

Виды исследований в доказательной медицине

Данишевский К. Д.

д.м.н., профессор Высшей школы экономики. Москва.

В статье описываются различные дизайны исследований, применяемые сегодня в медицинской науке, рассматривается иерархия их достоверности, сравнивается надежность получаемых данных.

Ключевые слова: дизайн исследования, описание случая, описание серии случаев, проспективное когортное исследование, рандомизированное исследование, систематический обзор

Существует множество классификаций исследований. Разделяют описательные и аналитические исследования. Биомедицинские исследования можно поделить на наблюдательные и экспериментальные, качественные и количественные, а также на лабораторные, на животных и на людях. Существуют эпидемиологические и социологические исследования, а также исследования по оценке экономической эффективности вмешательств. В современной классической эпидемиологии, иерархия основных эпидемиологических исследований, в порядке увеличения степени достоверности оценки причинно-следственной связи между явлениями, распределена следующим образом:

1. описание случая (клинического),
2. серии случаев, одномоментное,
3. случай-контроль,
4. проспективное когортное (или панельное),
5. экспериментальное, в частности, рандомизированное контролируемое испытание.

Наконец, доказательная медицина получила свой базовый инструментарий в результате разработки методик обобщающих исследований: систематических обзоров.

Одним из первых исследований с контрольной группы можно считать эксперимент, поставленный по легенде в конце XVIII века королем Швеции Густавом III. Считая, что кофе является страшным ядом, он решил провести «клиническое испытание»: приговорил одного заключенного к ежедневному потреблению кофе, а второго, представлявшего контрольную группу, к ежедневному употреблению чая. Два врача были приставлены для наблюдения за ходом исследования. Ожидалось, что потреблявший кофе заключенный должен был умереть в ближайшее время. Исследование не удалось по ряду причин. Во-первых, вскоре умерли оба врача следивших за ходом исследования. Затем король был убит. Оба заключенных жили долго, так,

пивший чай умер в возрасте 83 года, о втором заключенном данных нет. Стоит отметить, что результаты исследования не повлияли на принимающиеся решения: кофе в Швеции запрещали в 1794 и 1822 годах [1]. При том, что проведение исследований с контрольной группой всего на двух людях сейчас было бы неприемлемым, равно как и исследование факторов риска в экспериментальных испытаниях, стоит отметить, что для изучения сильных связей между явлениями с немедленно наступающим эффектом требуются небольшие выборки. Так, наиболее эффективные лекарства изучались в очень небольших исследованиях, а разница исходов в экспериментальной и контрольной группе порой может даже столь очевидной, что не требует серьезной статистической обработки.

Противофактическая теория (Counterfactual theory)

Идеальным аналитическим исследованием, однозначно доказывающим или отвергающим вред кофе, было бы такое, в котором мы бы получили информацию о том, что бы было с одним и тем же человеком, в зависимости от того, если бы он пил или не пил кофе. Однако, человек может либо потреблять, либо не потреблять кофе, делать и то и то в течение одного итого же периода времени невозможно.

В ряде исследований применяются «анти-противофактические» подходы. Во-первых, может измеряться так называемый ответ, зависящий от дозы. При этом исследуемые поделены на три или большее количество групп, например, на тех, кто пьют много кофе, потребляют его в небольшом количестве, и тех, кто вообще не потребляет кофе. Во-вторых, при изучении кратковременных обратимых эффектов используют так называемый перекрестный или кроссовер дизайн, при котором один и тот же человек в течение первого месяца пьет кофе, а затем в течение второго месяца не потребляет его вовсе. При этом можно, например, сравнивать артериальное давление или настроение. Однако, в большинстве случаев, единственное, что остается исследователям – это искать вторую наилучшую альтернативу: использовать в качестве контроля к человеку потребляющему кофе, другого человека максимально схожего во всем остальном кроме потребления кофе. Необходимость использования другого человека в качестве контроля в исследованиях, вовлекающих людей, связано с тем, что 1) большинство эффектов от лечения или образа жизни человека наступают отсрочено, порой через многие годы, и 2) часто являются дихотомическими: жив или умер, произошел инфаркт или не произошел. Кроме того, 3) большинство воздействий на важные (истинные) конечные точки в исследованиях, такие как риск смерти, которые изучаются в настоящее время, являются не очень сильными, а чаще столь слабыми как эффекты кофе на продолжительность жизни. Следовательно, чтобы дать требуемое количество человеко-лет воздействия, они требуют огромных выборок и длительного времени. Лишь на выборках, включающих тысячи человек, удастся обнаружить не очень большие, отсрочено наступающие эффекты многих современных лекарств. Время, когда ученые случайно натыкались в лабораториях

на чудодейственные эффекты пенициллина, к сожалению, проходит. Теперь для обнаружения препарата, который давал бы не столь ярко выраженный эффект при наиболее распространенных в 21 веке хронических неинфекционных заболеваниях, требуются огромные усилия.

Противофактическая теория диктует критерии того, исследования каких дизайнов могут считаться более или менее надежными. В реальности, все люди разные по неограниченно большому числу параметров, поэтому приходится использовать множество людей. Пользуясь «вторым наилучшим» вариантом контроля в аналитическом исследовании, ставящем целью не просто описать явление, а обнаружить причинно-следственные связи, было бы исследование всего человечества. Всех людей необходимо было бы случайным образом поделить на две равные части, одна из которых потребляла бы кофе в значительном количестве в приказном порядке, а вторая потребляла бы что то, что по вкусу, виду и запаху не отличалось бы от кофе, но при этом бы не являлось кофе.

Однако, в силу ряда существенных ограничений, эпидемиологам приходится использовать более реалистичные дизайны исследований, которые и описаны в данной работе. Эти исследования основаны на противофактической теории: ***хотя мы не можем знать, что было бы с одним и тем же человеком, если бы он пил или не пил кофе в течение всей жизни, мы можем выделить достаточно большую и репрезентативную выборку типичных потребителей кофе и к ней создать контрольную группу.***

Затем, сделав поправку на другие известные факторы риска, такие например, как курение и лишний вес, сравнить исходы, в частности ожидаемую продолжительность жизни, риск инфаркта или рака желудка. Именно таким образом, эпидемиологи подменяют наилучший контроль, которым для каждого индивидуума является тот же самый индивид.

Существует множество классификаций видов исследований. Разделяют описательные и аналитические исследования. Биомедицинские исследования можно поделить на наблюдательные и экспериментальные, качественные и количественные, а также на лабораторные, на животных и на людях. Существуют социологические и эпидемиологические исследования. В современной классической эпидемиологии иерархия основных эпидемиологических исследований распределена следующим образом: описание случая (клинического), серии случаев, одномоментное, случай-контроль, проспективное когортное (или панельное), рандомизированное контролируемое испытание. Наконец, доказательная медицина получила свой базовый инструментарий в результате разработки методик обобщающих исследований: систематических обзоров. Далее, мы коснемся основных дизайнов биомедицинской науки. Мы также опишем качественные или социологические методы исследований и клинко-экономические методы. Социологические методы исследования преимущественно классифицируются в зависимости от способа сбора данных, в то время как экономические разделяются в зависимости от того, с чем сравниваются затраты.

Так среди социологических исследований выделяют структурированные, полуструктурированные и глубинные интервью, отличающиеся по степени предопределенности последовательности задаваемых вопросов, а также, фокус-группы, антропологические исследования, в которых используется наблюдение. Данные методы используются в основном социологами медицины и не отличаются от методов, описанных в книгах по социологии.

Для начала разберемся с несколькими простыми базовыми подходами к классификации исследований. Исследования могут быть описательными (дескриптивными) или аналитическими. Дескриптивные исследования описывают распространенность, детали симптоматики заболевания, в то время как аналитические исследования изучают причинно-следственные связи. Например, изучение распространенности ВИЧ инфекции среди какой-либо группы риска является описательным, а вот исследование факторов риска заражения – это аналитическое исследование. Аналитическое исследование может быть обсервационным или экспериментальным. В частности, вышеприведенный пример аналитического исследования, ориентированный на изучения факторов риска заражения в группе риска является обсервационным, а вот изучение различных подходов к снижению риска передачи ВИЧ в ситуации, когда данные вмешательства проводятся в ходе исследования, является экспериментальным. В ряду случаев исследователи могут пользоваться «естественными экспериментами», а также квази-экспериментальными дизайном. Такие ситуации возникают, когда вмешательство производится не как часть исследования, порой без согласования с исследователями, однако, у них есть возможность провести сравнение с контрольной группой, где вмешательство не проводилось, или хотя бы динамику положения дел до и после¹.

Помимо этого исследования могут быть количественными, то есть отвечать на вопрос «сколько?» или «на сколько?», а могут быть качественными или социологическими и отвечать на вопрос «почему?» или «зачем?». Необходимо отметить, что социологические исследования недостаточно используются в биомедицине, хотя порой они дают ответы на наиболее важные вопросы. Многие исследователи собирают гигантские количество данных, пытаясь уточнить количественный показатель с точностью до долей процента, в то время как природа явления остается неясной. Например, крупное исследование международных агентств показало, что в Таджикистане 22,35% родов происходят на дому. В то же время, для чиновников Минздрава и международных организаций подобная точность в определении подобного показателя может быть важна только для мониторинга влияния реформ на его динамику. И даже для этих целей, возможно, такая точность не требуется. В тоже время, не до конца понятно, с чем именно связано то, что роды проходят на дому. Обычные количественные исследовательские подходы не обязательно являются наилучшими для определения ответа на вопрос «почему?». Описание историй, изучение комплекса факторов социологическими методами, часто

¹ Идеальным в любых экспериментах является сравнение «разницы-разниц», то есть изменения в группе вмешательства сравниваются с изменением в тот же период и, желательно, при наличии «поддельного» вмешательства в контрольной группе.

дает гораздо более точный и важный ответ на вопрос «почему?», чем, например данные о том, что в 70% случаев важную роль в решении рожать дома играли финансовые проблемы. Социологические исследования могли бы дать понимание того, что действует комплекс причин: врачи вымогают деньги за проведение родов и ведут себя крайне грубо, родственники со стороны мужа не считают, что рожать в стационаре приемлемо с религиозной точки зрения, потому что врач, работающий там – мужчина, в тоже время доступны недорогие и комфортные роды на дому с привлечением «Дои» – традиционной повитухи, которая вела роды у всех знакомых данной семьи. Такая комплексная картина, получить которую можно лишь в качественных исследованиях, позволяет разработать гипотезы в отношении возможных вмешательств, действенность которых в отношении снижения риска материнской и младенческой смертности и других негативных показателей затем стоит изучить в исследованиях экспериментального дизайна. Однако социологические исследования часто сложны с методологической точки зрения, и хотя могут быть не очень дорогими, требуют значительных трудозатрат, как на этапе сбора данных, так и на этапе анализа. Вероятно, именно в связи с этими ограничениями, они сильно недоиспользованы в биомедицинской науке, в отличие от, скажем, политологии или маркетинга.

Также существуют исследование экономической эффективности медицинских вмешательств, которые отвечают на вопросы: «за какую сумму?», «при каких ресурсных вложениях?», или «что произойдет с тем или иным показателем при инвестиции определенной суммы в данную технологию?». Эти исследования, хотя и не очень сложны, лишь изредка используются, преимущественно фарминдустрией в развитых странах, например, в качестве инструмента обоснования цены продукции. Вероятно, именно методы анализа экономической эффективности получают серьезное развитие в ближайшее время, не только как инструмент сравнения медицинской продукции, но и для оценки эффективности управленческих решений, правильности закупок и распределения средств.

Количественные методы исследования

Итак, самым простым эпидемиологическим исследованием можно считать *доклад случая*. Такие исследования в абсолютном большинстве случаев являются дескриптивными. Очень часто они служат скорее целям удовлетворения любопытства, нежели приносят научную пользу. Однако, именно с них обычно начинается привлечение общего внимания, сначала медиков, а затем и населения в целом, к какой либо проблеме. Так, возможно, мы бы не знали о существовании атипичной пневмонии, СПИДа, лихорадки Эбола, побочных эффектов лекарств, большинства орфанных патологий, если бы ни клиницисты, обнаруживающие некие симптомокомплексы или закономерности, которые представляются странными. Публикации их описаний в научной литературе, в свою очередь, влечет письма других клиницистов, которые сталкивались с подобными случаями. Хотя чаще описание новых заболеваний сразу начинается с *доклада серии*

случаев, когда одному клиницисту попадает ряд похожих пациентов, которые не подходят под описание заболеваний, с которыми данный врач знаком по литературным источникам или с которыми он сталкивается в своей практике. В общественном здоровье и организации здравоохранения доклад случаев часто используется и в качестве дополнительного метода для подтверждения, и в качестве иллюстрации каких-либо изучаемых закономерностей. Иногда популяционные проблемы здоровья трудно изучить, не прибегая к такому дизайну. Так доклады серий случаев помогли понять причины роста смертности пациентов с сахарным диабетом на территории бывшего СССР и их связь с реформами, проводившимися в 1990-е годы в России и странах СНГ.

Исследования с одномоментным дизайном, называются *кросс-секционными* или *исследованиями среза* и т.д. Чаще всего они бывают описательными, хотя в редких случаях могут использоваться и как аналитические. При этих исследованиях вся генеральная совокупность, или же выборка из нее, исследуется на предмет наличия или отсутствия тех или иных признаков. Возможно изучение как дихотомических признаков – мужчина-женщина, наличие или отсутствие заболевания, так и категориальных – цвет глаз карий, светло-серый или зеленый, группа крови, или же шкальных – например, по степени тяжести сердечной недостаточности или стадии рака, или же непрерывных – например, веса или роста. При этом в одномоментном исследовании нас могут интересовать как распространенность одной переменной, например, доля мужчин в популяции, так и распространенность комплексной переменной, например, доля мужчин старше 65 лет, у которых есть признаки облысения, так и сопряженность ряда переменных. В случае, если нас интересует сопряженность переменных, такое одномоментное исследование является аналитическим. Например, в одномоментном исследовании можно изучать связь вероятности наличия инфекций передаваемых половым путем с количеством половых партнеров. Такие аналитические исследования, однако, в большинстве случаев будут давать лишь ориентировочные данные, так как существует много причин, почему данная связь может существовать. Банальное определение направления связи становится проблемой в одномоментных исследованиях. Даже если мы исключим влияние случайности, систематических отклонений и конфаундинга, что в одномоментных исследованиях бывает непросто, точно сказать являются ли инфекции, передающиеся половым путем, толчком к тому чтобы иметь больше половых партнеров, или же большое количество половых партнеров приводит к инфекциям, только через изучение среза популяции мы однозначно не можем. В одномоментном исследовании присутствуют еще и такие интересные систематические отклонения выборки, как отклонения дифференциального выживания, при котором защитный фактор, снижающий смертность при наступлении заболевания, но не влияющий на вероятность возникновения самого заболевания, ошибочно принимают за фактор риска возникновения заболевания. Так, гипотетически, если прием препарата, например, гормонального контрацептива, до наступления инсульта, снижает риск смерти в случае его наступления, то среди тех, кто перенес инсульт и не погиб, процент людей получавших препарат будет выше, чем среди общей аналогичной популяции. В тоже время, среди людей умерших от инсульта, которые в выборку кросс-секционного

исследования не попадут, наоборот, количество людей получавших данный препарат будет ниже, чем среди тех, кто выжил и среди тех, у кого не было инсульта.

Повторенные более одного раза исследования среза становятся когортными или панельными (см. ниже), но только в случае, если они повторяются на одной и той же выборке. Встречаются ситуации многократного повторения исследования среза, когда данные собираются каждый раз на новой выборке из одной и той же генеральной совокупности. Иногда такие исследования делаются вместо когортных в силу трудностей повторной идентификации той же когорты, например в нестабильной популяции, скажем, с высоким уровнем миграции. Иногда же это делается не в силу методологических и логистических затруднений, а в силу особенности изучаемого вопроса. Так, если нас интересует процент кесаревых сечений среди первородящих, то делать это на одной и той же выборке первородящих дважды невозможно.

Кластерные или экологические исследования – это сравнение агрегированных статистических показателей в двух и более популяциях. Такие исследования чаще всего аналитические, при этом они дешевы и просты в применении при наличии данных, в тоже время подвержены большому количеству возможных ошибок. Таким исследованием может быть сравнение уровня смертности в северных странах в зависимости от уровня потребления алкоголя.

Одна из фундаментальных проблем кластерных исследования сводится к тому, что мы не знаем, являются ли те, кто подвержены следствию, теми же людьми на которых воздействовала причина. Так, если в стране с более высоким уровнем курения наблюдается более высокий показатель смертности от рака легкого, в отсутствии исследований с более скрупулезным дизайном, мы точно не знали бы, одни ли и те же или разные люди курят и умирают от рака легкого. Тем не менее, такие исследования, как правило, просты, дешевы, и в ряде случаев в комбинации с данными других исследований с более мощным дизайном, могут давать достаточно точные ответы на крайне важные вопросы.

Исследования *случай-контроль* часто делаются, после того как доступен доклад серии случаев. При этом исследования случай-контроль являются аналитическим продолжением доклада серии случаев и всегда направлены на изучение этиологических факторов. В исследовании случай-контроль к каждому заболевшему подбирается некоторое количество контролей, которые должны во всем кроме наличия заболевания быть схожими со случаями. Часто, так как в основном количество пациентов, изучаемых в таких исследованиях, бывает невелико, берется большее количество контролей, чем случаев. Два-четыре контроля на случай могут повысить статистическую мощность исследования, в то время как вовлечение более 4 контролей на случай, как правило, не имеет дополнительного смысла, с точки зрения повышения статистической силы исследования. Далее и «случаев», и «контролей» расспрашивают с применением структурированной анкеты или же без, или проводят сбор данных каким-либо иным

методом, и сравнивают частоту наличия тех или иных потенциальных факторов риска (или защитных факторов) у представителей обеих групп. Обычно, силу ассоциации фактора риска и наличия заболевания выражают как отношение шансов.

Для простоты расчета отношения шансов можно запомнить простое правило – в стандартной четырехпольной таблице цифры по диагонали перемножаются, затем получившиеся произведения делят одно на другое. Обычно для факторов риска деления проводят таким образом, чтобы получить число больше единицы.

Исследования случай-контроль идеальны для изучения факторов риска редких и долго развивающихся заболеваний и для расследования случаев вспышек заболеваний инфекционной и неинфекционной природы, когда ответ на вопрос о том, что вызвало заболевание надо получить быстро и дешево. Так, исследовать этиологию мочекаменной болезни или инсульта в проспективном исследовании будет крайне дорого, ведь на развитие этих заболеваний уйдут многие годы. Так же, в ситуации, когда в детском саду возникла вспышка пищевых отравлений, для расследования источника и предотвращения дальнейшего распространения возбудителя попросту не существует других приемлемых дизайнов исследования.

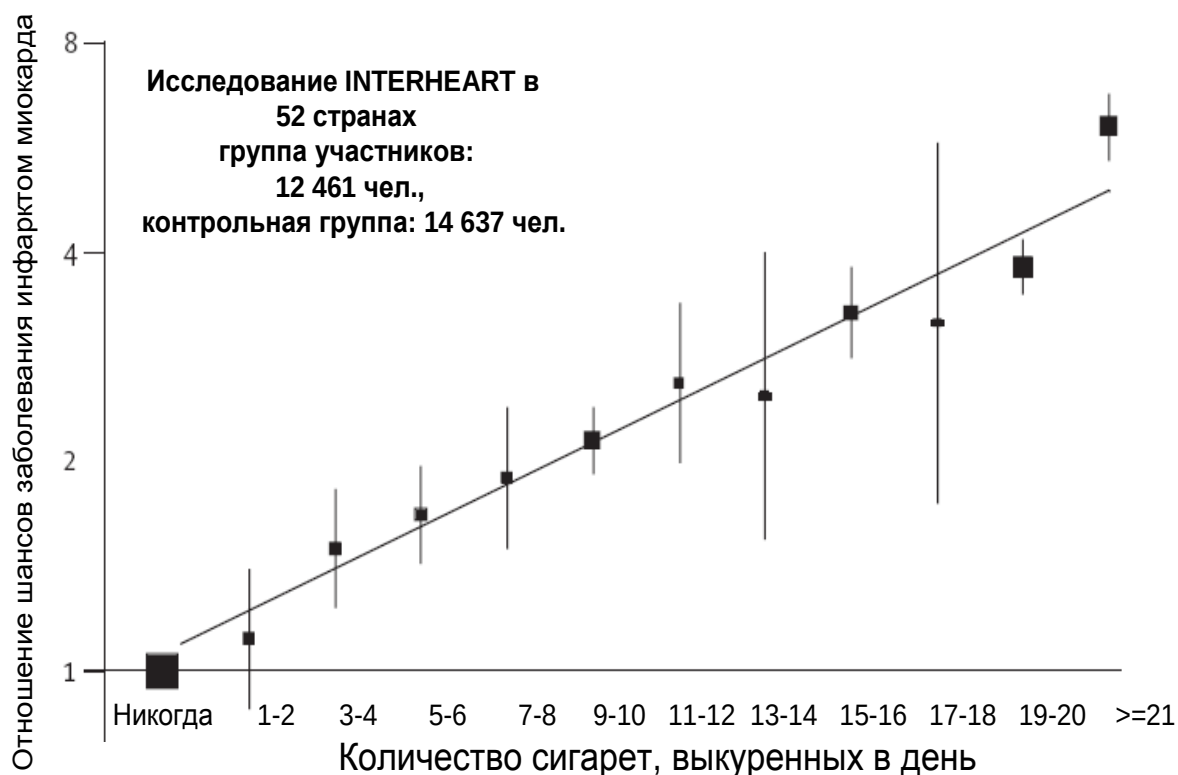
Недостатки исследования случай-контроль также многочисленны, как и достоинства. К ним относится невозможность вычисления относительного риска, и то, что чаще всего между воздействием фактора риска или защитного фактора и наступлением заболевания проходит значительное количество времени и, соответственно, появляется высокая вероятность информационных отклонений, связанных со свойствами памяти.

Исследования проспективного дизайна, которые обычно называют *когортными* или *панельными*, в основном делаются повторением одномоментных исследований на одной и той же выборке. Иногда, в качестве альтернативы используются регистры людей с определенным состоянием или подвергшихся какому-либо воздействию. В ряде случаев прибегают и к системе оповещения о возникновении нового случая заболевания в изучаемой замкнутой популяции. Когортные исследования являются наблюдательным вариантом клинических испытаний. В отличие от клинических экспериментов, изучающих эффективность лечения, диагностики или профилактики, их основной функцией, является изучение факторов с негативным воздействием на здоровье – факторов риска. Действительно, основная часть биомедицинских знаний об этиологии наиболее распространенных хронических заболеваний происходит из когортных исследований. Потенциальные факторы риска, даже в случае если данных о связанном с ними вреде для здоровья не много, этически неприемлемо изучать в экспериментах. Однако, если, например, человек уже курит и поделаться с этим ничего не получается, то можно, по крайней мере, отследить какие заболевания будут развиваться у курящих и с какой вероятностью по сравнению с группой не курящих. Подбор контрольной группы является почти столь же серьезной проблемой в проспективных исследованиях, как и в случае с исследованиями случай-контроль. Основным недостатком когортных исследований по

сравнению с рандомизированными клиническими испытаниями является невозможность контроля неизвестных конфаундеров. Так, если у курящих риск рака легкого выше примерно в 10 раз, объясняется ли эта разница тем, что курение вызывает рак, или же дело в том, что никотиновая зависимость и рак легкого определяются влиянием конфаундера, например, генетически обусловленного? Невозможность проведения исследований факторов риска в экспериментальных исследованиях в нынешнем обществе приводит к тому, что значительная часть медицинских знаний в отношении этиологии заболеваний не являются полностью «доказательными», так как не исключена вероятность того, что дело на самом деле во вмешивающихся факторах. Практически единственным вариантом более-менее полного доказательства того, что дело не в конфаундинге является изучения связи «дозы» воздействия фактора риска и размера эффекта, и изучение того, что происходит с теми, на кого фактор риска перестал действовать. Так, исследования изучавшие риск инфаркта у людей в зависимости от количества выкуриваемых сигарет, и исследования изучавшие снижение рисков заболеваний после отказа от курения показывают, что именно табакокурение является причиной целого ряда заболеваний. Хотя эпидемиологические исследования часто обнаруживают факторы вызывающие или снижающие риски заболевания, механизм действия которых неизвестен, лабораторные исследования, например, химического состава табака, во многом помогли понять, что он является ядом. В то же время, строго говоря, объяснение механизмов действия лекарств не является обязательным, важно лишь доказательство того, что они приносят пользу пациентам. Для факторов же риска понимание механизмов действия часто является желательным, так как может способствовать разработке методов противодействия их пагубному влиянию.

Существуют примеры крупных когортных исследований со сложным дизайном, в которых изучается воздействие целого ряда факторов на более чем одно заболевание. При этом в исследовании заранее не выделяются две отдельные когорты людей подверженных и не подверженных фактору риска, а собирается предельно полный массив информации обо всех поведенческих и иных особенностях людей, включенных в выборку. Это иногда позволяет исследователям впоследствии изучать новые гипотезы о причинах заболеваний, возникшие после начала исследования. В подобных исследованиях порой участвует население целых городов (например знаменитое Фрамингемское исследование), или же иные крупные когорты, например, 50 тысяч британских врачей. Сложным методологическим вопросом является определение степени репрезентативности богатого американского городка Фрамингем, хотя бы для населения США, и степени репрезентативности исследования проводимого на британских врачах хотя бы для представителей других профессий в Великобритании. Однако почти все, что мы знаем о вредоносных, и даже о многих протективных популяционных факторах, происходит именно из таких исследований.

Рис. 1. Связь количества выкуриваемых сигарет в день с отношением шансов инфаркта миокарда.



Источник: K. K. Teo et al Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study
Lancet 368 (9536):647-658, 2006 (К.К. Тео и др., Потребление табака и риск заболевания инфарктом миокарда в 52 странах, Исследование INTERHEART)

1

Одним из примеров того, какими серьезными могут быть ошибки в когортных исследованиях, является история гормонозаместительной терапии применяемой при менопаузе. В когортных исследованиях женщины принимавшие эстроген после наступления менопаузы имели более низкий риск ряда заболеваний, в частности, инфаркта. Однако, в исследовании с экспериментальным дизайном, когда женщины были рандомизированы – случайным образом разделены между группой получавшей и не получавшей гормонозаместительную терапию, положительные эффекты улетучились [2]. В когортных исследованиях, положительная связь эстрогенов с показателями здоровья была, вероятно, связана с тем, что более богатые и заботившиеся о своем здоровье женщины принимали терапию. В тоже время, они делали множество иных действий для поддержания своего здоровья, не все из которых были учтены, что в конечном итоге и давало положительные эффекты в обсервационных исследованиях. Как указано выше, исключить влияние неизвестных конфаундеров в когортных исследованиях мы не можем. Порой и известные вмешивающиеся факторы можно проконтролировать лишь частично. С этой задачей прекрасно справляется рандомизация в клинических исследованиях.

Еще одним «гибридным» дизайном является *гнездный случай-контроль*. В таких исследованиях случаи и контроли отбираются в рамках когортного исследования. Например, если в начале когортного исследования был произведен забор образцов крови

у всех людей, включенных в исследование, то затем можно сделать анализ крови тех людей, у которых возникло какое-либо состояние и у подобранных в рамках той же выборки контролей, у которых это состояние не возникло. При таком дизайне, с одной стороны экономятся средства, требуемые для анализа крови, а с другой стороны обеспечивается фактически проспективный дизайн, позволяющий избежать ошибок, связанных с тем, что в стандартных исследованиях случай-контроль данные собираются уже после возникновения заболевания.

Рандомизированные клинические испытания (РКИ) являются идеальным способом доказать причинно-следственную связь. Лучше чем РКИ для доказательства причинно-следственной связи в биомедицине, подходят лишь множество РКИ, повторенных разными исследователями на многих отличающихся выборках, в разное время, при различных условиях. В подобных исследованиях этически приемлемо изучать только воздействия факторов, которые как предполагается, приносят пользу пациенту. Тем не менее, иногда РКИ иногда дают возможность окончательно подтвердить и информацию об этиологии заболеваний. Так, РКИ, изучавшие эффективность противовирусных препаратов, позволили поставить точку в спорах о том, является ли вирус иммунодефицита человека причиной СПИДа или же оппортунистической инфекцией, возникающей на фоне снижения иммунитета.

В рандомизированном клиническом исследовании группа людей, как правило, с определенным патологическим состоянием или наличествующим фактором риска, (за исключением исследований методов первичной профилактики), случайным образом делится на группу вмешательства и контрольную группу. Контрольная группа при этом получает плацебо – «подделку» вмешательства. Так в случае, если группа вмешательства получает витамин С, контрольная группа должна получать таблетки такого же вкуса, цвета и запаха, но без витамина С. В РКИ с ослеплением пациенты, врачи, исследователи, в том числе те, кто проводят анализ данных, не должны знать какая из групп получает препарат, а какая плацебо. Лишь после того, как исследование остановлено и проведен анализ данных, протокол, в котором содержится информация о том, какая из групп получала изучаемый препарат, должен быть вскрыт. Исследование прекращается, если набрано запланированное количество больных. Также, исследование может быть остановлено досрочно из-за явного расхождения двух групп по изучаемому исходу или, например, по уровню смертности. Размер требуемой выборки рассчитывается так, чтобы получить статистически значимый результат для определения эффекта вмешательства определенной предполагаемой силы.

Правильно построенное РКИ позволяют контролировать практически все виды ошибок. Так достаточная выборка делает вероятность того, что результат случаен, ничтожно малым. Рандомизация позволяет контролировать некоторые отклонения выборки и полностью исключить влияние конфаундеров, даже неизвестных. Правильно сформулированные критерии включения людей в исследования позволяют избежать другой части ошибок выборки, которые нельзя поправить рандомизацией. Ослепление

позволяет контролировать ошибки информации. Проспективный дизайн экспериментов позволяет безошибочно определить направление связи, то есть выявить, что лекарство является причиной, а выздоровление следствием. Тем не менее, часть ошибок выборки полностью не исключается. Так, большинство РКИ в мире делаются на относительно молодых больных без сопутствующих заболеваний, а потребляют их люди старшего возраста, как правило, страдающие целым букетом заболеваний.

Ошибки, которые РКИ не в состоянии контролировать – это отклонения связанные с человеческим фактором. Несмотря на то, что РКИ представляют наилучший дизайн для контроля всех возможных ошибок исследования, даже они зачастую дают противоречивые результаты. Иногда расхождение данных РКИ объясняется вмешательством спонсоров исследований, подделкой данных самим исследователем, нарушениями протокола или изменениями протокола для получения «нужного» результата и т.д. Предварительная регистрация протоколов исследования, с размещением их в открытом доступе в сети Интернет позволяет частично контролировать и человеческий фактор, так как исследователи все же считают менее приемлемым идти на фальсификацию данных, по сравнению «с небольшими изменениями в подходах к проведению эксперимента». Небольшие «трюки» с данными, такие как исключение отдельных «странных», «случайных» результатов или выбор «подходящего» метода расчета окончательных результатов, считаются порой не просто вполне приемлемыми, но даже признаком профессионализма эпидемиолога.

Именно в связи с тем, что даже РКИ не до конца точны и были разработаны методы доказательной медицины, такие как *систематические обзоры*, частным случаем которых является *мета-анализ*. Систематические обзоры – это исследования исследований. В них по заранее оговоренным критериям из всех обнаруженных опубликованных и неопубликованных исследований на заданную тематику, проводится отбор наилучших, с наиболее сильным дизайном, наиболее скрупулезно выполненных. Затем результаты всех прошедших отбор исследований обобщаются, чтобы дать однозначный ответ на клинический вопрос. Как правило, систематические обзоры касаются вопросов о наилучших клинических тактиках лечения какого-либо состояния. В мета-анализе первичные данные нескольких исследований объединяются так, как если бы это было одно крупное исследование.

В основу иерархии доказательств ложится представление о разных дизайнах исследования. В иерархии наиболее надежными являются систематические обзоры или данные нескольких крупных, многоцентровых РКИ, а самыми слабыми данными являются экспертное мнение и личный опыт. Это не означает тем не менее, что опыт стал менее важен для клинициста, чем был раньше. Дело лишь в том, что опыт не в состоянии дать однозначный достоверный ответ на клинически значимые вопросы, так как подвержен максимальному количеству ошибок исследования.

Иерархия научных данных

I	1. Систематический обзор нескольких испытаний, проводившихся с контролем, рандомизацией и двойным слепым методом. 2. Одно или несколько крупных испытаний, с контролем, рандомизацией и двойным слепым методом.
II	1. Одно или несколько хорошо проведенных когортных исследований. 2. Одно или несколько хорошо проведенных исследований по типу случай-контроль. 3. Эксперимент без контроля, но с очень четкими результатами
III	Обзор, которым занималась комиссия специалистов; учитывалось мнение ведущего специалиста
IV	Личный опыт

Литература

1. Weinberg, Bennett Alan; Bealer, Bonnie K. (2001). The world of caffeine: the science and culture of the world's most popular drug. Psychology Press. pp. 92–3.
2. Writing Group for the Women's Health Initiative Investigators (2002). "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial". JAMA 288 (3): 321–333.

Types of studies in evidence-based medicine

Danishvskiy K. D.

MD, Professor, Higher School of Economics, Moscow

The article describes different study designs used today in medical science, the hierarchy of their reliability is considered, the validity of the obtained data is compared.

Key words: study design, case, case series, prospective cohort study, randomized study, systematic review.

References

1. Weinberg, Bennett Alan; Bealer, Bonnie K. (2001). The world of caffeine: the science and culture of the world's most popular drug. Psychology Press. pp. 92–3.
2. Writing Group for the Women's Health Initiative Investigators (2002). "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial". JAMA 288 (3): 321–333.