случайной ошибки, тем больше людей надо набрать в группы сравнения, чтобы повысить надежность выводов. Даже выраженная случайная ошибка не смещает (не искажает) результат исследования, но может не позволить выявить статистическую достоверность полученных выводов.

Контроль за уровнем системных ошибок осуществляется с помощью строгого выполнения требований дизайна исследования. КИ, благодаря особенностям своего дизайна, способны в разной степени контролировать влияние системных ошибок, поэтому они имеют определенные ограничения в степени достоверности. Надо отметить, что некоторые факторы, такие как использование несоответствующих статистических методов анализа, отсутствие поправок на систематические и случайные ошибки, недобросовестное обращение с данными, могут исказить результаты исследования вне зависимости от выбранного дизайна.

В научной медицине используются КИ, различающиеся по дизайну. Среди них можно выделить три основных дизайна, задачей которых являются выявление и изучение причинно-следственных связей. К ним относятся исследования типа «случай-контроль», когортные исследования и рандомизированные КИ [8, 9].

ДИЗАЙН И ДОКАЗАТЕЛЬНАЯ ЦЕННОСТЬ ИССЛЕДОВАНИЯ ТИПА «СЛУЧАЙ-КОНТРОЛЬ»

Этот вид КИ относится к наблюдательным исследованиям, при которых исследователи не вмешиваются в естественное течение процессов возникновения и распространения заболеваний. Они лишь наблюдают за развитием ситуации, которая никак не зависит от них, проводят сбор данных по изучаемому вопросу и делают выводы [10].

Исследования «случай-контроль» применяют для выявления неизвестных факторов риска известных заболеваний. Для того, чтобы выявить связь между клиническим исходом и предшествующим воздействием предполагаемого фактора, в исследовании набирают две группы людей. В основную группу включают тех, у которых имеется интересующая исследователей болезнь

или состояние (исход). Эта группа носит название «случаи». В группу «контроли» набирают людей, у которых данной болезни или состояния нет. Затем в анамнезе всех участников исследования определяют наличие или отсутствие факторов, которые могли бы быть причиной развития изучаемой болезни. Далее обе группы сравниваются по частоте встречаемости потенциальных факторов риска развития данного исхода и определяют статистическую значимость различий.

Особенность исследования «случай-контроль» состоит в том, что этот дизайн не предполагает проведения рандомизации при формировании групп сравнения, что приводит к неполной сопоставимости основной и контрольной групп в результате действия систематических ошибок.

Группа «случаев» выбирается из популяции больных, имеющих изучаемое заболевание или состояние, и на которую исследователь хотел бы распространить выводы, которые он предполагает получить. Группа «случаев» всегда должна быть репрезентативна исследуемой популяции. Недостаточная репрезентативность группы «случаев» (ошибка выборки) может привести к необоснованному обобщению результатов исследования.

Группу «контролей» исследователь подбирает в соответствии с характеристиками группы «случаи», а не в результате рандомизации, что и является источником систематической ошибки отбора. Основным условием при отборе контрольных лиц является их максимальная сопоставимость группе «случаев» по всем основным характеристикам, за исключением изучаемого заболевания. Для того, чтобы получить более достоверный результат, группа «контролей» должна быть максимально сопоставима с группой «случаи» [11]. Для этого подбор «контролей» должен проводиться из той же популяции, что и «случаи», желательно в тот же самый период времени. Например, и «случаи», и «контроли» должны быть набраны из людей, поступивших в одну больницу, или обслуживающихся в одной поликлинике, или проживающих в одном районе, или работающих на одном предприятии. При недостаточной сопоставимости «случаи» и «контроли» могут различаться по тяжести состояния, сопутствующей патологии, социальному статусу, наличию вредных привычек, использованию лекарственных средств, влияющих на состояние здоровья и пр. [12]. Для уменьшения ошибки отбора используют метод подбора пар, который заключается в индивидуальном подборе каждому участнику группы «случаи», соответствующему ему по набору отличительных признаков участника контрольной группы [13]. В итоге исследователи получают примерно одинаковые группы сравнения с единственным различием: наличие или отсутствие изучаемой болезни.

Одна из систематических ошибок отбора, по которой полученные в исследовании результаты могут не верно отражать истинный результат, может быть обусловлена действием неизвестного или неучтенного фактора, который способен влиять одновременно как на исход, так и на изучаемый фактор заболевания. Такой фактор называется «спутывающим или вмешивающимся фактором» или «спутывающей переменной» (конфаундером) [14].

Примером может служить исследование, изучающее связь между очередностью рождения ребенка (ребенок, родившийся первым по счету, 2-м, 3-м и т. д.) и наличием у него синдрома Дауна. В этом исследовании возраст матери будет спутывающей переменной, так как он одновременно влияет и на исход — более высокий возраст матери напрямую связан с вероятностью развития синдрома

Дауна у ребенка, и он также связан с очередностью рождения — каждый следующий ребенок, за исключением двойни, рождается, когда мать старше, чем она была на момент рождения 1-го ребенка.

Присутствие спутывающих факторов может быть или не быть распознано. В результате, выводы, полученные на основании наблюдательных исследований, могут не отражать действительного эффекта применения исследуемого вмешательства.

Ретроспективным исследованиям характерны систематические **ошибки на этапе сбора информации и ошибки памяти**. В исследовании типа «случай-контроль» поиск причинно-следственных связей всегда идет в направлении от следствия к предполагаемой причине, т. е. ретроспективно. При проведении ретроспективного исследования на момент его начала исследователь уже имеет информацию об интересующем его исходе и собирает информацию о событиях (возможных факторах риска), которые имели место в прошлом участников. Источником информации в данном исследовании выступают истории болезни или амбулаторные карты, находящиеся в организациях здравоохранения (т. е. вторичная информация), а также воспоминания пациентов, их родственников в рамках интервью или по результатам анкетирования. И с этим связано возникновение информационных систематических ошибок и ошибок памяти. Зарегистрированные в медицинских документах данные собирались для других целей и задач, исследователь не участвовал в их сборе и чаще всего не знает, кем и когда собиралась информация. Архивная информация может не в полной мере соответствовать цели проводимого исследования, может быть собрана некачественно, часть данных может отсутствовать. Информация, собранная со слов людей, может недостаточно точно отражать события прошлого. Может иметь значение и избирательность памяти больного и здорового человека. Например, больной лучше, чем здоровый, может вспомнить события, потенциально связанные с возникновением заболевания, и не вспомнить о некоторых фактах, представляющих интерес для исследователя. Ошибки памяти особенно актуальны, если они касаются информации о воздействии изучаемого фактора риска, что является главным недостатком всех ретроспективных исследований [5].

Наряду с ошибками выборки, ошибками отбора и ошибками сбора информации исследований «случай-контроль» не защищены и от случайных ошибок, что в целом создает много возможностей альтернативных объяснений полученным результатам. По этой причине доказательность этого типа исследований не очень высока.

Результатом исследования является формулировка гипотез о факторах риска заболеваний и состояний. Эти гипотезы затем должны быть подтверждены в более точных когортных исследованиях.

Несмотря на то, что дизайн исследования «случай-контроль» не дает возможности доказать наличие причинно-следственной связи, исследования этого типа являются единственно подходящими для изучения факторов риска редких заболеваний [7].

ДИЗАЙН И ДОКАЗАТЕЛЬНАЯ ЦЕННОСТЬ КОГОРТНЫХ ИССЛЕДОВАНИЙ

Когортные исследования также являются наблюдательными, т. е. данные собираются путем наблюдения за событиями без вмешательства исследователя [8].

Целью исследования являются поиск и выявление неизвестных последствий воздействия предполагаемых факторов риска на здоровье человека и изучение связи между ними. Для выполнения исследования из генеральной совокупности (популяции) отбирают группу лиц (когорту), которая должна представлять собой репрезентативную выборку популяции. Под когортой понимают группу людей, объединенную общими характеристиками или опытом в течение определенного периода времени, в которой ожидается возникновение новых случаев болезни. Объединяющей характеристикой может быть проживание людей в одном городе, подверженность воздействию вредных веществ, прохождение определенной медицинской процедуры, отношение к представителям одной профессии или социальной группы, факт рождения в определенный период времени и т. п. Исследуемую когорту составляют люди, подвергающиеся влиянию изучаемого фактора риска, а контрольную когорту — люди, на которых изучаемый фактор не действовал [15]. Контрольная группа подбирается из той же популяции, из которой составляется когорта, или это может быть другая когорта лиц, которые, как считается, мало или вообще не подвергались изучаемому воздействию, но в остальном по всем характеристикам максимально похожи с исследуемой группой. Эти когорты наблюдают в течение некоторого периода времени, чтобы выяснить, к каким исходам может привести данный фактор риска. Обязательным условием включения людей в исследуемую и контрольную когорты является отсутствие изучаемого заболевания у всех из них на момент начала наблюдения.

Далее обе группы сравнивают по частоте развития заболевания, определяют показатель относительного риска, величина которого свидетельствует о связи фактора риска с вероятностью исхода, и оценивают статистическую значимость различий.

Когортные исследования называются *проспективными*, если поиск причинно-следственных связей идет в направлении от причины к предполагаемому следствию. Другими словами, за когортой наблюдают с момента начала исследования, когда заболевания еще нет, и продолжают в течение времени, достаточном для развития предполагаемого исхода. При этом исследователь не может заранее знать его итоги, что исключает субъективизм в подборе исследуемых. В этом случае источником информации являются данные, которые предполагается получать в ходе исследования и которые исследователи регистрируют сами, а поэтому они более надежны и больше соответствуют целям исследования.

Когортные исследования могут быть и ретроспективными, когда на момент его начала исследователь уже чаще всего имеет информацию об интересующем его исходе и собирает информацию о событиях, которые имели место в прошлом этих людей, но формирование групп, тем не менее, осуществляется в зависимости от наличия или отсутствия факторов риска. Как и в других ретроспективных исследованиях, информация собирается по архивным документам (истории болезни, анкеты, результаты опроса участников и т. д.). Исследователь анализирует эти данные группы в прошлом, прослеживая развитие заболеваемости или смертности для всех членов исследуемой группы вплоть до настоящего времени [15].

Когортные исследования не свободны от действия систематических и случайных ошибок. **Ошибки, связанные с репрезентативностью** когорты, могут возникать, если ее

состав недостаточно полно соответствует той популяции, из которой она была выбрана [16]. Такая ситуация возможна, например, при формировании когорты из числа посетителей определенного медицинского центра, куда пациенты могут попасть не случайно, а в силу близкого проживания, или могут направляться в связи с более тяжелым состоянием либо в силу возможности оплачивать медицинские услуги, в то время как общая популяция включала не только пациентов медицинских центров, но также и больных городских стационаров, и поликлиник. Различия могут касаться возраста, пола, социально-экономического статуса, бытовых условий, состояния здоровья и т. д. Нередко бывает трудно обобщить результаты даже крупных КИ. Например, трудно определить степень репрезентативности богатого американского городка Фрамингем (Фрамингемское исследование факторов риска ИБС) хотя бы для населения США или степени репрезентативности исследования, проводимого на британских врачах хотя бы для представителей других профессий в Великобритании (исследование связи между курением и развитием рака легких).

На возможность перенесения полученных в исследовании данных на исходную популяцию, а также на популяции, имеющие схожие характеристики, оказывает влияние правильное определение численности когорты. Чем больше ее размер, тем точнее получаемые данные, тем больше они соответствуют генеральной совокупности [17].

Ошибки информации и памяти при проведении когортных исследований с ретроспективным сбором информации, как и при других ретроспективных исследованиях, состоят в том, что события прошлого сложно реконструировать без искажений. Часть документов, свидетельствующих о воздействии, например, вредного фактора в прошлом, может быть утеряна, а воспоминания родственников неточны. Ошибки сбора информации и памяти приводят к маскировке эффекта воздействия и искажению выводов.

Еще одной ошибкой проспективных когортных исследований является **ошибка выбывания из исследования**. В зависимости от изучаемого заболевания проспективные когортные исследования могут длиться достаточно долго: годы и даже десятилетия. При такой продолжительности наблюдений часть пациентов может «выпасть» из исследования вследствие переезда на другое место жительства, отказа от участия, смерти, потери контактов и других причин. Уменьшение численности когорт связано со снижением статистической мощности и, как следствие, с уменьшением надежности исследования. Считается, что потеря более 10% от размера когорты делает результаты исследования сомнительными, а выбывание более 20% участников приводит к их недостоверности [9, 18].

При проведении когортных исследований, как и при других дизайнах, могут возникать **ошибки отбора**, к которым относятся все факторы, помимо изучаемых, которые при неравномерном распределении между исследуемой и контрольной когортами способны приводить к их несопоставимости и оказывать влияние на результаты исследования. Примерами может служить различия по лечению, по числу посещений врача или любым другим параметрам. К значимым различиям групп сравнения может привести включение больных в исследование в разное время. Например, при смешанных ретроспективно-проспективных исследованиях могут быть не учтена разница в сроках и точности постановки

диагноза, а также подходах к терапии в прошлом, скажем, 15 лет назад, и в настоящее время. В этом случае изменения в исходах могут в большей степени объясняться разницей в оценке тяжести заболевания, чем в эффекте лечения.

Среди факторов, которые могут стать источником систематической ошибки, можно выделить неучтенные или неизвестные спутывающие факторы (конфаундеры). В результате действия спутывающих факторов эффект изучаемого фактора может быть переоценен или недооценен. Для исключения влияния известных конфаундеров обе группы должны быть максимально сопоставимы по как можно большему количеству параметров, за исключением исследуемых [19]. Существуют приемы, позволяющие при анализе данных учесть влияние всех факторов, о существовании которых нам известно. Но даже после всех поправок могут остаться неучтенными те спутывающие переменные, о существовании которых мы просто не догадываемся. Баланс неизвестных конфаундеров достигается с помощью рандомизации. Но в когортных исследованиях рандомизацию провести нельзя, поскольку наблюдательный подход к изучению связей между событиями исключает случайное распределение людей в сравниваемые группы.

Невозможность контроля неизвестных конфаундеров является серьезным недостатком, который отличает наблюдательные исследования от рандомизированного эксперимента. К сожалению, полностью устранить этот недостаток наблюдательных исследований невозможно. Отсюда и неполная степень доказательности наблюдательных исследований, в частности и когортных [14].

Несмотря на то что проспективные когортные исследования не исключают всех возможных ошибок, они являются наиболее доказательными среди наблюдательных исследований и точнее отражают причинно-следственные отношения. Когортный дизайн является наилучшим видом исследования в случаях, когда необходимо изучить влияние потенциально вредных факторов риска на возникновение болезни, т. е. в случаях, когда проведение эксперимента на людях невозможно.

ДИЗАЙН И ДОКАЗАТЕЛЬНАЯ ЦЕННОСТЬ РАНДОМИЗИРОВАННЫХ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

Рандомизированное клиническое исследование (РКИ) — это экспериментальное исследование, в котором исследователь моделирует клиническую ситуацию, в наилучшей степени подходящей для изучения причинно-следственных отношений между изучаемыми явлениями. Экспериментальные исследования, как правило, проводятся для проверки причинно-следственных гипотез при изучении эффективности различных методов лечения и профилактики, как медикаментозных, так и немедикаментозных. В экспериментальных исследованиях этически приемлемо изучать только воздействия факторов, которые, как предполагается, приносят пользу пациенту, поэтому искусственное вмешательство в естественный ход событий происходит за счет или устранения подозреваемых факторов, вызывающих болезни, или за счет назначения лекарственных средств, методов или мероприятий, способных благоприятно воздействовать на изучаемое заболевание [20–22].

Дизайн РКИ во многом напоминает дизайн когортных исследований. Из генеральной совокупности (популяции)

согласно строгим критериям включения и невключения отбирают группу лиц, которая должна представлять собой репрезентативную выборку популяции. Затем пациенты, включенные в исследование, случайным образом, т. е. независимо от воли исследователя, распределяются в группу исследования (получают изучаемое вмешательство) и группу контроля (получают плацебо или известное вмешательство с доказанной эффективностью). В период исследования за участниками проводят планомерное наблюдение с регистрацией их субъективного и объективного состояния. В конце исследования проводится оценка различий результатов обеих групп сравнения и оценивается их статистическая значимость.

Экспериментальные исследования могут быть проспективными, ретроспективными и смешанными (исследование с историческим контролем). При проспективном исследовании сбор и регистрацию информации о пациенте проводит сам исследователь, при ретроспективном — информация собирается по архивным медицинским документам или опросам больных, что снижает ее надежность.

Дизайн РКИ отличается от других видов исследований возможностью проведения процедуры рандомизации. Именно рандомизация позволяет нивелировать значимость большинства систематических ошибок, которые встречаются при проведении КИ. Сюда относятся **систематические ошибки, создающие дисбаланс групп сравнения, включая конфаундинг**. Именно поэтому вероятность того, что полученные результаты не являются следствием действия изучаемого вмешательства, а имеют альтернативное объяснение, очень низка. Однако это справедливо только в том случае, если исследователь не нарушает основной принцип рандомизации, который заключается в том, что каждый член выборки должен иметь равные шансы попасть как в группу исследования, так и в группу контроля [23]. Примером неправильно проведенной рандомизации является отбор больных в группу сравнения по показаниям, по порядку отбора, по дням недели, по номеру истории болезни, страхового полиса или дате рождения. Эти основания сами вносят в процесс формирования групп сравнения систематическую ошибку. Лучше всего пользоваться таблицей случайных чисел, методом конвертов или путем централизованного компьютерного распределения вариантов лечения. При нарушении принципа равных шансов равномерное распределение влияния систематических ошибок не происходит и доказательность такого исследования опускается до уровня когортного наблюдения [20].

Рандомизированные исследования с историческим контролем считаются менее точными по сравнению с проспективными из-за **ошибок при сборе информации и ошибок памяти**, а также из-за возможных различий в критериях диагностики и точности обследования больных контрольной группы. **Систематическая ошибка, связанная с выбыванием** больных из исследования при его длительном проведении, требует коррекции на этапе оценки результатов.

В рандомизированных исследованиях не полностью исключаются **ошибки выборки**, снижающие возможность обобщать полученные результаты на более широкую популяцию больных. Например, большинство РКИ делаются на относительно молодых больных без сопутствующих заболеваний, а изученные в этих условиях лекарственные средства потребляют люди старшего возраста, как правило, страдающие целым набором

заболеваний. Низкой репрезентативностью обладают рандомизированные исследования, проведенные на селективных группах. Использование селективных групп оправдано при изучении нового лекарственного препарата с целью подтверждения его фармакологической активности и определения его безопасных доз на первых фазах КИ.

Некоторые **систематические ошибки, связанные с позитивным ожиданием** больных от участия в КИ (плацебо эффект), желательно выявлять и устранять на этапе отбора. Это необходимо поскольку неодинаковые ожидания пациентов в сравниваемых группах могут в значительной мере повлиять на результаты исследования. Психологические установки и ожидания возникают не только у больных, но и у медицинского персонала, проводящего исследование, что может проявиться предвзятостью исследователя при отборе больных и субъективностью в оценках пограничных результатов исследования. Для исключения этих психологических влияний необходимо ограничить информированность исследователей о назначаемых лекарственных средствах в группах сравнения (слепые, двойные слепые исследования). Показано, что отсутствие двойного ослепления может преувеличивать эффективность лекарственных средств в среднем на 15–20%. [21]. Применение слепого метода в отношении больных, врачей, исследователей, оценивающих клинические исходы, и даже статистиков позволяет заметно снизить вероятность возникновения систематической ошибки такого типа.

Несмотря на рандомизацию, сравниваемые группы могут быть неодинаковыми из-за недостаточного размера выборки и связанным с этим усилением влияния **случайной ошибки**. Вероятность случайной ошибки также повышается при высокой гетерогенности (неоднородности) популяции, из которой составляется выборка. Поэтому небольшие РКИ или РКИ, проведенные только в одном медицинском центре, как правило, обладают недостаточной репрезентативностью (неоднородность выборки), пониженной внутренней валидностью (дисбаланс сравниваемых групп) и недостаточной надежностью (повышенная вероятность случайной ошибки). Поскольку случайная ошибка и гетерогенность выборки уменьшаются при увеличении ее размера, доверие к крупным многоцентровым РКИ всегда выше. Для большей надежности с целью доказательства причинно-следственной связи результаты РКИ необходимо перепроверять много раз. Желательно, чтобы исследование было повторено разными исследователями на многих отличающихся выборках, в разное время, при различных условиях. Полностью исключить влияние случайной ошибки нельзя, поэтому всегда есть 5-процентная вероятность того, что полученный в исследовании результат будет обусловлен стечением случайных обстоятельств [24].

Несмотря на возможные проблемы, правильно спланированные и проведенные РКИ позволяют получить высокодостоверные выводы и являются золотым стандартом доказательных КИ.

СТРУКТУРА И ДОКАЗАТЕЛЬНАЯ ЦЕННОСТЬ СИСТЕМАТИЧЕСКОГО ОБЗОРА И МЕТААНАЛИЗА

Именно в связи с тем, что даже РКИ не до конца точны, были разработаны методы доказательной медицины, такие как систематические обзоры, включающие или не включающие метаанализ.

Систематический обзор (СО) представляет собой аналитическое исследование имеющихся в литературе аналитических наблюдательных и экспериментальных исследований и служит инструментом вторичного анализа научных публикаций.

Исследование начинается с формулировки клинического вопроса, на который хотят получить ответ. Обычно он касается эффективности методов лечения, профилактики или диагностики. Затем проводятся поиск и отбор наилучших работ, которые исследуют одну и ту же проблему, схожи по структуре исследования, обладают наиболее сильным дизайном и выполнены наиболее тщательным образом. Отбор исследований базируется на определенных, четких критериях включения и исключения, которые должны быть обоснованы и определены заранее. Затем проводят обобщение результатов всех исследований, прошедших отбор, и на основании этих обобщенных результатов формулируется ответ на поставленный клинический вопрос. Он может быть выражен как подтверждение наличия причинно-следственной связи, или ее отрицание или возможна ситуация, когда качественно проведенных первичных исследований недостаточно, чтобы дать однозначный ответ на поставленный вопрос [25].

Источником информации для СО служат все обнаруженные опубликованные аналитические наблюдательные и экспериментальные исследования, касающиеся изучаемого клинического вопроса. Обычно поиск данных проводят в электронных информационных базах данных, в которых включаются только материалы, отвечающие определенным критериям методологического качества. К ним в первую очередь относятся Medline, Embase, Cochrane Library, отечественная база данных eLibrary.ru.

Однако не все исследования могут быть включены в СО, поскольку СО обобщает результаты только относительно однородных исследований. Считается неправомерным обобщение результатов исследований, существенно различающихся по характеристикам пациентов, по различным аспектам применения сравниваемых ЛС, а также по критериям оценки изучаемого исхода, так как эти различия повышают неоднородность (гетерогенность) обобщаемых данных и снижают достоверность выводов.

СО может включать применение статистического метода, который дает возможность обобщить результаты нескольких первичных исследований, как если бы это было одно крупное исследование, и сделать на их основании общий статистический вывод. Такой метод называется метаанализом. Объединение исследований обеспечивает большую выборку для анализа и большую статистическую мощность. Это позволяет повысить точность оценки эффекта анализируемого вмешательства и поднимает доказательность данных систематического обзора с метаанализом на более высокий уровень, чем у отдельных экспериментальных или описательных исследований. Метаанализ может обнаружить эффект, который отдельные эксперименты не обнаружили из-за нехватки статистической мощности (малое число участников в каждом из них по отдельности), а также помогает сделать общий вывод на основе нескольких исследований, давших разные и даже противоречивые результаты [26, 27].

Несмотря на все достоинства, метаанализ также не свободен от влияния систематических ошибок и может содержать ложные выводы. Среди них такие систематические ошибки, как ошибки включения работ в СО и публикационное смещение [28].

Ошибки включения отражают низкое качество систематического обзора. Известно, что качество метаанализа существенно зависит от качества включенных в него исходных исследований и статей, то есть от качества самого систематического обзора, на основании которого он проводится. Метаанализ несет в себе систематические ошибки всех тех первичных работ, из которых он состоит. В случаях, когда публикуемая научная литература отражает ложные утверждения, проведенный на ее основе метаанализ также подтвердит ложные результаты.

Публикационное смещение возникает вследствие того, что некоторые проведенные исследования, результаты которых не выявили статистически значимой разницы между группами сравнения или не отличаются от уже известных данных, остались неопубликованными, и, следовательно, не могли быть включены в метаанализ. Тогда доля публикаций с положительными результатами становится больше, чем есть на самом деле, что приводит к переоценке усредненного эффекта.

Нарушение методологии проведения СО — недостаточно полный поиск данных, несоблюдение строгих критериев отбора и включение исследований низкого качества, ведет к накоплению систематических ошибок и снижает достоверность результатов СО. Так, большое высококачественное РКИ может дать более надежные результаты, чем метаанализ нескольких маленьких.

Тем не менее, систематические обзоры и метаанализы высокого качества являются основой аналитической базы доказательной медицины и очень ценным инструментом при принятии решения о выборе наиболее эффективных и безопасных методов лечения и профилактики.

ИЕРАРХИЯ ДОКАЗАТЕЛЬСТВ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

В настоящее время для формирования клинических рекомендаций по вопросам профилактики, диагностики, лечения и реабилитации используются результаты КИ различного дизайна. Для ориентации в относительной силе их доказательности была предложена иерархия доказательств, которую можно определить, как ранжирование различных дизайнов КИ по степени их подверженности систематическим ошибкам [29]. На вершине иерархии находится метод с наибольшей свободой от системного смещения, что с большой уверенностью предполагает, что истинный эффект близок к полученному в исследовании. В основании иерархии находятся виды исследований, не свободные от влияний многих систематических ошибок, что значительно снижает уверенность в истинности полученных в этих исследованиях результатов.

В различных странах и крупных медицинских организациях разработаны и используются классификации уровней доказательности, имеющие некоторые различия в протоколах оценки КИ. В Российской Федерации оценка уровней достоверности КИ, включаемых в клинические рекомендации, проводится на основании результатов одного или нескольких КИ наиболее высокого ранга в соответствии с единой шкалой, предусмотренной требованиями, утвержденными приказом Минздрава России от 28.02.2019 № 103н.

Уровень достоверности доказательств (УДД) трактуется как степень уверенности в том, что найденный эффект от применения медицинского вмешательства является истинным [30]. Выделяют пять уровней достоверности доказательств (табл. 1).

Таблица 1. Шкала определения уровней достоверности доказательств (УДД) для лечебных, реабилитационных и профилактических вмешательств

УДД	Иерархия дизайнов клинических исследований (по убыванию уровня достоверности доказательств от 1 до 5)
1	Систематический обзор РКИ с применением метаанализа
2	Отдельные РКИ и систематические обзоры исследований любого дизайна (за исключением РКИ с применением метаанализа)
3	Нерандомизированные сравнительные исследования, в том числе когортные
4	Несравнительные исследования, описание клинического случая или серии случаев, исследование «случай-контроль»
5	Доклинические исследования (с обоснованием механизма действия вмешательства) или мнение экспертов

Таблица 2. Шкала определения уровня убедительности рекомендаций

УУР	Расшифровка
А	Сильная рекомендация (при одновременном выполнении двух условий): <ul style="list-style-type: none"> • все рассматриваемые критерии эффективности (исходы) являются важными, • все исследования имеют высокое или удовлетворительное методологическое качество, • их выводы по интересующим исходам являются согласованными
В	Условная рекомендация (при выполнении хотя бы одного из условий): <ul style="list-style-type: none"> • не все рассматриваемые критерии эффективности (исходы) являются важными, • не все исследования имеют высокое или удовлетворительное методологическое качество, • их выводы по интересующим исходам не являются согласованными
С	Слабая рекомендация — отсутствие доказательств надлежащего качества (при выполнении хотя бы одного из условий): <ul style="list-style-type: none"> • все рассматриваемые критерии эффективности (исходы) являются неважными, • все исследования имеют низкое методологическое качество, • их выводы по интересующим исходам не являются согласованными

В свою очередь сформулированные на основании результатов КИ рекомендации также ранжируются по **уровню их убедительности** (УУР), которая определяется как степень уверенности в достоверности эффекта вмешательства и в том, что следование рекомендациям принесет больше пользы, чем вреда в конкретной ситуации.

Определение уровня убедительности рекомендаций строится на оценке методологического качества КИ, согласованности результатов КИ, на основании которых оценивался УДД, и важности исходов.

Методологическое качество КИ оценивается по соответствующим балльным вопросам, разработанным отдельно для СО, РКИ, когортных исследований и исследований «случай-контроль». Результаты КИ считаются согласованными, если во всех КИ выявлено одинаковое направление эффекта и, как следствие, сделаны одинаково направленные выводы, то есть преимущество вмешательства А перед вмешательством В выявлено во всех КИ наиболее высокого дизайна [31]. Важность (значимость) исходов по результатам КИ определяется как важные и неважные. К важным исходам относятся: все клинические исходы («твердые конечные точки»), суррогатные исходы, оцененные по валидизированным шкалам, суррогатные исходы, для которых доказана ассоциированность с клиническими исходами по результатам КИ. К неважным исходам относятся суррогатные исходы при отсутствии КИ, подтверждающих ассоциированность с клиническими исходами («твердыми конечными точками»). Это показатели невалидизированных клинических шкал, лабораторные показатели, субъективные оценки пациентов (в том числе по визуальным аналоговым шкалам), продолжительность симптомов.

Оценка уровня убедительности рекомендаций для диагностических, лечебных, профилактических вмешательств и реабилитационных мероприятий также проводится в соответствии с единой шкалой, предусмотренной требованиями, утвержденными приказом Минздрава России от 28.02.2019 № 103н. По убедительности выделяют рекомендации сильного,

условного и слабого уровней, которые обозначаются буквами латинского алфавита А, В, С (табл. 2).

Правильная оценка уровней убедительности рекомендаций и уровней достоверности КИ, на которых построены рекомендации, должна обеспечивать их высокую научную обоснованность, что соответствует требованиям медицины, основанной на доказательствах.

ЗАКЛЮЧЕНИЕ

В качестве инструмента для получения новых знаний в области медицины применяют различные клинические эпидемиологические исследования, предназначенные для достижения различных целей и задач. КИ различаются по своей структуре и точности, с которой они способны оценивать причинно-следственные связи между явлениями. Поэтому к оценке достоверности их выводов надо подходить с пониманием тех ограничений, которые характерны для различных дизайнов. Точность КИ зависит от влияния многих факторов, которые способны приводить к искажению получаемых результатов по сравнению с их истинными значениями. Влияние этих факторов — систематических и случайных ошибок — позволяет делать альтернативные выводы о причинах обнаруженных различий. Дизайны различных КИ допускают влияние большего или меньшего числа таких факторов, что отражается на достоверности результатов КИ. Ни одно исследование не свободно от действия систематических и случайных ошибок. Однако наблюдательные исследования подвержены им в большей степени, чем экспериментальные. Это объясняется тем, что в силу особенностей дизайна наблюдательные исследования не могут контролировать ошибки, связанные с возможной несопоставимостью групп сравнения. Они способны выявлять наличие статистической связи между явлениями, но доказать, что связь носит причинно-следственный характер, могут только РКИ. Точность РКИ может быть повышена с помощью систематических обзоров с метаанализом.

Литература

1. Марцевич С. Ю., Дроздова Л. Ю., Воронина В. П. Здоровье и образование врача: две составляющие успеха. Рациональная фармакотерапия в кардиологии. 2010; 6 (1): 73–76.
2. Гланц С. Медико-биологическая статистика. М.: Изд-во «Практика», 1999; 459 с.
3. Беграмбекова Ю. Л. Методология анализа клинических исследований и определения их связей с клинической действительностью. Часть I. Кардиология: Новости. Мнения. Обучение. 2016; 4 (11): 69–75.
4. Everitt BS, Howell DC. Encyclopedia of Statistics in Behavioral Science. Chichester: John Wiley & Sons Ltd., 2005; 1451–62.
5. Vandembroucke JP, Von Elm E, Altman DG и др. Повышение качества отчетов о наблюдательных исследованиях в эпидемиологии (STROBE): разъяснения и уточнения. Вопросы современной педиатрии. 2022; 21 (3): 173–208. DOI 10.15690/vsp.v21i3.2426.
6. Марцевич С. Ю., Кутишенко Н. П., Лукина Ю. В., Лукьянов М. М., Драпкина О. М. Клинические исследования по оценке лекарственной терапии. Виды, достоверность результатов, место в современной доказательной медицине. Рациональная фармакотерапия в кардиологии. 2021; 17 (2): 249–255.
7. Баврина А. П. Наблюдательные эпидемиологические исследования и особенности представления результатов в научном отчете. Медицинский альманах. 2021; 2 (67): 83–89.
8. Наркевич А. Н., Виноградов К. А. Дизайн медицинского исследования. Социальные аспекты здоровья населения. 2019; 5 (65): 13. [Электронный ресурс]. Режим доступа: URL: <http://vestnik.mednet.ru/content/view/1108/30/lang, ru. DOI: 10.21045/2071-5021-2019-65-5-13>.
9. Simpson MR. Kohortstudier [Cohort studies]. Tidsskr Nor Laegeforen. 2021 Oct 25; 141 с. Norwegian. DOI: 10.4045/tidsskr.21.0511. PMID: 34726047.
10. Гржибовский А. М., Иванов С. В., Горбатова М. А. Исследования типа «случай-контроль» в здравоохранении. Наука и здравоохранение. 2015; 4: 5–17.
11. Pearce N. Analysis of matched case-control studies. BMJ. 2016 Feb 25; 352: i969. DOI: 10.1136/bmj.i969. PMID: 26916049; PMCID: PMC4770817.
12. Котеров А. Н. Критерии причинности в медикобиологических дисциплинах: история, сущность и радиационный аспект. Сообщение 1. Постановка проблемы, понятие о причинах и причинности, ложные ассоциации. Радиационная биология. Радиоэкология. 2019; 59 (1): 5–36. DOI: org/10.1134/S0869803119010065.
13. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 2018 Jan; 33 (1): 5–14. DOI: 10.1007/s10654-017-0325-0. Epub 2017 Nov 3. PMID: 29101596; PMCID: PMC9188848.
14. Thoresen M. Konfundering — et tilbakevendende problem [Confounding — a recurring problem]. Tidsskr Nor Laegeforen. 2018 Feb 5; 138 (3). Norwegian. DOI: 10.4045/tidsskr.17.0795. PMID: 29411587.
15. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract. 2009; 113 (3): c214–7. DOI: 10.1159/000235241. Epub 2009 Aug 18. PMID: 19690438.
16. Холматова К. К., Харькова О. А., Гржибовский А. М. Особенности применения когортных исследований в медицине и общественном здравоохранении. Экология человека. 2016; 4: 56–64.
17. Cortegiani A, Absalom AR. Importance of proper conduct of clinical trials. Br J Anaesth. 2021 Feb; 126 (2): 354–356. DOI: 10.1016/j.bja.2020.09.030. Epub 2020 Oct 26. PMID: 33121749.
18. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 407–416. DOI: 10.1111/aogs.13319. Epub 2018 Mar 5. PMID: 29415329.
19. Pripp AH. Matching i kohortstudier [Matching in cohort studies]. Tidsskr Nor Laegeforen. Norwegian. 2020 Jan 27; 140 (2). DOI: 10.4045/tidsskr.19.0308. PMID: 32026854.
20. Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. 2020 Jul; 158 (1S): S79–S87. DOI: 10.1016/j.chest.2020.03.013. PMID: 32658656; PMCID: PMC8176647.
21. Kennedy-Martin T, Curtis S, Faries D, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015; 16: 495. DOI 10.1186/s13063-015-1023-4.
22. Петровская Ю. А., Огородова Л. М., Петровский Ф. И. Принципы принятия решений о назначении лекарственных средств. Что еще надо знать при интерпретации результатов клинических исследований? Педиатрическая фармакология. 2018; 15 (4): 343–348. DOI: org/10.15690/pf.v15i4.1949.
23. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials. 2017; 18 (1): 498–510. DOI: 10.1186/s13063-017-2240-9.
24. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 380–387. DOI: 10.1111/aogs.13309. Epub 2018 Feb 27. PMID: 29377058.
25. Donato H, Donato M. Etapas na Condução de uma Revisão Sistemática [Stages for Undertaking a Systematic Review]. Acta Med Port. 2019 Mar 29; 32 (3): 227–235. Portuguese. DOI: 10.20344/amp.11923. Epub 2019 Mar 29. PMID: 30946795.
26. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. Urol Oncol. 2008 May-Jun; 26 (3): 320–9. DOI: 10.1016/j.urolonc.2006.12.001. Epub 2007 Nov 7. PMID: 18452828.
27. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018 Sep; 74 (3): 785–794. DOI: 10.1111/biom.12817. Epub 2017 Nov 15. PMID: 29141096; PMCID: PMC5953768.
28. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008 Aug 28; 3 (8): e3081. DOI: 10.1371/journal.pone.0003081. PMID: 18769481; PMCID: PMC2518111.
29. Stegenga, J. Down with the Hierarchies. Topoi. 2014; 33 (2): 313–322.
30. Приказ Министерства здравоохранения Российской Федерации от 28.02.2019 № 103н «Об утверждении порядка и сроков разработки клинических рекомендаций, их пересмотра, типовой формы клинических рекомендаций и требований к их структуре, составу и научной обоснованности включаемой в клинические рекомендации информации». [Электронный ресурс]. Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_324276/.
31. Омеляновский В. В., Авксентьева М. В., Железнякова И. А., Сухоруких О. А., Хачатрян Г. Р., Галеева Ж. А. [и др.]. Клинические рекомендации как инструмент повышения качества медицинской помощи. Онкопедиатрия. 2014; 4 (4): 246–259.

References

1. Martsevich Slu, Drozdova Llu, Voronina VP. Zdorov'e i obrazovanie vracha: dve sostavliaiushchie uspekha. Ratsional'naia Farmakoterapiia v Kardiologii. 2010; 6 (1): 73–76. Russian.
2. Glants S. Mediko-biologicheskaja statistika. M.: Izd-vo «Praktika», 1999; 459 p.
3. Begrambekova lu L. Metodologija analiza klinicheskikh issledovaniij i opredeleniia ikh sviazei s klinicheskoi deistvitel'nost'iu. Chast' I. Kardiologija: Novosti. Mneniia. Obuchenie. 2016; 4 (11): 69–75. Russian.
4. Everitt BS, Howell DC. Encyclopedia of Statistics in Behavioral Science. Chichester: John Wiley & Sons Ltd., 2005; 1451–62.
5. Vandembroucke JP, Von Elm E, Altman DG, et al. Povyshenie kachestva otchetov o nabliudatel'nykh issledovaniakh v epidemiologii (STROBE): raz'iasneniia i utochneniia. Voprosy sovremennoj pediatrii. 2022; 21 (3): 173–208. DOI 10.15690/vsp.v21i3.2426. Russian.
6. Martsevich Slu, Kutishenko NP, Lukina luV, Lukianov MM, Drapkina OM. Klinicheskie issledovaniia po otsenke lekarstvennoi

- terapii. Vidy, dostovernost' rezul'tatov, mesto v sovremennoi dokazatel'noi meditsine. Ratsional'naia Farmakoterapiia v Kardiologii. 2021; 17 (2): 249–255. Russian.
7. Bavrina AP. Nabludatel'nye epidemiologicheskie issledovaniia i osobennosti predstavleniia rezul'tatov v nauchnom otchete. Meditsinskii al'manakh. 2021; 2 (67): 83–89. Russian.
 8. Narkevich AN, Vinogradov KA. Dizain meditsinskogo issledovaniia. Sotsial'nye aspekty zdorov'ia naseleniia. 2019; 5 (65): 13. Available from URL: <http://vestnik.mednet.ru/content/view/1108/30/lang,ru>. DOI: 10.21045/2071–5021–2019–65–5–13. Russian.
 9. Simpson MR. Kohortstudier [Cohort studies]. Tidsskr Nor Laegeforen. 2021 Oct 25; 141. Norwegian. DOI: 10.4045/tidsskr.21.0511. PMID: 34726047.
 10. Grijbovskii AM, Ivanov SV, Gorbato MA. Case-control studies in health sciences. Nauka i Zdravookhranenie [Science & Healthcare]. 2015; 4, 5–17. Russian.
 11. Pearce N. Analysis of matched case-control studies. BMJ. 2016 Feb 25; 352: i969. DOI: 10.1136/bmj.i969. PMID: 26916049; PMCID: PMC4770817.
 12. Koterov A N. Kriterii prichinnosti v medikobiologicheskikh distsiplinakh: istoriia, sushchnost' i radiatsionnyi aspekt. Soobshchenie 1. Postanovka problemy, poniatie o prichinakh i prichinnosti, lozhnye assotsiatsii. Radiatsionnaia biologiiia. Radioekologiiia. 2019; 59 (1): 5–36. DOI: org/10.1134/S0869803119010065. Russian.
 13. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 2018 Jan; 33 (1): 5–14. DOI: 10.1007/s10654–017–0325–0. Epub 2017 Nov 3. PMID: 29101596; PMCID: PMC9188848.
 14. Thoresen M. Konfundering — et tilbakevendende problem [Confounding — a recurring problem]. Tidsskr Nor Laegeforen. 2018 Feb 5; 138 (3). Norwegian. DOI: 10.4045/tidsskr.17.0795. PMID: 29411587.
 15. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract. 2009; 113 (3): c214–7. DOI: 10.1159/000235241. Epub 2009 Aug 18. PMID: 19690438.
 16. Kholmatova KK, Khar'kova OA, Grzhibovskii AM. Osobennosti primeneniia kogortnykh issledovaniia v meditsine i obshchestvennom zdravookhraneni. Ekologiiia cheloveka. 2016; 4: 56–64. Russian.
 17. Cortegiani A, Absalom AR. Importance of proper conduct of clinical trials. Br J Anaesth. 2021 Feb; 126 (2): 354–356. DOI: 10.1016/j.bja.2020.09.030. Epub 2020 Oct 26. PMID: 33121749.
 18. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 407–416. DOI: 10.1111/aogs.13319. Epub 2018 Mar 5. PMID: 29415329.
 19. Pripp AH. Matching i kohortstudier [Matching in cohort studies]. Tidsskr Nor Laegeforen. 2020 Jan 27; 140 (2). Norwegian. DOI: 10.4045/tidsskr.19.0308. PMID: 32026854.
 20. Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. 2020 Jul; 158 (1S): S79–S87. DOI: 10.1016/j.chest.2020.03.013. PMID: 32658656; PMCID: PMC8176647.
 21. Kennedy-Martin T, Curtis S, Faries D, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015; 16: 495. DOI 10.1186/s13063–015–1023–4.
 22. Petrovskaya YuA, Ogorodova LM, Petrovskiy FI. Principles of Decision Making on the Prescription of Medicines. What else Do You Need to Know when Interpreting the Results of Clinical Trials? Pediatric pharmacology. 2018; 15 (4): 343–348. DOI: org/10.15690/pf.v15i4.1949. Russian.
 23. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials. 2017; 18 (1): 498–510. DOI:10.1186/s13063–017–2240–9.
 24. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 380–387. Available from URL: DOI: 10.1111/aogs.13309. Epub 2018 Feb 27. PMID: 29377058.
 25. Donato H, Donato M. Etapas na Condução de uma Revisão Sistemática [Stages for Undertaking a Systematic Review]. Acta Med Port. 2019 Mar 29; 32 (3): 227–235. Portuguese. DOI: 10.20344/amp.11923. Epub 2019 Mar 29. PMID: 30946795.
 26. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. Urol Oncol. 2008 May-Jun; 26 (3): 320–9. DOI: 10.1016/j.urolonc.2006.12.001. Epub 2007 Nov 7. PMID: 18452828.
 27. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018 Sep; 74 (3): 785–794. DOI:10.1111/biom.12817. Epub 2017 Nov 15. PMID: 29141096; PMCID: PMC5953768.
 28. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Gherzi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008 Aug 28; 3 (8): e3081. DOI: 10.1371/journal.pone.0003081. PMID: 18769481; PMCID: PMC2518111.
 29. Stegenga J. Down with the Hierarchies. Topoi. 2014; 33 (2): 313–322.
 30. Prikaz Ministerstva zdravookhraneniia Rossiiskoi Federatsii ot 28.02.2019 № 103n "Ob utverzhdenii poriadka i srokov razrabotki klinicheskikh rekomendatsii, ikh peresmotra, tipovoi formy klinicheskikh rekomendatsii i trebovaniia k ikh strukture, sostavu i nauchnoi obosnovannosti vkluchaemoi v klinicheskie rekomendatsii informatsii". Available from URL: http://www.consultant.ru/document/cons_doc_LAW_324276/.
 31. Omelianovskii VV, Avksenteva MV, Zhelezniakova IA, Sukhorukikh OA, Khachatryan GR, Galeeva ZhA, et al. Klinicheskie rekomendatsii kak instrument povysheniia kachestva meditsinskoi pomoshchi. Onkopediatriia. 2014; 4 (4): 246–259. Russian.