

Рандомизированные контролируемые испытания (РКИ)

Биомедицинские исследования данной разновидности являются *экспериментальными* (т.е. в процессе исследования производится активное вмешательство экспериментаторов в ход изучаемого процесса), *проспективными* и, обычно, *продольными*. Чаще всего они используются для сравнения эффективности различных способов оказания медицинской помощи (схем терапии, методов ухода за больными) либо медицинских технологий (сравнение эффективности фармацевтических препаратов, оперативных вмешательств либо методов диагностики).

Особенности:

1. Исследования являются сравнительными – производится сравнение двух и более лечебных подходов между собой;
2. Для доказательства того, что улучшение состояния больных связано именно с проводимой терапией, а не с какими-либо неучтенными факторами, вводится контрольная группа;
3. Контрольная группа может состоять из нелеченых лиц (получающих плацебо);
4. Распределение больных по группам осуществляется путем т.н. рандомизации. Рандомизация – наиболее важная особенность РКИ, благодаря ее наличию РКИ являются «золотым стандартом» экспериментальных биомедицинских исследований;
5. На момент начала исследования все испытуемые группы должны быть очень похожими друг на друга по величинам всех показателей, могущих повлиять на результат; благодаря рандомизации любые межгрупповые различия могут объясняться только случайностью (но не систематической)

Ограничения РКИ:

1. Данный дизайн пригоден только для экспериментальных исследований;
2. Принципы ведения больных, включенных в РКИ (объем обследований, частота осмотров) могут существенно отличаться в лучшую сторону от принятых для больных с данной нозологией, что может искусственно увеличить приверженность лечению;
3. Выборка больных, включенных в РКИ, может оказаться нерепрезентативной, поскольку участники отбираются в исследование не случайно, а в соответствии с т.н. «критериями включения» (inclusion criteria) и наличием информированного согласия, и, соответственно, их показатели могут существенно отличаться от средних в генеральной совокупности;
4. РКИ стоят дороже всех прочих типов биомедицинских исследований, поэтому их продолжительность ограничена (обычно = 48/96 неделями) и в рамках одного испытания

5. Ввиду вынужденной кратковременности РКИ в качестве регистрируемых исходов («конечных точек», endpoints) исследования обычно используются суррогатные маркеры (какие-либо лабораторные и/или инструментальные показатели). Более важные для понимания результатов изучаемого вмешательства «истинные» исходы за время проведения РКИ обычно не успевают развиться;

6. В случае, если исследуемое вмешательство существенно эффективнее того, с которым сравнивается (или наоборот), неэтично отказывать части больных в назначении более эффективного лечения. Кроме того, многие больные могут не согласиться отдать назначение своей терапии на волю случая;

7. РКИ малопригодны для исследования эффективности вмешательств при редких заболеваниях и состояниях, поскольку для этого необходимы значительные по объему исследуемые группы и продолжительный период наблюдения; для указанных исследований существует

Наиболее важным этапом РКИ является процедура *рандомизации*. Данная процедура позволяет добиться:

1. Сравнимые группы имеют одинаковый размер;
2. Каждая из сравниваемых групп является случайной выборкой из генеральной совокупности лиц с данной нозологией/состоянием;
3. На момент начала исследования все испытуемые группы имеют одинаковые (или почти одинаковые) характеристики;
4. Результаты РКИ можно распространить на всех индивидуумов с тем же состоянием;
5. Сравнение исходов вмешательства в исследуемых группах свободно от систематических ошибок.

В качестве примера систематических ошибок, могущих иметь место при отсутствии рандомизации, можно привести исследование, проведенное в Мозамбике в 1996 г: анализировались исходы кесарева сечения, выполненного опытными акушерами-гинекологами (113 случаев), в сравнении с исходами аналогичного вмешательства, проведенного фельдшерами (958 случаев). В итоге, вопреки здравому смыслу, никаких различий в исходах операций не было выявлено. В данном случае в группу беременных, которым кесарево сечение выполняли фельдшеры, попали в основном женщины в удовлетворительном состоянии, при отсутствии тяжелых сопутствующих заболеваний, могущих привести к неблагоприятному исходу вмешательства. В случае наличия таковых операцию выполняли врачи. В итоге оказалось, что беременных с благоприятным прогнозом вмешательства оперировали фельдшера, а с неблагоприятным – врачи, и разница в квалификации нивелировалась тяжестью состояния рожениц.

Проводится т.н. маскирование («ослепление») терапевтических вмешательств:

Субъекты исследования не знают, какое именно вмешательство будет им произведено (что исключает «эффект плацебо» либо давление на медицинский персонал с целью получения определенной схемы терапии); то же может касаться и лечащих врачей («слепое исследование» и «двойное слепое исследование», соответственно). Бывает «тройное слепое исследование», когда даже фармацевт, выдающий препараты, не знает, что именно он выдает. В идеале и статистик, обрабатывающий результаты РКИ, не знает расшифровки рандомизационных кодов, что полностью исключает субъективный подход при анализе результатов – субъективные симпатии больных и/или медперсонала к какому-либо препарату не могут повлиять на учет его эффективности в сравнении с конкурирующей схемой лечения, и только организаторы исследования, владеющие рандомизационными кодами, могут располагать истинными

Тот факт, что лечение назначается больным *случайным образом*, позволяет производить сравнение терапевтических схем, **не опасаясь систематических ошибок**. Именно поэтому РКИ обладают большей доказательностью по сравнению с прочими типами дизайна биомедицинских исследований, а полученные в процессе выполнения РКИ результаты имеют больший шанс повлиять на общепринятые терапевтические подходы и практику здравоохранения.

При проведении РКИ **встает ряд проблем этического толка**. Так, новый способ терапии может оказаться как лучше, так и хуже предшествующих. В обоих случаях больные либо в контрольной, либо в опытной группе будут получать не самую лучшую из возможных схем лечения, что неминуемо отразится на динамике и прогнозе заболевания. В этой связи РКИ обычно предусматривают проведение промежуточных анализов результатов исследования через заранее определенные интервалы времени; в случае, если по данным промежуточных анализов один из способов лечения окажется достоверно более эффективным, чем другие, исследование должно быть досрочно завершено, а все больные – переведены на наиболее эффективную схему терапии. При этом выполнение этапов РКИ должно контролироваться комитетом по этике того учреждения, на базе которого проводится исследование.

РКИ включает в себя ряд последовательных этапов:

1. Разработка протокола исследования

Протокол исследования – официальный документ, в деталях описывающий цели, задачи и методы исследования. Обычно включает в себя *следующие разделы*:

- обоснование, гипотеза и цели исследования;
- методы исследования (дизайн испытания, критерии включения/исключения, планируемые вмешательства, метод рандомизации участников, сокрытия результатов рандомизации, первичные и вторичные «конечные точки» исследования, способ определения размера исследуемых групп);
- анализ результатов исследования (планируемые сравнения, используемые статистические методы);
- процедуры исследования (получение информированного согласия участника, регистрация участников, процедура рандомизации, схема наблюдения за участниками);
- форма одобрения этического комитета, план наблюдения за ходом исследования, формы сбора данных и опросные листы, решение административных вопросов).

2. Выбор конкретной разновидности дизайна РКИ

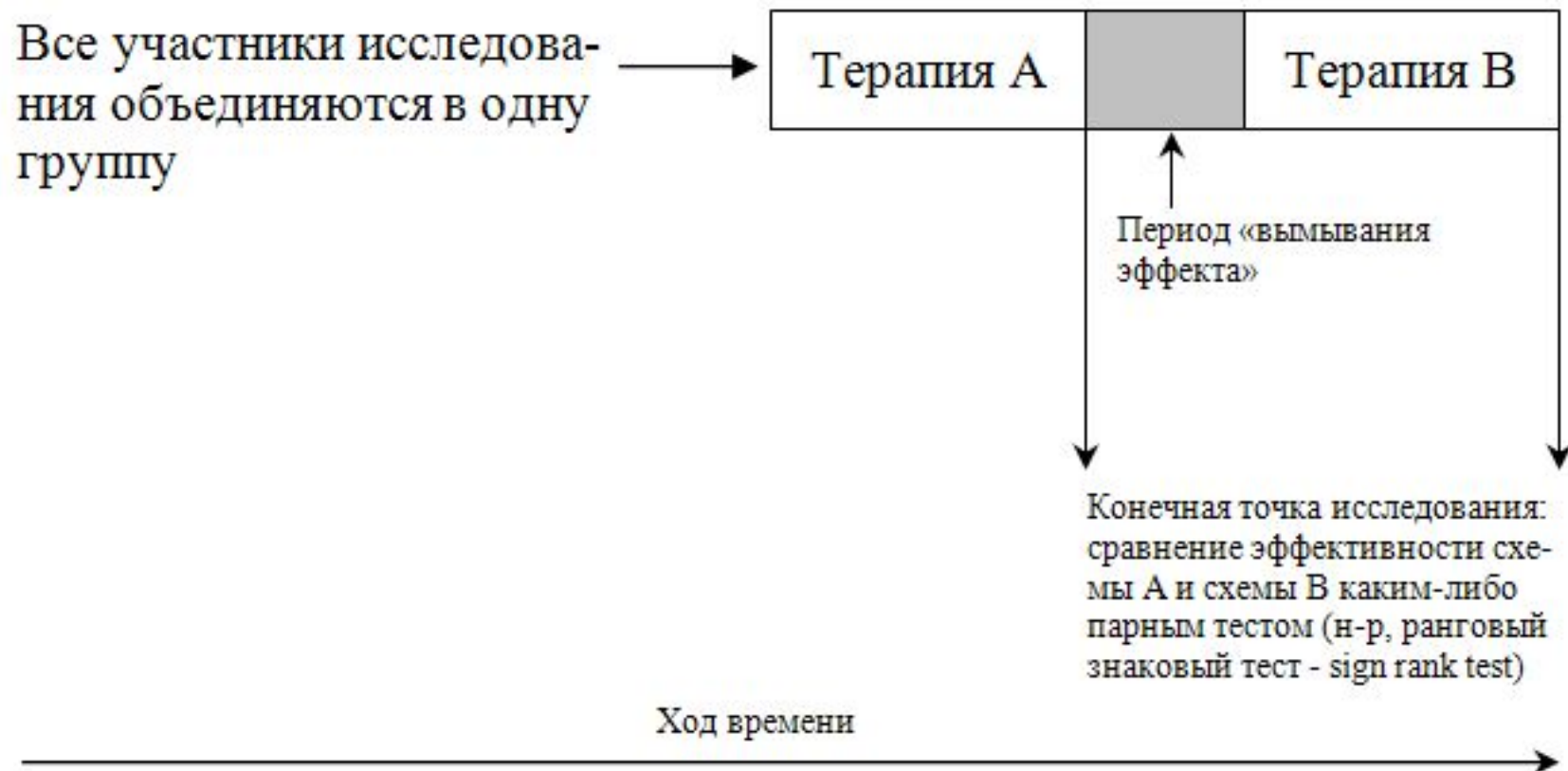
А) Параллельный дизайн

В этом случае все участники исследования рандомизируются либо в группу А, либо в группу В, при этом в каждой из групп испытывается только один метод терапии. Соответственно, каждый отдельно взятый больной получает только одну схему лечения. Такой подход сокращает сроки исследования, но требует набора большого количества добровольцев для достижения однородности сравниваемых групп; кроме того, возможна этическая проблема, связанная с тем, что один из сравниваемых способов лечения может оказаться более эффективным или безопасным, чем другой.



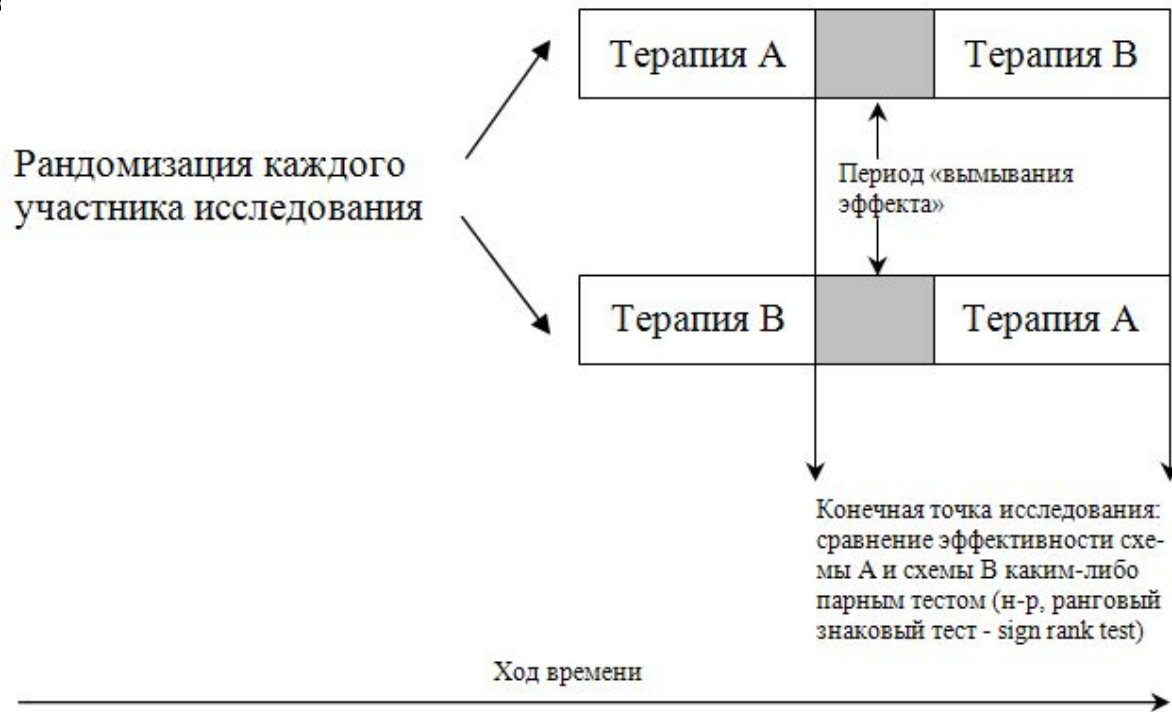
Б) Последовательный дизайн

Данная схема подразумевает назначение одной из сравниваемых схем терапии всем добровольцам, включенным в РКИ, а затем, по истечении заранее оговоренного «отмывочного периода» – второй, третьей схемы терапии и т.д. Сроки исследования при этом удлиняются, но зато уменьшается требуемое количество участников исследования. Этическая проблема в данном случае состоит в том, что во время «отмывочного периода» все больные-участники РКИ не получают никакого лечения.



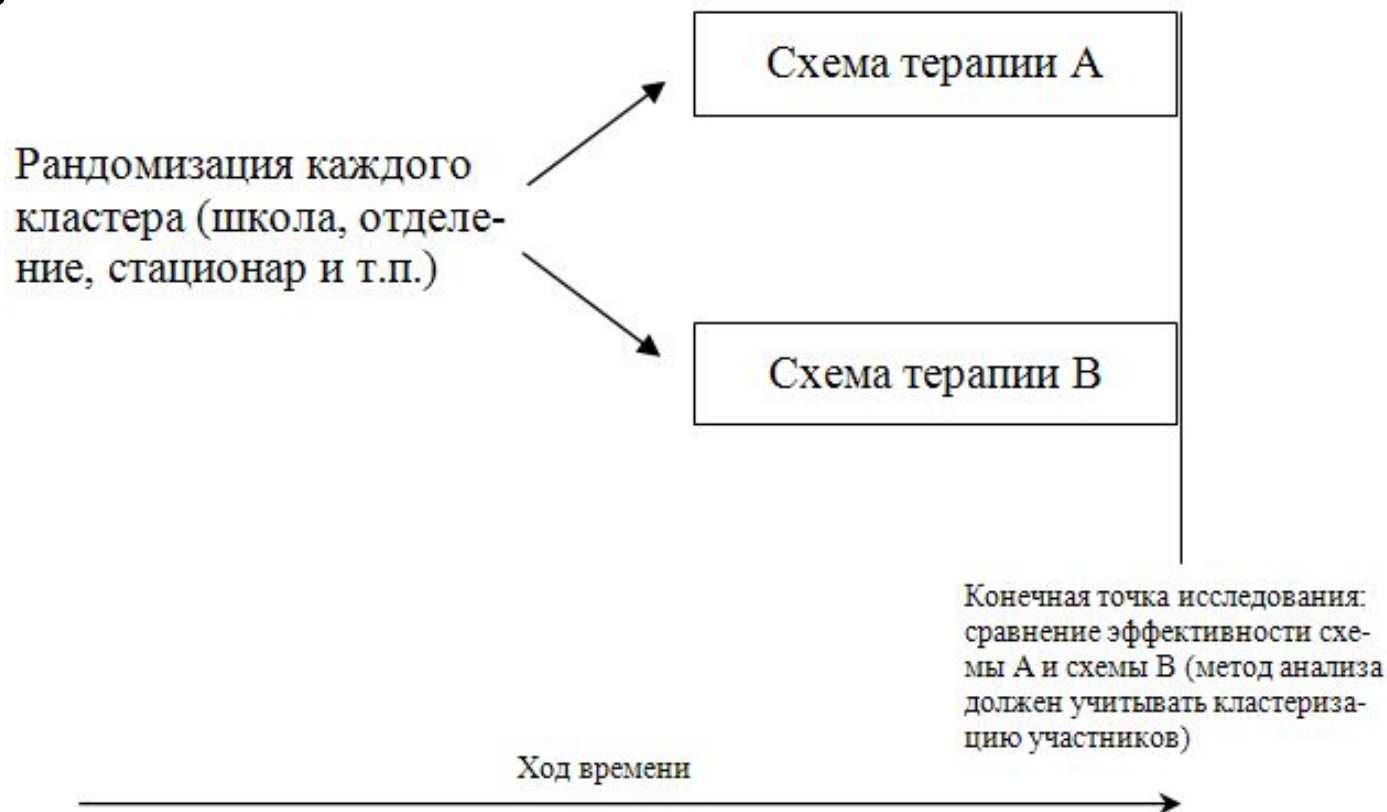
В) Перекрестный дизайн

Каждый участник исследования получает вначале одну схему терапии, а затем, после некоторого промежутка времени – другую, причем порядок назначения схем лечения рандомизируется для каждого отдельно взятого больного. В этом случае каждый участник исследования является «контролем» сам для себя. Данный тип дизайна РКИ требует меньших размеров испытуемых групп, но пригоден лишь для оценки краткосрочных эффектов терапии у больных хроническими заболеваниями. При этом указанный дизайн сохраняет недостатки последовательного: во время отмывочного периода участники РКИ не получают никакого лечения. В то же время устранена возможная неэквивалентность схем терапии, поскольку одни и те же больные по очереди подвергаются всем испытуемым вмешательствам



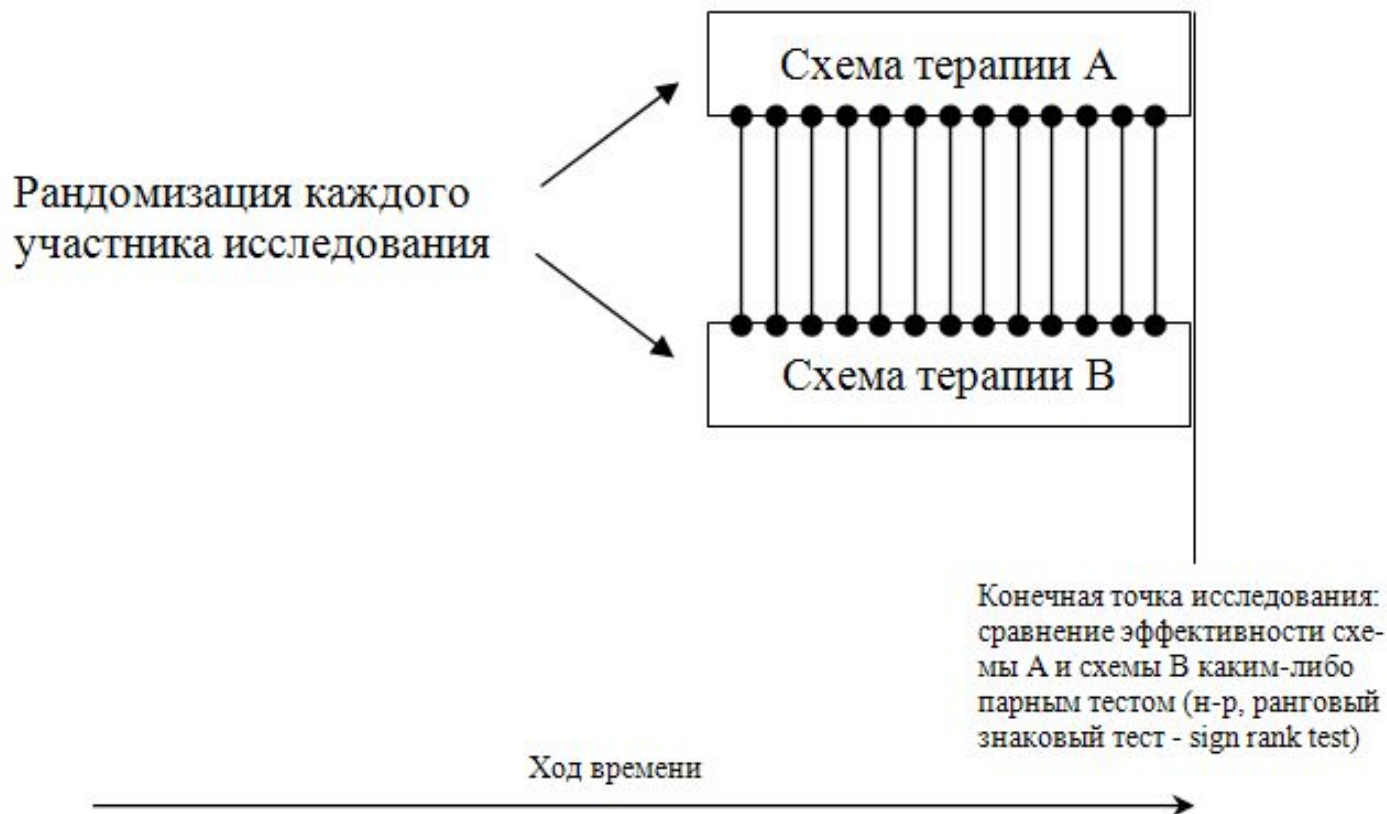
Г) Кластерный дизайн

В этом случае дизайн исследования напоминает РКИ с параллельными группами, но рандомизации в исследуемые группы подвергаются не отдельные лица, а т.н. «кластеры» участников (например, все больные, госпитализированные в определенное отделение стационара, на базе которого проводится РКИ, будут включены в состав одного кластера; в другой кластер могут быть включены все больные, обслуживаемые семейным врачом – участником РКИ). При этом каждый отдельный кластер (все его участники) получает только одну из возможных схем терапии. Данный тип дизайна используется при проведении больших многоцентровых РКИ



Д) Парный дизайн

Данный тип дизайна подразумевает рандомизацию больных в несколько групп, получающих различные схемы терапии (как в случае параллельного дизайна), но учет результатов производится строго в парах пациентов; при этом можно выявить межиндивидуальную вариабельность эффекта лечения.



Е) Адаптивный дизайн

При использовании адаптивного дизайна сравниваемые группы формируются по сложным алгоритмам, исходя из прогнозируемой и текущей эффективности исследуемых воздействий; формирование групп происходит таким образом, чтобы в группе, где применяют наиболее эффективное лечение, оказалось больше добровольцев, чем в остальных. При этом отчасти решается этическая проблема неэквивалентных по эффективности схем терапии, и, кроме того, такая схема позволяет уменьшить влияние межиндивидуальной вариабельности на результат исследования.



Ж) Факторный дизайн («латинский квадрат»)

В этом случае количество формируемых групп соответствует количеству ячеек латинского квадрата, построенного с использованием сравниваемых терапевтических схем. Каждая ячейка латинского квадрата (т.н. блок) образуется на пересечении двух символов; в результате общее количество блоков соответствует полному набору возможных комбинаций всех сравниваемых опций (т.е. схем терапии и контрольных групп). Метод позволяет учесть воздействие каждого из учитываемых факторов (в том числе – плацебо-эффекта) на исход лечения. Оценку результатов исследования с факторным дизайном выполняют с использованием метода многомерного анализа (ANOVA)

Рандомизация каждого участника исследования



	Лечение А	Плацебо (К)
Лечение В	АВ	ВК
Плацебо (К)	АК	КК

Конечная точка исследования: сравнение эффективности схемы А и схемы В при помощи дисперсионного анализа

Ход времени



3) *Дизайн Зелена*

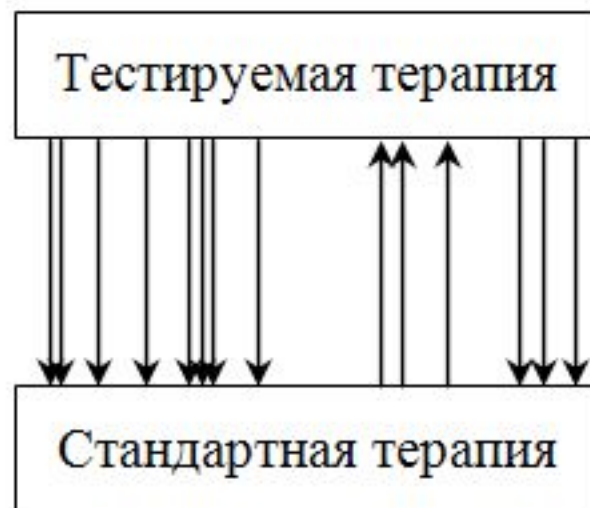
Данная разновидность дизайна была предложена статистиком Мартином Зеленом. Суть ее состоит в том, что рандомизацию больных в сравниваемые группы осуществляют до подписания ими информированного согласия. При этом те больные, которые по результатам рандомизации должны получать стандартное, общепринятое лечение, вообще не информируются о том, что они участвуют в клиническом испытании. В то же время больных, распределенных в группу с экспериментальным вмешательством, извещают о том, что они участвуют в РКИ; им также сообщают и о существовании стандартной схемы лечения их заболевания и дают возможность перейти в группу, ее получающую. Если больные, вопреки результатам рандомизации, перейдут в группу, получающую стандартное вмешательство, результаты их лечения все равно анализируются так, как если бы вмешательство было экспериментальным.

Преимуществом данного дизайна является возможность выбора, предоставляемая участникам РКИ, что обеспечивает включение в исследование максимального количества добровольцев. *Недостатки:* 1) РКИ с дизайном Зелена вынужденно являются открытыми исследованиями; 2) возможен недостаток статистической мощности исследования вследствие того, что значительная часть пациентов предпочтет стандартное вмешательство экспериментальному; 3) имеется этическая проблема, т.к. больные, получающие стандартную терапию, не информируются об их участии в РКИ. Эту проблему было предложено решать путем предоставления этим больным возможности перейти в группу, в которой применяется экспериментальное вмешательство (т.н. «дизайн с двойной рандомизацией»). Естественно, **сокрытие результатов**

Считается, что дизайн Зелена наиболее точно моделирует реальные условия применения нового метода, а также *позволяет оценить приверженность* больных назначенной терапии. Полученные результаты оцениваются с применением кластерного анализа.

Известен случай, когда дизайн Зелена был использован в РКИ для испытания нового метода экстракорпоральной мембранной оксигенации у детей с фатальными заболеваниями бронхолегочной системы только для того, чтобы не давать родителям половины детей, получающих обычную схему ИВЛ, необоснованной надежды на существование лучшего метода респираторной поддержки.

Рандомизация каждого участника исследования



Конечная точка исследования: сравнение эффективности стандартной и новой схем лечения с использованием кластерного анализа

Ход времени



3. Подбор участников исследования

Необходимо разработать четкие, объективные критерии включения/ исключения. Узкие, строгие критерии включения позволяют сфокусировать внимание на тех больных, у которых испытываемое лечение будет наиболее эффективно; при этом вариабельность исходов терапии уменьшится. В то же время, если мы хотим, чтобы полученные результаты можно было экстраполировать на генеральную совокупность лиц с данной патологией, исследуемая выборка должна быть репрезентативной. Вследствие указанного ограничения *РКИ часто критикуют за излишнюю «эсклюзивность»* (например, из рандомизированных исследований часто исключаются больные с сопутствующей патологией или не говорящие на государственном языке). Кроме того, многие больные могут вообще отказаться от рандомизации из-за личных терапевтических предпочтений.

4. Разработка метода назначения терапевтических схем (т. е. собственно рандомизация)

Различают **следующие виды рандомизации:**

1) ***Статическая*** – метод определяется до начала формирования сравниваемых групп, в дальнейшем не меняется. Этот тип распределения используется почти во всех проводимых РКИ.

Разновидности статической рандомизации:

- простая (simple), или полная (complete) рандомизация;
- перестановка блоков (permuted blocks);
- модель урны для голосования (urn model);
- асимметричная монета (biased coin).

2) ***Ковариативная адаптивная*** (covariate adaptive) – метод рандомизации изменяется *в зависимости от наблюдаемого соотношения различных существенных признаков* (ковариат) у лиц, уже вовлеченных в исследование.

Разновидности:

– минимизация

3) ***Реактивная адаптивная*** (response adaptive) – метод рандомизации изменяется *в зависимости от особенностей реакции на изучаемые воздействия* уже вовлеченных в исследование лиц.

Разновидности:

– «игра на лидера» (play the winner).

В том случае, когда наряду с получением информации о сравнительной эффективности различных методов лечения необходимо (с этических позиций) вовремя перевести больных на схему терапии, оказавшуюся наиболее эффективной, реактивная адаптивная схема (т.е. «игра на лидера») вынужденно оказывается более предпочтительной.

Простая рандомизация (также называемая полной) – рандомизация без ограничений, предполагаемых процессом распределения испытуемых лиц по группам, за исключением предварительного определения общего размера исследуемой выборки. В частности, простая рандомизация имеет место в случае, когда общий размер выборки (n) точно определен заранее, причем схему лечения А получает подгруппа, состоящая из $n/2$ случайно отобранных индивидуумов, в то время как оставшиеся $n/2$ больных получают схему лечения Б. Последующее проведение стратификации обеспечивает баланс ковариат в формируемых группах.

Существует много *разновидностей простой рандомизации*:

Бросание монетки («орел или решка»). Позволяет рандомизировать больных в две сравниваемые группы. Образно говоря, *перед назначением больного в одну из испытуемых групп исследователь подбрасывает монетку*, и если выпадает орел – то больной приписывается к группе А, а если решка – то к группе Б. Метод удовлетворительно работает при больших размерах групп, но на его эффективность могут оказывать влияние случайные факторы - в частности, результат зависит от силы и направления броска, а также от степени симметричности монеты и ее сбалансированности по весу, т.к. при наличии эксцентриситета центра масс вероятность выпадения одной из сторон при равной силы бросках будет значимо выше, чем другой (систематически действующий фактор).

Бросание игральной кости. Позволяет рандомизировать больных в четное число групп (до шести). При этом определенные выпадающие номера соотносятся с распределением испытуемых в определенную группу (например, четные номера – распределение в группу А, нечетные номера – в группу Б; или номера 1-2 – группа А, 3-4 – группа Б, 5-6 – группа В). Недостатки данного метода такие же, как и у предыдущего – он обеспечивает приблизительно равную численность групп только при их больших размерах, кроме того, возможны систематические ошибки распределения выпадающих номеров в зависимости от силы и направления бросков, а также балансировки кости. **В настоящее время методы 1 и 2 имеют историческое значение либо применяются в странах с ограниченными ресурсами.**

Алфавитный метод. Если фамилия или имя больного начинается на букву в интервале А-М, он попадает в группу А, а если в интервале Н-Я – то в группу Б; число групп и размеры интервалов могут варьироваться. Ненадежный способ рандомизации, поскольку в обществе могут существенно преобладать всего несколько десятков фамилий (самые распространенные фамилии в России на текущий момент в порядке убывания – Смирнов (1,86% населения), Иванов (1,33%), Кузнецов (1,0%), Соколов (0,86%), Попов (0,80%). Для имен это еще более справедливо – существенное преобладание всего нескольких имен на протяжении нескольких лет как дань своеобразной моде известно всем. Соответственно, избавиться от систематической ошибки при формировании групп в принципе невозможно, т.к. процесс присвоения имен испытуемым (и наследования фамилий) не находится во власти исследователей.

Рандомизация по номеру телефона / социальной страховки. Является вариантом рандомизации по таблице случайных номеров, с поправкой на то, что номера телефонов либо страховых свидетельств могут быть неслучайными (например, в некий стационар обращаются в основном жители примыкающих микрорайонов, вследствие чего их номера телефонов принадлежат всего 1-2 пулам и начинаются на одинаковые цифры). Соответственно, пользоваться данным способом рандомизации *не рекомендуется*, поскольку будет трудно избавиться от систематической ошибки в формировании сравниваемых групп.

Последовательная рандомизация. При этом больные, поступившие в течение часа (вариант – до обеда), определяются в группу А, а поступившие в течение следующего часа (соответственно, после обеда) – в группу Б и т.п. Проблема данного метода состоит в том, что время поступления больных может быть неслучайным (ночью и в ранние утренние часы экстренно поступают тяжелые больные, с утра до обеда госпитализируются пенсионеры и школьники, после обеда и до позднего вечера – работающие люди и т.д.). Вследствие этого возможны систематические ошибки при формировании исследуемых групп – в группу, набранную до обеда, попадут преимущественно дети и пожилые люди, а после обеда – рабочие и служащие средних лет. Данный метод также не рекомендуется к практическому применению.

Использование таблицы или последовательности случайных номеров (например, таблицы В13 Альтмана или последовательности псевдослучайных чисел).

Определенный пул номеров соотносится с распределением больных в определенную группу. При этом новые номера в таблице считываются в определенной исследователями очередности (скажем, слева направо сверху вниз), один за другим, и сравниваются с заранее заданными интервалами, закрепленными за испытуемыми группами; использованные номера зачеркиваются. Если используется не таблица, а генератор псевдослучайных чисел (ГПСЧ), то новые числа выдаются им последовательно, одно за другим, после чего алгоритм действий такой же (например, если ГПСЧ выдает случайные числа из интервала $0 \dots 1$, то выпадение числа в интервале $0 \dots 0,5$ может соответствовать распределению в группу А, а в интервале $0,51 \dots 1$ – в группу Б, ГПСЧ же запускается при включении в исследование каждого нового больного).

Все ГПСЧ продуцируют не случайные, а псевдослучайные числа (каждое последующее число отчасти определяется предыдущим). Важнейшей характеристикой ГПСЧ является его «период» – количество итераций, по прошествии которого генератор «заиклиивается», т.е. начинает вновь выдавать ранее сгенерированную числовую последовательность. Период разных ГПСЧ различен и составляет от 2^{32} для генератора, встроенного в MS Excel, до $2^{19937}-1$ для алгоритма «вихрь Мерсенна». Ранние ГПСЧ грешили коротким периодом (рекорд здесь принадлежит, вероятно, генератору, реализованному на ДВК Basic, который заиклиивался через 8192 итерации); короткий период повторения числовой последовательности является систематически действующим фактором, влекущим за собой ухудшение качества рандомизации. Тем не менее, в настоящее время практически все современные ГПСЧ удовлетворяют требованиям биомедицинских исследований. К числу надежных ГПСЧ относятся, помимо «вихря

Все вышеперечисленные методы рандомизации являются «почти случайными», поскольку несвободны от систематических ошибок. Разница между методами состоит в том, что в некоторых из них влияние систематических ошибок может быть сведено к *пренебрежимо малому* (например, в методе с применением таблиц или последовательностей случайных номеров – при использовании качественного ГПСЧ с надежным алгоритмом и значительной длиной псевдослучайной последовательности). В остальных случаях систематические ошибки являются *неотъемлемой, неконтролируемой и неустранимой чертой* описанных методик (неидеальная балансировка монет и игральных костей, неконтролируемо разные сила и направление бросков, преобладание в популяции некоторых имен и фамилий и т.д.), что препятствует их практическому использованию, несмотря на простоту, доступность для понимания и дешевизну.

Полная (простая) рандомизация – единственная схема, при которой распределение больных по группам полностью свободно от систематических ошибок. Каждый больной, включаемый в исследование, имеет 50% вероятность оказаться в группе А и 50% – в группе Б. Тем не менее, при наличии ограничений на количество пациентов в группах может иметь место существенный дисбаланс по этому показателю. Например, если необходимо набрать 2 группы по 5 больных в каждой, существует возможность, что в каждую из них попадет от 0 до 10 человек. Для группы в 10 человек вероятность оптимального баланса (по 5 человек в каждой из групп) составляет 24,61%. Если принять за «значительный дисбаланс» разницу в численности формируемых групп в 20% и более, то его вероятность для данной группы составит 2,16%. Дисбаланс количественного состава групп влечет за собой снижение статистической мощности исследования. Ввиду этого, полная рандомизация рекомендуется для исследований, включающих не менее (лучше – более) 200 б

Существует *вариант полной рандомизации*, называемый «**асимметричная монета**» (biased coin), который специально предназначен для ликвидации количественного дисбаланса в формируемых группах непосредственно в процессе рандомизации. В случае, если дисбаланс состава групп выходит за некоторые приемлемые для исследователей пределы, вводится поправочный коэффициент, повышающий (или понижающий) шансы больного оказаться в той или иной группе испытуемых.

Наиболее показательное применение данного метода при использовании рандомизации с помощью последовательности случайных чисел.

Предположим, что рандомизация производится в группы А и В; генератор случайных чисел выдает целые числа в диапазоне от 1 до 10, и если выпавшее число меньше или равно 5, больной попадает в группу А, иначе – в группу В. Предположим, что на каком-либо этапе рандомизации в группе А оказалось 7 человек, а в группе В – только 3. В таком случае вводится поправочный коэффициент 1,5, на который умножается выпадающее случайное число; в результате шансы больного попасть в группу А уменьшаются на 20%, и в процессе дальнейшей рандомизации количественный состав формируемых групп постепенно выравнивается. Образно говоря, мы начинаем подбрасывать *асимметричную монету* вместо «идеальной». При этом распределение участников исследования по группам остается в целом случайным, метод прост в исполнении, но малоприменим для работы с небольшими группами.

Перестановка блоков (permuted blocks) – пример исходно ограниченной схемы рандомизации. Данный метод предотвращает дисбаланс количественного состава групп посредством введения ограничительного условия – полного равенства размера образующихся при рандомизации групп сравнения. Принцип метода: **из всех вариантов распределения больных по группам, повторенных требуемое количество раз, формируется т.н. блок.**

Размер блока произволен, но общее количество вариантов в блоке должно быть кратно количеству сравниваемых вмешательств. Например, если требуется быстро распределить больных по группам А и В так, чтобы размер обеих групп все время оставался одинаковым, можно распределять в обе группы по 2 больных на каждом этапе рандомизации; при этом размер блока будет равен 4.

Существует 6 вариантов взаимного расположения «А» и «В» в таком блоке: ААВВ, АВАВ, АВВА, ВВАА, ВАВА, ВААВ.

В процессе распределения больных по группам один из таких блоков выбирается случайным образом (например, компьютером, либо экспериментатор, непосредственно осуществляющий рандомизацию, выбирает неподписанный конверт из тщательно перетасованной пачки таких же, вскрывает его и достает оттуда карточку – схему блока рандомизации). В данном случае проходят процедуру рандомизации одновременно 4 человека, причем в каждую из формируемых групп попадут 2 из них. При размере блока 2 и наличии двух альтернативных вмешательств А и В рандомизация перестановкой блоков сводится к *простой* рандомизации. Размер блока может быть и более четырех – 6, 8, 10 и т.п., в соответствии с потребностями исследователей; при этом он должен оставаться кратным количеству сравниваемых вмешательств (в данном случае 2 – А и В).

Рандомизация участников исследования последовательными блоками сохраняет баланс сравниваемых групп на протяжении всего исследования. Тем не менее, обеспечивается только баланс количественного состава групп, но не важных характеристик больных. Кроме того, данный дизайн может вносить элемент предопределенности в формирование групп *при несоблюдении условия сокрытия рандомизации из-за наличия периодически повторяющихся элементов в конце каждого блока*. Например, представим себе клиническое испытание, где производится рандомизация в 3 группы блоками по 6 больных. В этом случае, если первые 5 номеров в блоке имеют последовательность 2, 3, 1, 1, 2, то шестым с неизбежностью окажется номер 3, что делает распределение предсказуемым.

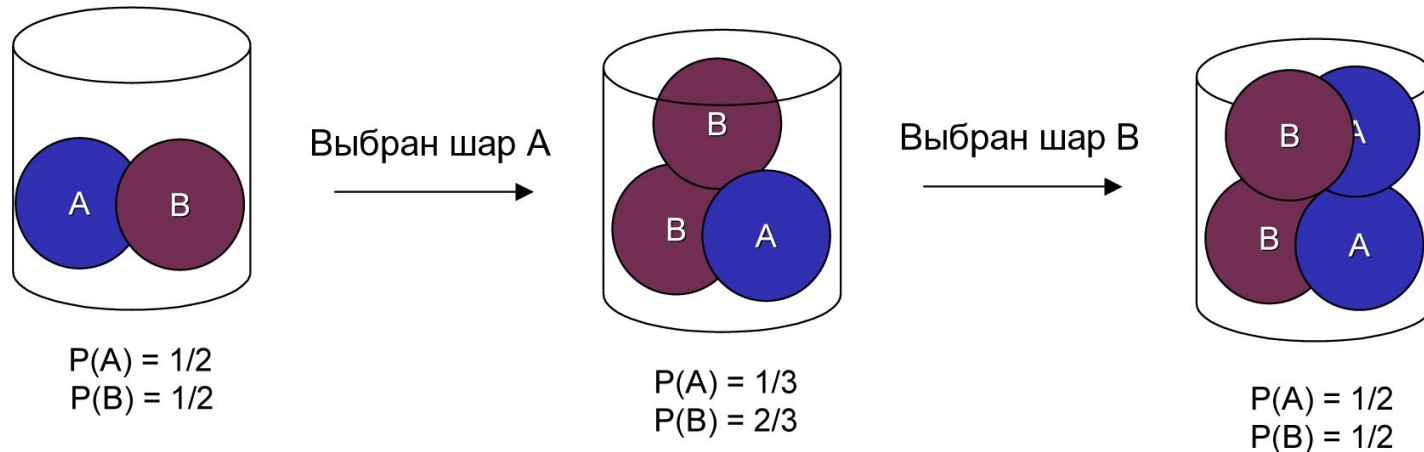
Модель избирательной урны (urn model) – одна из наиболее сложных схем рандомизации; тем не менее, при правильном применении она обеспечивает лучший баланс численного состава групп, чем полная рандомизация (особенно при работе с малыми группами), и в то же время не страдает излишней предсказуемостью, как метод перестановки блоков. Модель избирательной урны также удобна для тех случаев, когда размер формируемых групп на момент начала исследования еще не определен.

Принцип модели: исследователи используют непрозрачную емкость («урну»), реальную или компьютерную модель. В урне имеются шары различных цветов, каждый из которых соответствует одной из формируемых групп. Исходное количество шаров разного цвета в урне всегда одинаково (может быть равным нулю). В наиболее простой ситуации мы формируем две группы. В этом случае у нас в урне имеются шары двух цветов – например, красного и синего. Если исследователь вытаскивает из урны красный шар – больной определяется в группу А, если синий – в группу В. Если из урны извлечен шар красного цвета, то больной направляется в группу А, красный шар возвращается обратно в урну, после чего в урну добавляются один или несколько (всегда одинаковое, заранее оговоренное количество) шаров синего цвета.

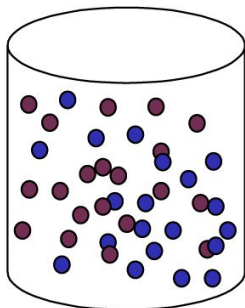
После распределения больного в одну из групп шансы следующего больного попасть в другую группу искусственно несколько повышаются (степень повышения зависит от количества добавленных шаров соответствующего цвета), что и обеспечивает существенно бóльшую вероятность баланса количественного состава формируемых групп, чем в случае простой рандомизации. Если формируется *более двух групп*, принцип не меняется – извлеченный шар определенного цвета возвращается назад, после чего в урну вносится определенное одинаковое количество шаров *всех остальных цветов*. По сути, модель избирательной урны – усложненная разновидность

Общее количество шаров *каждого цвета*, находящихся в урне на момент начала рандомизации, принято обозначать как α , а количество шаров *другого цвета* (цветов), добавляемое после каждой итерации рандомизации – как β .

Пример нестратифицированной рандомизации ($\alpha=1, \beta=1$)



Последовательная рандомизация больных в группы А и В...



Вероятность того, что будет выбран шар А ($P(A)$) равна отношению количества синих шаров к общему числу шаров в урне

Вероятность того, что будет выбран шар В ($P(B)$) равна отношению количества красных шаров к общему числу шаров в урне

В случае, когда $\alpha=0$, $\beta>0$, т.е. исходно урна шаров не содержит, первый больной распределяется в какую-либо из групп по результату броска «идеальной», т.е. симметричной, монеты. После этого в урну добавляется заданное количество шаров другого цвета (цветов), и процесс рандомизации продолжается.

Согласно модели Фридмана, вероятность, с которой больной попадает в группу А,

вычисляется по формуле:

$$\varphi(D_n, n) = \frac{1}{2} \left(1 - \frac{\beta D_n}{2\alpha + \beta n} \right)$$

где:

$\alpha, \beta \geq 0$, причем либо α , либо β обязательно больше 0;

$\varphi(D_n, n)$ – вероятность распределения в группу А;

n – предполагаемое количество больных в обеих формируемых группах;

D_n – величина дисбаланса между формируемыми группами (человек), $D_n = n(A) - n(B)$;

α – общее количество шаров каждого цвета, находящихся в урне на момент начала рандомизации;

β – количество шаров другого цвета (цветов), добавляемое после каждого шага

Для случая, когда $\alpha=0$, $\beta>0$, величина $\varphi(D_n, n)$ не зависит от величины β , а схема рандомизации может быть сведена к перестановке блоков, когда имеется только один блок с размером, увеличивающимся на одну и ту же величину при каждом шаге рандомизации. Такой вариант метода обеспечивает хороший количественный баланс формируемых групп, особенно при небольших n . В этой связи вариант метода ($\alpha=0$, $\beta>0$) предпочтительнее при рандомизации больных в небольшие группы

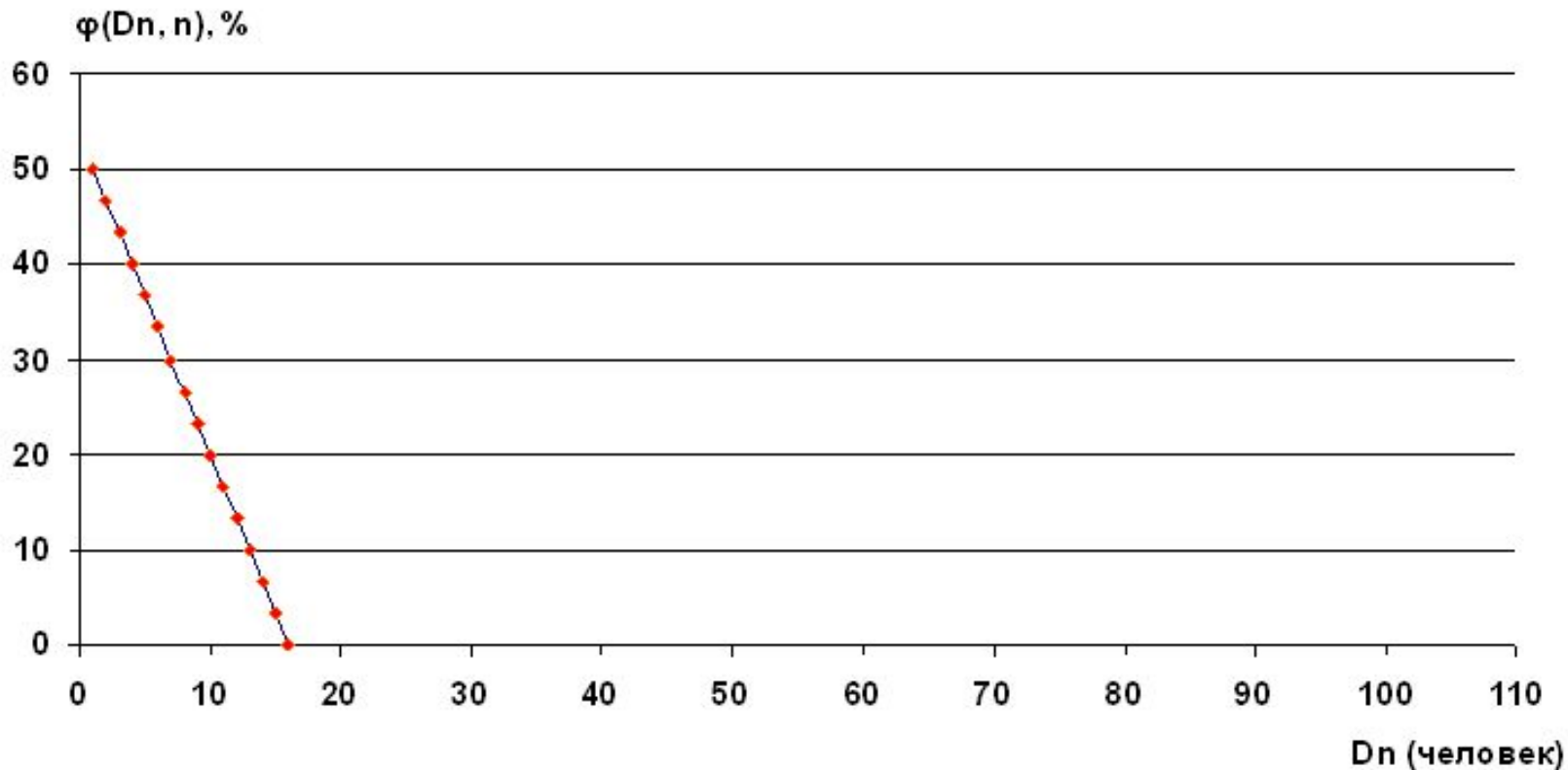


График вероятности попадания произвольного больного в группу A.
 $\alpha = 0$, $\beta = 5$, $n = 15$.

В случае, когда $\alpha > 0$, $\beta = 0$, модель избирательной урны сводится к схеме простой (полной) рандомизации, и $\varphi(D_n, n) = 0,5$ независимо от величины D_n . Та же ситуация наблюдается при $\beta > 0$, если n стремится к бесконечности.

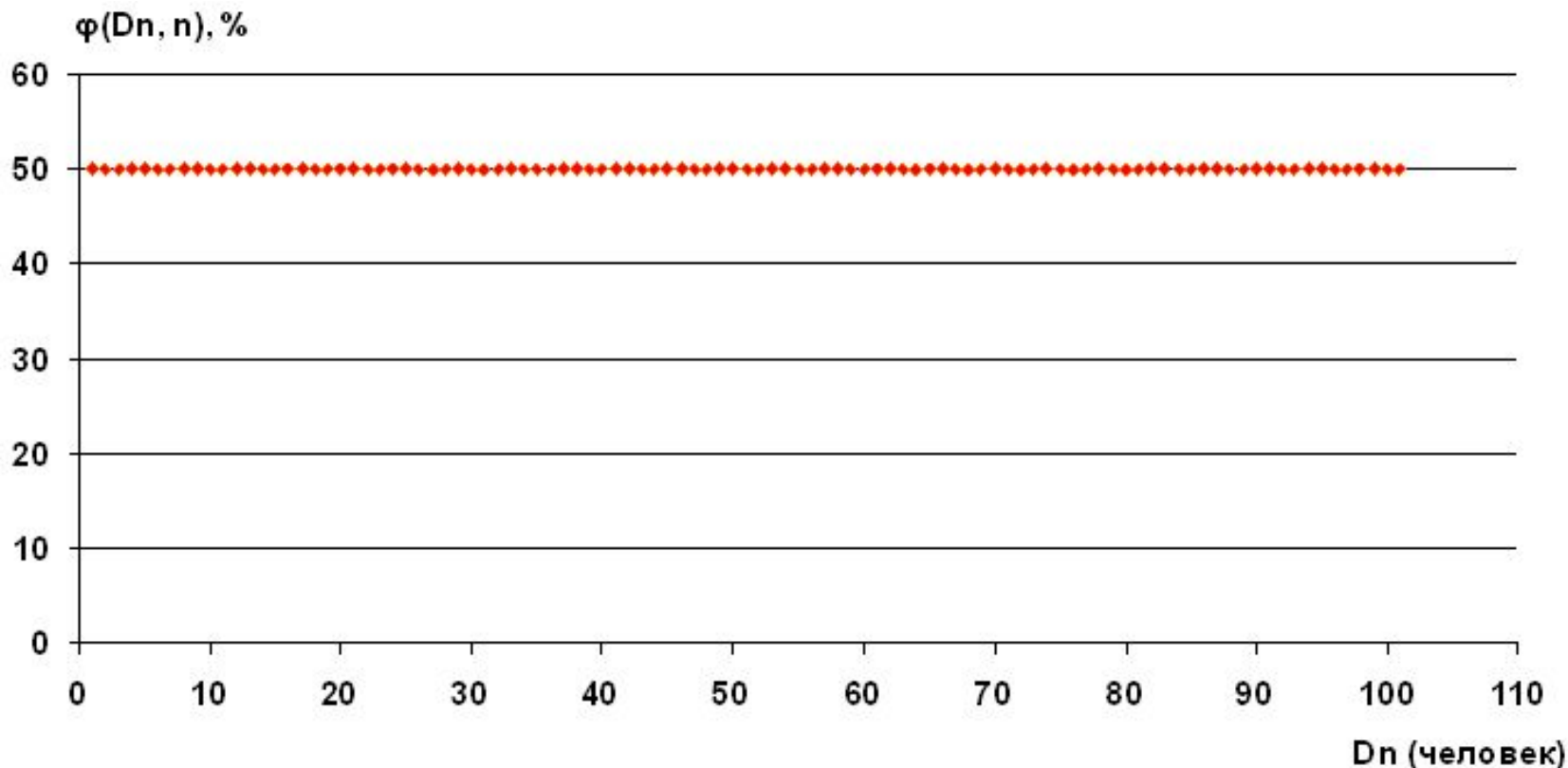


График вероятности попадания произвольного больного в группу А.
 $\alpha = 100$, $\beta = 0$, $n = 50$.

Вариант модели избирательной урны с ($\alpha=0$, $\beta>0$) выгодно использовать для рандомизации больных в небольшие группы, поскольку он обеспечивает лучший количественный баланс малых групп, чем полная рандомизация, и в то же время бóльшую случайность распределения больных по группам, чем метод перестановки блоков.

При $\alpha=1$ увеличение β с 1 до 5 приводило к повышению баланса групп, при $\beta>5$ существенного улучшения баланса не наблюдалось. При больших величинах α (вплоть до 500) улучшение балансировки групп наблюдалось при возрастании β до 12-13, при дальнейшем повышении β разница дисбалансов всякий раз оказывалась незначимой.

Метод минимизации (ковариативная адаптивная рандомизация)

Предполагает построение функций индивидуального баланса для каждого из учитываемых существенных факторов и функции общего баланса для всего исследования, причем переменными этих функций являются, в числе прочих, значения вероятности (p) назначения больному той или иной терапевтической схемы. При этом *каждый новый больной, включаемый в исследование, оценивается на предмет уровней мониторируемых ковариат*, данные уровни подставляются в вышеописанные формулы, и больной распределяется в ту из формируемых групп, где величина рассчитанного по соответствующей формуле дисбаланса ковариат окажется минимальной, а вероятность p – соответственно, максимальной. Данный способ позволяет сбалансировать формируемые группы по 10-20 существенным факторам (ковариатам), что абсолютно недостижимо при

К недостаткам метода минимизации относятся:

- Предопределенность (детерминированность) получаемых результатов: в каждый момент времени можно предсказать, в какую из групп будет распределен конкретный пациент.
- Данный метод рандомизации сложен на этапе подготовки к проведению исследования и требует наличия в исследовательском штате квалифицированных математиков и статистиков.
- Перечень рандомизации невозможно сгенерировать заранее, перед началом исследования. Это – *общий недостаток динамических моделей рандомизации*: каждая последующая итерация распределения участников по группам зависит от всех предыдущих; при каждом новом эпизоде рандомизации приходится вновь запускать алгоритм минимизации.

- При неправильной оценке уровней мониторируемых ковариат у включаемого в исследование больного он будет распределен в неверную группу; при этом общий баланс ковариат в формируемых группах нарушится, причем неявно для исследователей.
- В процессе исследования больные могут по разным причинам выйти из него, покинув свои группы, что явным образом приведет к нарушению достигнутого баланса ковариат.
- Алгоритм минимизации может быть реализован с ошибками на программном уровне; описан случай дефектного распределения более чем 1000 женщин в сравниваемые группы до выявления ошибки в алгоритме. Данная проблема отчасти преодолевается неоднократным тестовым запуском рабочего алгоритма до начала исследования с вдумчивым анализом получаемых результатов.

Реактивная адаптивная рандомизация предполагает целенаправленное изменение правил распределения новых участников исследования по группам сравнения в зависимости от особенностей реакции уже включенных в исследование лиц на изучаемые воздействия. Данная разновидность рандомизации предпочтительна с этической точки зрения в тех клинических испытаниях, где производится сравнение методов диагностики и, особенно, лечения серьезных жизнеугрожающих заболеваний, причем один из методов может оказаться существенно более эффективным, чем другие. В этом случае неэтично подвергать опасности жизнь и здоровье остальных участников исследования, и схема рандомизации новых участников должна предусматривать искусственное повышение вероятности их попадания в группу лиц, получающих наиболее эффективное лечение из сравниваемых; приходится считаться с тем, что на момент начала испытания соотношение эффективности

«Игра на лидера»

Наиболее простая ее разновидность, описанная Zelen, предусматривала рандомизацию в 2 группы, причем имелось 2 взаимоисключающих исхода лечения – «успех» и «неудача». Распределение каждого нового участника исследования (n) в одну из двух групп определялось результатом исследуемого вмешательства у предыдущего ($n-1$) участника: если оно было успешным, то участник n распределялся в ту же группу, что и $n-1$, а если безуспешным – то в другую группу (схема «все или ничего»). Известен ряд модификаций данного подхода, допускающих сравнение 3-х и более схем лечения, а также учитывающих отдаленные эффекты лечения (не всегда в клинической практике успех или неудача терапии очевидны сразу после прохождения курса лечения; кроме того, учитываемый исход может не успеть развиться за время непосредственного наблюдения за субъектом « $n-1$ »).

Несмотря на очевидный детерминизм, «игра на лидера» остается стохастическим процессом, так как определяется успехом либо неудачей исследуемых вмешательств в каждом

К недостаткам описанного способа рандомизации относятся некоторые неотъемлемые его черты, как-то:

– Преимущественная ориентация на методы лечения, дающие *немедленный положительный либо отрицательный исход* (к моменту окончания курса терапии).

– *Значительная продолжительность* процесса рандомизации, т.к. каждый последующий участник испытания может быть распределен в какую-либо из сравниваемых групп только после формирования определенного исхода вмешательства у предыдущего участника; этот факт также затрудняет регистрацию отдаленных результатов терапии.

– Адекватное использование адаптивных моделей рандомизации требует наличия в исследовательской группе *квалифицированного математика* со знанием прикладной медицинской статистики, а также подготовленного

Важным элементом рандомизации является сокрытие ее результатов, или т.н. «ослепление» (blinding). РКИ признаются «золотым стандартом» доказательной медицины именно потому, что включают в себя этап рандомизации, который максимально устраняет систематические ошибки при формировании сравниваемых групп. Указанные ошибки заключаются прежде всего в дисбалансе количественного и качественного состава сравниваемых групп, а также дисбалансе ковариат. Главную роль в возникновении указанного дисбаланса играют персональные пристрастия врачей-экспериментаторов, которые по каким-либо причинам, нередко – совершенно субъективным, питают те или иные личные симпатии или антипатии к определенным медикаментозным препаратам, а также вкусы больных, предпочитающих тот или иной метод лечения другому, нередко на основании нелепых домыслов, догадок и слухов. С целью исключения личного субъективизма всех участников испытания и производится сокрытие (ослепление)

1. Соккрытие рандомизационных кодов

а) участник исследования, непосредственно осуществляющий рандомизацию в группы при регистрации новых участников (например, врач приемного покоя стационара) получает из центра управления ходом РКИ немаркированные заклеенные конверты, в каждом из которых содержится один рандомизационный код. Указанный код не должен иметь ничего общего с названием препаратов или инструментальных вмешательств; в идеале он должен представлять собой внешне бессмысленное буквенно-цифровое сочетание. При регистрации каждого нового участника исследования экспериментатор извлекает из контейнера очередной конверт, вскрывает его, читает рандомизационный код и распределяет больного в соответствующую группу. Как вариант, возможно использование специально сгенерированной таблицы псевдослучайных чисел, которые последовательно используются для рандомизации участников РКИ в группы и

б) В случае, если рабочее место экспериментатора, осуществляющего рандомизацию, оснащено компьютером, связанным с сервером центра управления РКИ через Интернет, возможно использование клиентской программы, при помощи которой сотрудник отправляет серверу запрос; по получении запроса на сервере запускается алгоритм рандомизации, и результат (сгенерированный в реальном времени рандомизаци-онный код) отправляется обратно на компьютер исследователя, после чего его можно визуализировать при помощи клиентской части программы. Такого рода подход удобен при осуществлении динамических схем рандомизации, где результат каждой следующей итерации зависит либо от совокупности характеристик больных, ранее включенных в исследование, либо от исхода исследуемого вмешательства у ранее включенного в испытание участника, вследствие чего рандомизационные коды нельзя подготовить заранее, до начала исследования.

В идеале расшифровка рандомизационных кодов должна находиться в руках только организаторов или даже только спонсоров исследования (если речь идет об испытаниях лекарственного препарата).

Раскрытие рандомизационных кодов до полного завершения РКИ и обработки его результатов не допускается; в исключительных случаях, оговоренных в протоколе исследования, такое раскрытие (с обязательным официальным извещением организаторов и/или спонсоров исследования) все же может иметь место; обычно это происходит в том случае, если предварительный (промежуточный) анализ результатов эксперимента показал явное, существенное и статистически значимое преимущество какого-либо из сравниваемых вмешательств, либо выявил высокую частоту неблагоприятных исходов, побочных эффектов и опасных осложнений при применении какого-либо из тестируемых методов лечения.

В обоих случаях дальнейшее проведение исследования с использованием прежнего протокола неэтично и сопряжено с опасностью для жизни и здоровья участников эксперимента, вследствие чего целесообразно раскрыть рандомизационные коды и либо:

- 1) назначить всем больным лечение, оказавшееся наиболее эффективным, либо
- 2) отменить препарат / прекратить вмешательство, ведущие к неблагоприятным исходам и осложнениям исследуемого заболевания либо состояния.

В целях своевременного выявления подобного развития событий в протоколе РКИ *обязательно необходимо предусмотреть этапы исследования, на которых будет производиться промежуточный анализ его результатов, и действия, которые будут предприняты экспериментаторами при обнаружении соответствующих тенденций.*

Соккрытие применяемых схем терапевтических вмешательств

Ни больные, ни врачи, выполняющие сравниваемые вмешательства, не должны догадаться по виду и количеству препаратов, какое именно лечение получают подопытные в той или иной группе, включая *контрольную*. Необходимо учитывать также т.н. «плацебо-эффект», имеющий место у большинства людей при назначении им лекарств: у многих больных состояние может несколько (у наиболее внушаемых – существенно) улучшиться под влиянием самого факта назначения лечения независимо от его реальной эффективности. Плацебо-эффект может вносить существенный вклад в клинические проявления активности препаратов: так, существует довольно обширная группа (до 30-35% популяции) т.н. «*плацебо-реакторов*», которые могут реагировать на назначение любой медикаментозной терапии появлением тошноты и рвоты, повышенной утомляемости, головной боли и т.п., что симулирует побочные эффекты. В основе плацебо-эффекта лежит самовнушение, в некоторых

Для нивелирования «эффекта плацебо» больные в контрольной группе должны получать такие же по виду и количеству «препараты», что и во всех опытных группах. Соответственно, в опытных группах количество и вид получаемых каждым больным препаратов / инвазивных манипуляций также должны быть абсолютно идентичны. Маскирование вида хирургических вмешательств проблематично; что же касается испытаний медикаментозных препаратов, то фармацевтические компании, заказывающие такое испытание, специально для РКИ выпускают партии сравниваемых лекарств в идентичной упаковке, маркированной только рандомизационным кодом; сами таблетки или капсулы выглядят абсолютно идентично, независимо от содержащегося в них препарата, и либо маркированы рандомизационным кодом, непонятным непосвященному, либо вообще не маркированы.

Для назначения в контрольной группе специально выпускается т.н. **плацебо** – вещество, не имеющее никаких терапевтических действий, но неотличимое от лекарства по виду, вкусу, массе, цвету, запаху и т.д. В идеальном случае плацебо действительно не содержит никаких медикаментозных препаратов (основу его может составлять, например, молочный сахар); такое средство предназначено исключительно для моделирования плацебо-эффекта в контрольной группе и называется инертным плацебо. В том случае, если плацебо содержит какой-либо действующий препарат, бесполезный или неэффективный при изучаемом заболевании, но в целом безвредный или условно-полезный для больного (например, витамины), оно называется активным плацебо. В случае, если лица в контрольной группе получают плацебо, говорят о РКИ с плацебо-контролем.

Использование плацебо-контроля, являясь отраслевым стандартом, тем не менее, порождает этическую проблему, поскольку больные в контрольной группе фактически *остаются вообще без лечения* на все время исследования. В связи с этим в ряде испытаний применялся т.н. активный, или позитивный контроль. В этом случае в контрольной группе назначают схему лечения, включающую ранее разработанные препараты с известной эффективностью против изучаемого заболевания, и сравнение новой схемы / новых препаратов производится с уже известными. При этом данные препараты должны быть внешне неотличимыми от используемых в опытной группе (т.н. двойное маскирование). *РКИ с активным контролем постоянно используют при оценке клинической эффективности лекарств-дженериков в сравнении с оригинальными препаратами.*

Все РКИ подразделяются на:

1. Слепые – о результатах рандомизации и применяемых схемах терапии не знают только пациенты – участники испытания;
2. Двойные слепые – ни пациент, ни врачи-исследователи не владеют информацией о результатах рандомизации и применяемых вмешательствах;
3. Тройные слепые – ни пациент, ни исследователи, ни фармацевты, выдающие препараты для испытания, не знают результатов рандомизации. В этом случае рандомизационные коды известны только организаторам и/или спонсорам исследования; нередко их не знает даже статистик, производящий промежуточный и окончательный анализ результатов РКИ.

Отраслевым стандартом являются двойные слепые РКИ с плацебо-контролем; считается, что такой уровень сокрытия результатов рандомизации является достаточным для устранения субъективности при получении и анализе результатов исследования, а более строгое сокрытие

5. Следующий этап планирования РКИ – определение размера сравниваемых групп (т.е объема формируемой выборки)

Размер формируемых групп сравнения имеет принципиальное значение при последующем анализе полученных в РКИ результатов. В процессе РКИ мы всегда проверяем некую гипотезу, например, что тестируемый препарат А лучше (или не хуже) ранее созданного препарата В, или что препарат А обладает клинической эффективностью, значительно превышающей эффект плацебо. При этом мы не можем включить в наши группы всю интересующую нас популяцию – мы вынуждены делать выборки из нее, вследствие чего получаемые нами результаты всегда характеризуются элементом неопределенности: мы никогда не можем быть уверенными на 100%, что сделанное нами заключение об истинности или ложности проверяемой гипотезы безошибочно.

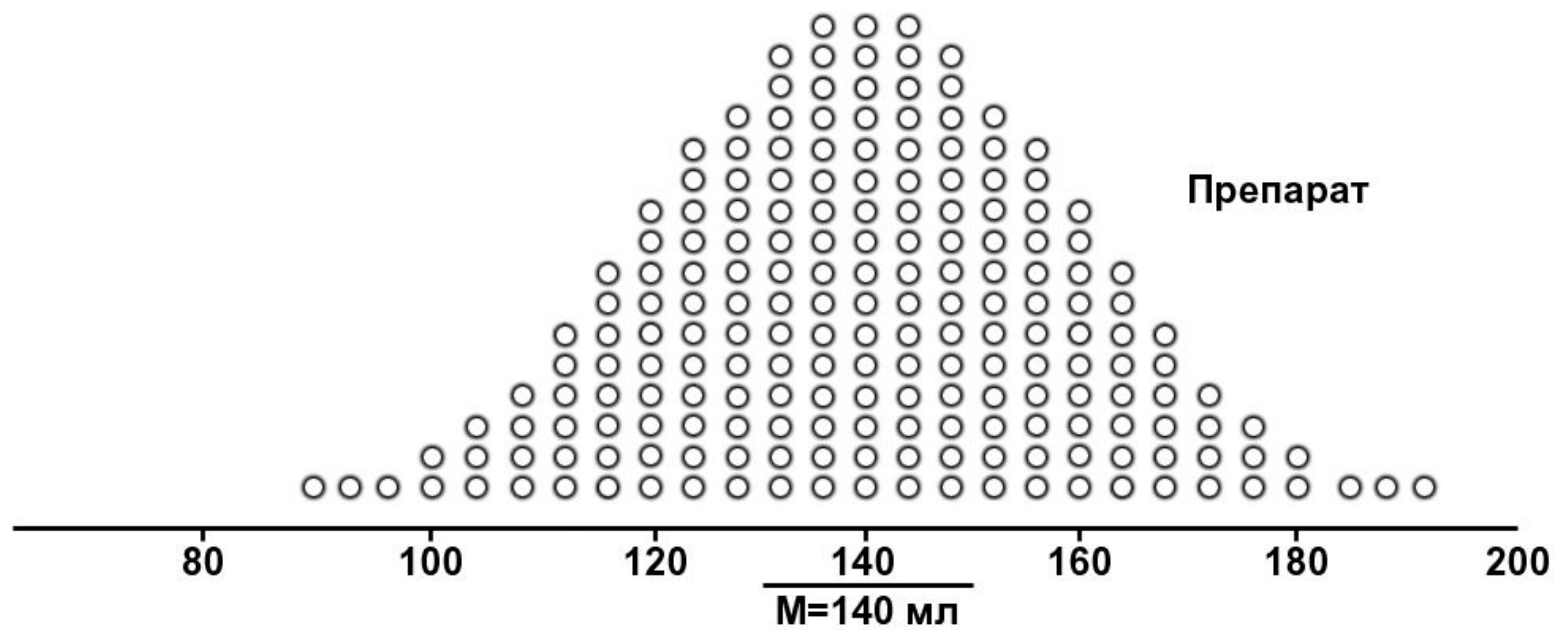
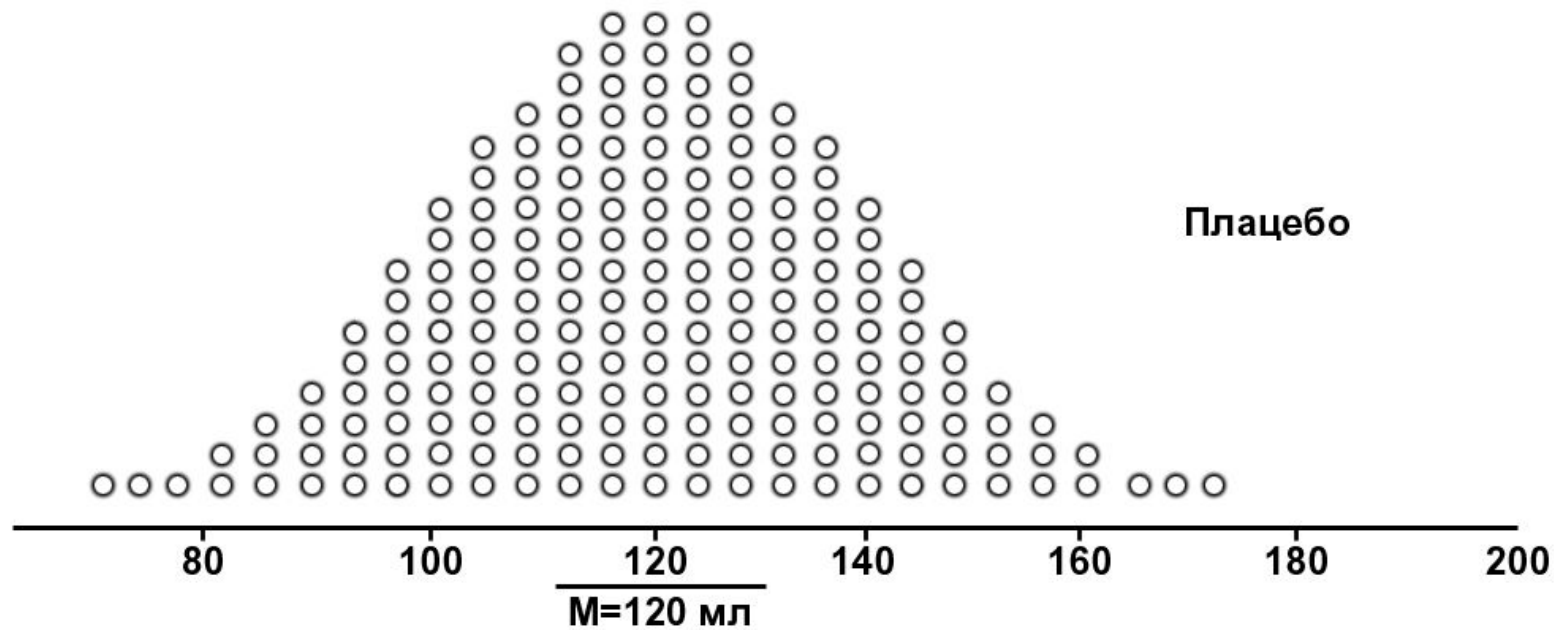
($n=10 \times 2$; $A=5$, $B=6$, $p=0,5$ – точный тест Фишера)

($n=100 \times 2$; $A=50$, $B=60$, $p=0,15$)

($n=1000 \times 2$; $A=500$, $B=600$, $p < 0,0001$)

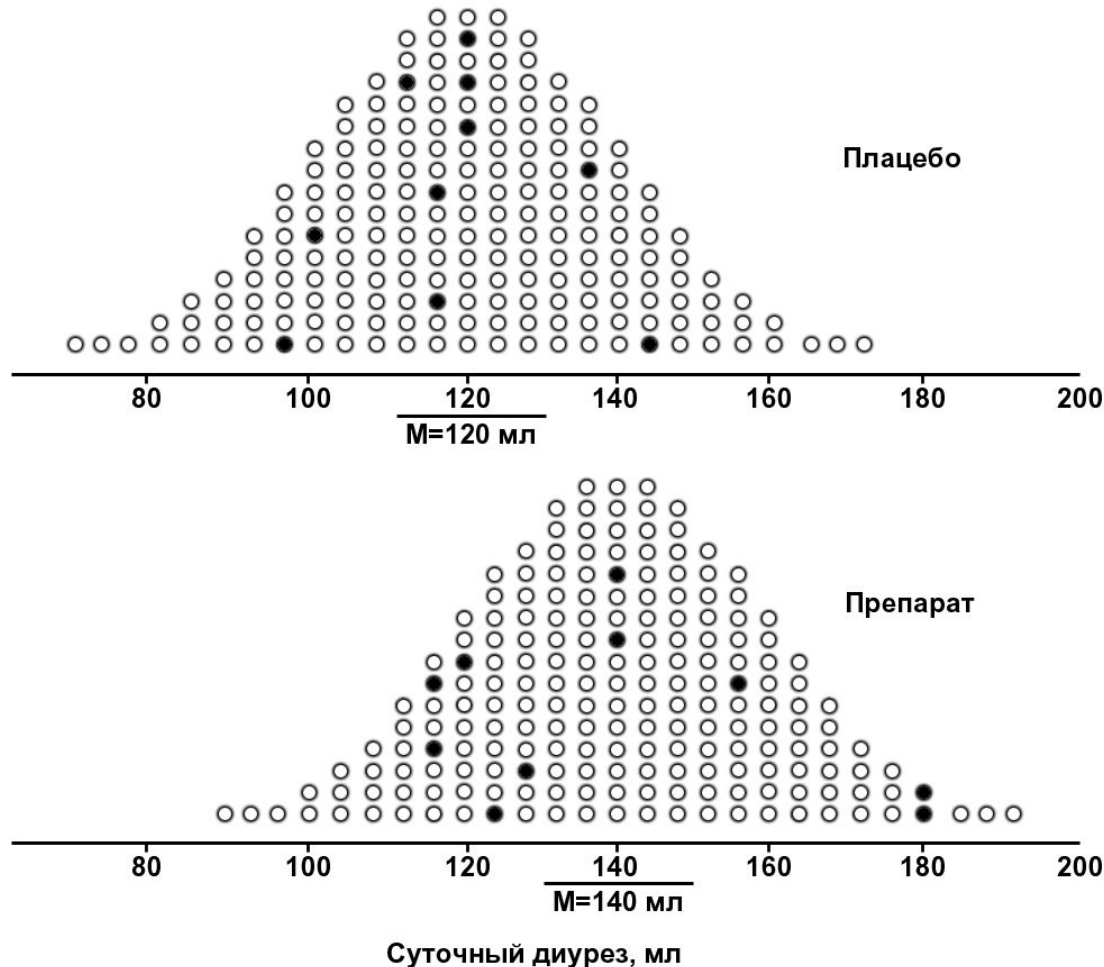
Понятно, что с увеличением размеров сравниваемых групп статистическая значимость наблюдаемых в них различий в результатах сравниваемых вмешательств будет нарастать. Параллельно будет нарастать стоимость исследования и сложность его организации. Размер сравниваемых групп в 10000 человек каждая даст еще более значимые результаты сравнения, но будет разорительным для исследователей. Итак, при формировании групп мы должны достичь некоей золотой середины: когда размер групп уже достаточно велик для получения значимых результатов сравнения, и в то же время достаточно мал для того, чтобы стоимость исследования осталось в рамках сметы.

Вторым аспектом данной проблемы является тот факт, что выборка (т.е. группа сравнения) формируется путем привлечения индивидуумов из популяции случайным образом. В связи с этим никто не может гарантировать, что картина распределения эффекта исследуемого вмешательства, наблюдаемая в генеральной совокупности (т.е. во всей популяции), хоть сколько-нибудь точно будет воспроизведена в сформированной выборке; при этом чем меньше размер формируемых групп, тем большее влияние на наблюдаемые в них результаты оказывает фактор случайности. Одной из типичных иллюстраций данного феномена является испытание гипотетического диуретика в изолированной популяции, насчитывающей 200 человек, причем в ходе РКИ формируется 2 группы добровольцев – плацебо-контроль и опытная группа, включающие по 10 человек каждая. Нам, как представителям высшей истины, исходно известно, что испытуемый препарат эффективен, и его действие превосходит плацебо-эффект на 20 мл/сут.

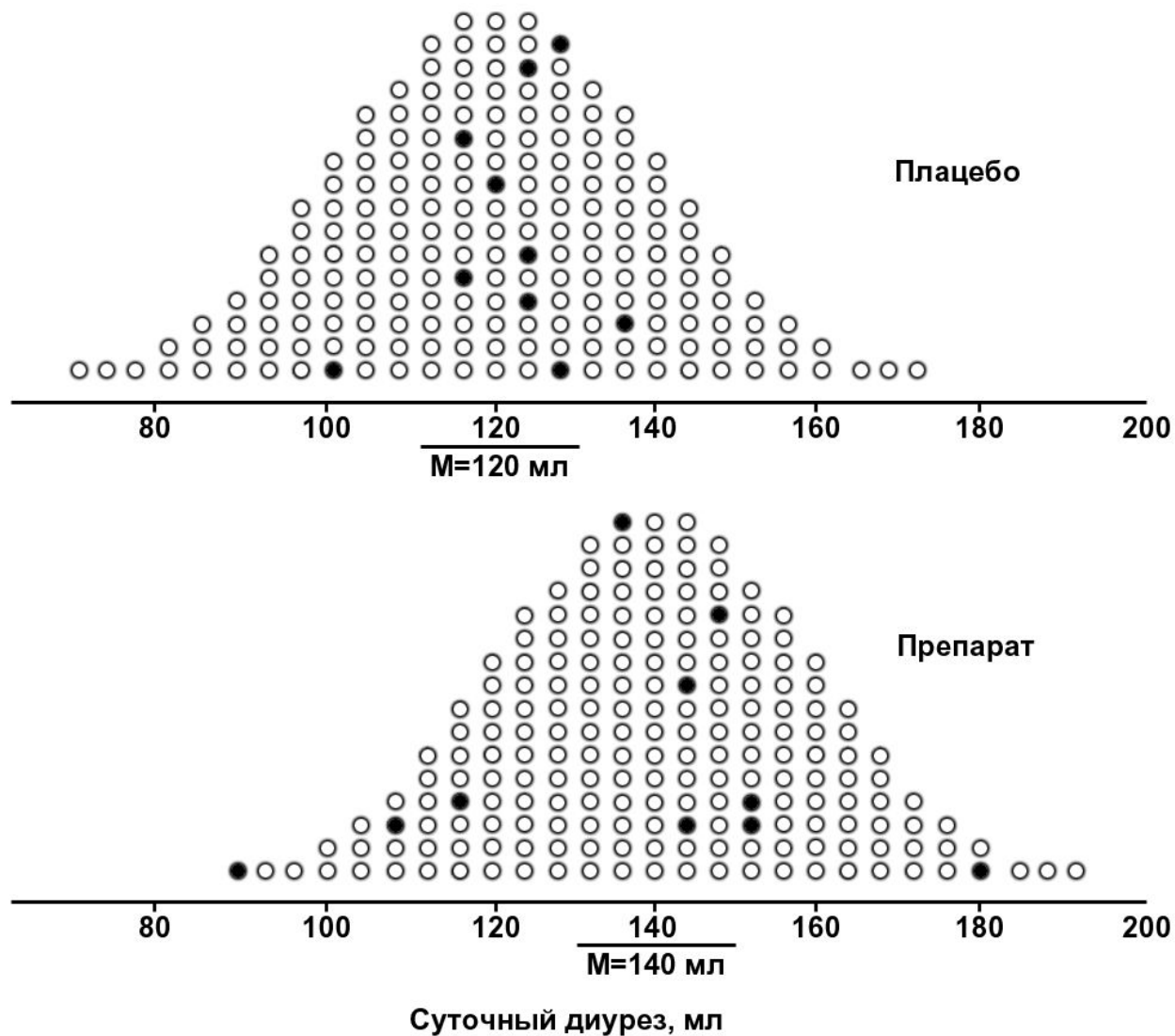


Суточный диурез, мл

Мы выполняем две случайные выборки из генеральной совокупности, каждая по 10 человек – для контрольной и опытной групп; таких выборок (по 10 человек из 200) может существовать 10^{27} . В показанном случае средняя арифметическая суточного диуреза в опытной группе составит $140,0 \pm 24,5$ мл, а в группе плацебо-контроля – $118,0 \pm 14,4$ мл, при этом их разница оказывается статистически значимой (U-тест Манна-Уитни, $p < 0,05$) → испытуемый диуретик более эффективен, чем плацебо



Тем не менее, случайно набранные в группы сравнения добровольцы могут распределиться в генеральной совокупности и по-другому:



В этом случае средняя арифметическая суточного диуреза в опытной группе составит $136,8 \pm 26,3$ мл, а в группе плацебо-контроля – $121,6 \pm 9,7$ мл, при этом их разница оказывается статистически незначимой (U-тест Манна-Уитни, $p > 0,05$), гипотеза о том, что испытуемый диуретик более эффективен, чем плацебо, не подтверждается, и в целом неясно, какой из проведенных экспериментов (№1 или №2) в большей степени заслуживает доверия, поскольку проанализировать всю генеральную совокупность по условию задачи невозможно. Вообще, расчеты показывают, что в данном случае из всех возможных выборок по 10 человек только в 55% случаев можно выявить статистически значимый эффект диуретика.

Формируемая выборка должна быть репрезентативной, т.е. отражать структуру изучаемого явления в генеральной совокупности настолько точно, насколько это возможно, и при этом следовать принципу минимальной достаточности.

Проверка гипотезы	Решение, принятое исследователями:	
Действительное положение дел:	Отличается ($p < 0,05$)	Не отличается ($p > 0,05$)
Новое вмешательство отличается от существующего	Цель достигнута	Ошибка 2 рода (β): ложноотрицательное заключение
Новое вмешательство не отличается от существующего	Ошибка 1 рода (α): ложноположительное заключение	Цель достигнута

Статистическая ошибка 1 рода (α) – вероятность получить ложноположительный результат. Допустимая величина α , принимаемая в большинстве клинических испытаний $\leq 5\%$ (реже $\leq 1\%$), т.е. $p \leq 0,05$ или $p \leq 0,01$. Статистическая ошибка 2 рода (β) – это вероятность прийти к ложноотрицательному результату. В РКИ величину β обычно принимают $\leq 20\%$ (реже $\leq 10\%$). Величина $100 - \alpha$ является *специфичностью* проводимого исследования, а величина $100 - \beta$ – *чувствительностью*. Чувствительность исследования называют также его мощностью (power). Соответственно, при величине ошибки 2 рода 20% мощность исследования составляет $100 - 20 = 80\%$

Для того, чтобы рассчитать оптимальный размер выборки, необходимо заранее знать:

- 1) требуемое допустимое значение α (обычно в биомедицинских исследованиях $\alpha \leq 0,05$,
- 2) требуемую мощность исследования (обычно в биомедицинских исследованиях $100 - \beta = 70 \dots 90\%$),
- 3) уровень наименьших различий в эффективности сравниваемых вмешательств, имеющий значение для экспериментаторов (эту величину также называют «минимальная разрешающая способность исследования»).

Данная величина зависит сугубо от произвола экспериментаторов и в реальных РКИ обычно составляет не менее 5-10% (т.е. 0,05-0,1).

Подсчет оптимального размера выборки может производиться с использованием номограмм, формул, компьютерных программ (в т.ч. онлайн-овых).

Для разных видов исследований используются различные формулы подсчёта n:

Учитываемый показатель – одиночное среднее арифметическое (ситуация, не встречающаяся при РКИ, но типичная при описаниях серии случаев или срезовых исследованиях):

$$n \geq \frac{(u + v)^2 \sigma^2}{(\mu - \mu_0)^2}$$

где:

n – оптимальный объем выборки;

$\mu - \mu_0$ – разница между средним арифметическим (μ) и средним арифметическим нулевой гипотезы (μ_0);

σ – стандартное отклонение;

u – односторонний процентный пункт (percentage point) нормального распределения, соответствующий мощности исследования;

v – процентная доля нормального распределения, соответствующая требуемому (двустороннему) уровню значимости (α), т.е. если $\alpha = 0.05$, то $v = 1.96$.

u – односторонний процентный пункт (percentage point) нормального распределения, соответствующий мощности исследования. Вычисляется при помощи т.н. квантильной функции или пробит-функции (в программе Statistica 6-8 вычисляется в диалоговом окне, находящемся в меню Statistics → Probability Calculator → Distributions → Z (normal), в диалоговом окне «p» указывается мощность исследования (в долях от единицы)). Например, если заданная мощность нашего исследования – 90% (0,9), то соответствующий ей $u=1,28$;

v – процентная доля нормального распределения, соответствующая требуемому (двустороннему) уровню значимости (α), т.е. если $\alpha=0,05$, то $v=1,96$. Вычисляется при помощи той же функции Statistica Z (normal), причем необходимо выбрать опцию «Two-tailed», а в поле «p» ввести величину, равную $100-\alpha$. Результат считывается в окошке «X».

Учитываемый показатель – одиночная доля или процент:

$$n \geq \frac{(u + v)^2 \mu^2}{(\mu - \mu_0)^2}$$

где:

μ – частота встречаемости;

μ_0 – частота встречаемости, соответствующая нулевой гипотезе;

u, v – аналогично описанному выше.

Данная формула возвращает величину выборки в тех единицах, в которых был выражен знаменатель (например, если частота события рассчитана на человеко-год, то и размер выборки будет подсчитан в требуемом количестве человеко-лет).

Учитываемый показатель – одиночная пропорция:

$$n \geq \frac{\{u\sqrt{[\pi(1-\pi)]} + v\sqrt{[\pi_0(1-\pi_0)]}\}^2}{(\pi - \pi_0)^2}$$

где:

π – пропорция;

π_0 – пропорция в нулевой гипотезе;

u, v – аналогично описанному выше.

Сравнение двух средних арифметических (формула возвращает размер каждой из групп; предполагается, что количество добровольцев в сравниваемых группах одинаково):

$$n \geq \frac{(u + v)^2 (\sigma_1^2 + \sigma_0^2)}{(\mu_1 - \mu_0)^2}$$

где:

$\mu_1 - \mu_0$ – разница между средними сравниваемых групп;

σ_1, σ_0 – стандартные отклонения соответствующих групп;

u, v – аналогично описанному выше.

Сравнение двух долей или процентов (формула возвращает размер каждой из групп; предполагается, что количество добровольцев в сравниваемых группах одинаково).

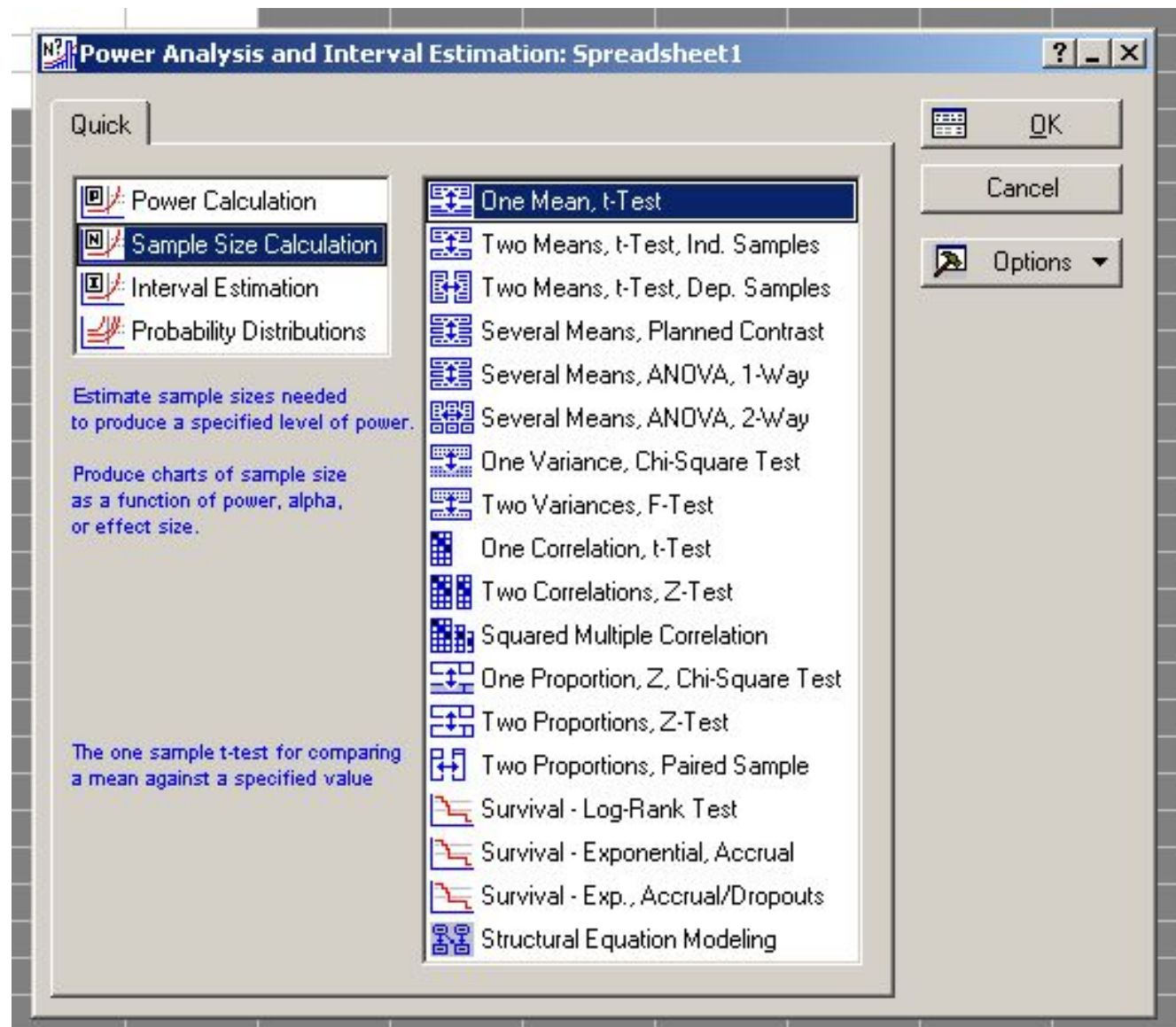
Данная формула возвращает величину выборки в тех единицах, в которых был выражен знаменатель доли.

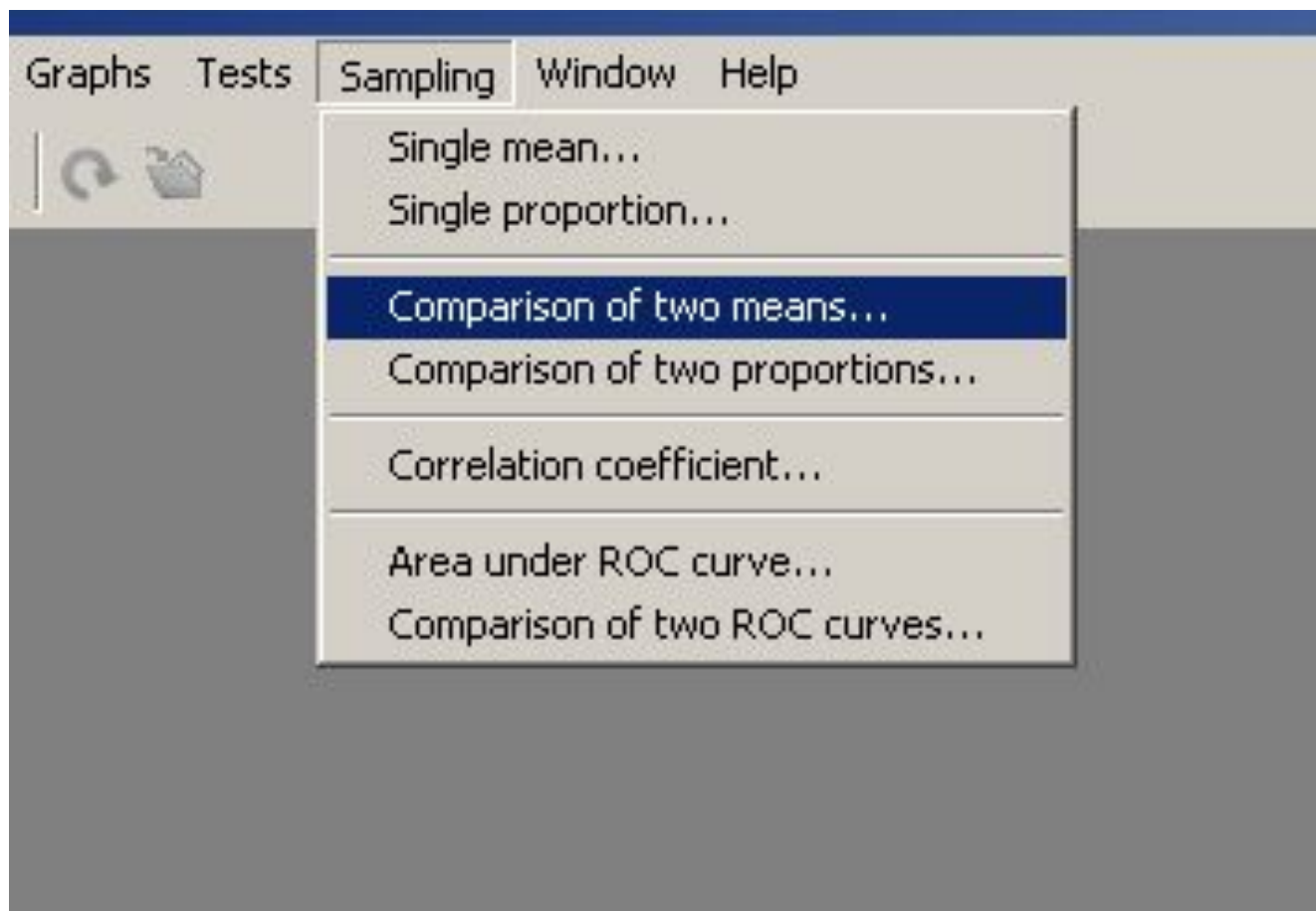
$$n \geq \frac{(u + v)^2 (\mu_1^2 + \mu_0^2)}{(\mu_1 - \mu_0)^2}$$

где:
 μ – частота встречаемости;
 μ_0 – частота встречаемости, соответствующая нулевой гипотезе;
 u, v – аналогично описанному выше.

...и так далее.

Statistics → Power Analysis → Sample Size Calculation





Возвращаясь к примеру испытания диуретика на острове с населением в 200 человек, легко подсчитать, что величина ошибки 2 рода в данном случае составляет 55%, и, соответственно, мощность исследования равна 45%, что существенно ниже требуемых 80-90%. Можно подсчитать и оптимальный размер выборки для данного случая (сравнение двух средних арифметических, в обоих сравниваемых группах имеет место нормальное распределение данных): **по 26 человек в каждой из групп** (что обеспечит мощность исследования 80%), или **по 34 человека в каждой из групп** (что соответствует мощности исследования 90%).

Если оптимальный размер сравниваемых групп определить невозможно из-за недостатка исходных данных, рекомендуется проведение «пилотного» исследования с набором произвольного (разумного) количества добровольцев. Обычно такое исследование имеет дизайн срезового или описания серии случаев. Исходя из полученных предварительных результатов, можно рассчитать оптимальный размер выборки для проведения исследования с более высоким уровнем доказательности. Истинные результаты такого исследования могут существенно отличаться от предварительных, поэтому мощность исследования по мере накопления данных может оказаться недостаточной. Промежуточный подсчет мощности исследования по мере его выполнения с учетом накопленных данных может использоваться для пересчета и коррекции оптимального размера формируемой выборки (если РКИ – открытое, т.е. в ходе исследования продолжается набор добровольцев в сравниваемые

Все предыдущие вычисления исходили из условия равенства размеров сравниваемых групп. В то же время возможны ситуации (адаптивный дизайн), когда размер сравниваемых групп обязательно окажется неодинаковым. В этом случае поступают следующим образом:

1. Одним из вышеперечисленных способов определяют оптимальный для данного случая размер выборки (n);
2. Определяют отношение наибольшей группы к наименьшей (c);
3. Подсчитывают фактор коррекции (correction factor): $f = \frac{(c + 1)}{2c}$
4. Оптимальный размер меньшей группы = ($f \times n$)
5. Оптимальный размер большей группы = ($f \times c \times n$)

В практической работе неизбежны ситуации, когда определенное количество добровольцев по разным причинам покидают исследование до его завершения. Мощность исследования при этом падает по сравнению с запланированной. Наиболее простым выходом является введение поправочного коэффициента на возможное убытие больных (**обычно – 10-15%**); соответственно, оптимальный размер выборки при вычислениях дополнительно увеличивается на эти 10-15%. Данный коэффициент каждый исследователь определяет самостоятельно исходя из личного опыта и данных о предыдущих РКИ со сходным дизайном, выполненных ранее на той же клинической и популяционной базе.

Типичная запись об определении оптимального размера выборки в отчете о реальном исследовании выглядит следующим образом: «Было подсчитано, что выборки по 60 человек в каждой из двух групп достаточно для обеспечения мощности исследования 80% с целью регистрации снижения ВИЧ-1-вирусной нагрузки в плазме крови в 5 раз и более через 72 недели после начала терапии. За базовый был принят уровень стандартного отклонения средних значений вирус-нагрузки в $0,9 \times \log_{10}$; поправочный коэффициент на возможное самовольное прекращение лечения участниками испытания был установлен равным 8%. В ходе исследования размер обеих групп был увеличен до 72 человек, чтобы скомпенсировать неожиданно высокую частоту самовольного отказа от лечения, составившую 16%»

6. Выбор ожидаемых исходов (конечных точек, endpoints) РКИ

Ввиду вынужденной кратковременности РКИ в качестве регистрируемых исходов обычно используют подходящие суррогатные маркеры. Более важные для понимания результатов вмешательства «истинные» исходы за время проведения РКИ обычно не успевают развиться, кроме сравнения терапевтических схем, используемых для лечения острых заболеваний, где истинные исходы (полное выздоровление, хронизация либо смерть больного) наступают в сроки, сравнимые с общей продолжительностью курса терапии.

Исходы необходимо подбирать, исходя из целей и задач планируемого исследования; при этом сами цели и задачи должны быть четко сформулированы, реальны, разумны и достижимы, а квалификации и оснащения исследователей

7. Выбор методов статистической обработки результатов РКИ

Основа анализа данных РКИ – непосредственное сравнение частоты/выраженности исходов в сравниваемых группах. Поскольку считается, что рандомизация «выравнивает» сравниваемые группы по всем принципиально важным показателям, для их сравнения рекомендуется использовать простые статистические тесты. В сложных методах анализа нет необходимости, *поскольку вся разница между группами на момент окончания исследования считается проистекающей исключительно из различий в применяемых вмешательствах*. Соответственно, нет смысла учитывать прочие факторы, кроме собственно сравниваемых схем терапии. При публикации результатов необходимо представить читателю «эффективность лечения» (разницу в исходах между сравниваемыми группами), ее 95%

Для непосредственного сравнения эффективности вмешательств в анализируемых группах рекомендуются следующие статистические методы:

Вид данных	Параметрический тест	Непараметрический тест
Непрерывный ряд данных (количественные и порядковые переменные)		
Попарные сравнения (н-р, параметр до- и после вмешательства) в пределах одной группы испытуемых	Парный (одновыборочный) t-тест	Ранговый знаковый тест Уилкоксона (Wilcoxon)
Непарные данные (одна и та же переменная, две независимые выборки)	Непарный (двухвыборочный) t-тест	U-тест Манна-Уитни (Mann-Whitney)
Непарные данные (одна и та же переменная, три и более независимые выборки)	ANOVA (дисперсионный анализ)	Тест Краскела-Уоллиса (Kruskal-Wallis)
Бинарные данные (номинальные переменные)		
Сравнение двух и более групп	Тест по критерию Chi-квадрат (χ^2), точный тест Фишера	

- 1) Доказать превосходство схемы А над схемой В (superiority test),
- 2) Доказать эквивалентность схем А и В (equivalence test),
- 3) Доказать, что схема А не хуже схемы В (non-inferiority test).

Обычно при анализе результатов РКИ используется принцип «каждая попытка засчитывается» (intention to treat). Данный принцип включает 2 положения:

1. При анализе результатов учитываются все участники РКИ, прошедшие рандомизацию. Указанное положение может потребовать использования в ходе исследования допущений вида «неявка = неудача вмешательства» и «данные последнего (перед исчезновением больного) обследования = данные на момент окончания исследования»;

2. Даже если доброволец, рандомизированный в одну из сравниваемых групп, в дальнейшем отказался от участия в исследовании до его завершения либо настоял на смене схемы терапии (т.е. переходе в другую группу), анализ производится так, как если бы данный больной входил в ту группу, в которую он был исходно распределен.

Альтернативный, реже используемый принцип учета результатов клинических испытаний «в соответствии с протоколом» (“**per protocol**”), или «пролеченные» (“**on treatment**”). В этом случае финальному анализу подвергаются только данные, полученные от пациентов, идеально соответствующих всем требованиям утвержденного протокола РКИ. Добровольцы, прошедшие рандомизацию, но отказавшиеся в дальнейшем от участия в исследовании по любой причине, а также лица, настоявшие на смене терапии на альтернативную (перешедшие в другую группу сравнения), удаляются из анализируемого массива; при этом изучаемая выборка уменьшается, а мощность исследования падает. Данный подход полезен в том случае, когда необходимо изучить особенности действия принципиально нового препарата либо эффекта от впервые вводимого в практику вмешательства; при этом очень важно, чтобы анализировались только данные, полученные от лиц, действительно получивших данные препарат / вмешательство, причем полным курсом / в полном объеме

Следование принципу «каждая попытка засчитывается» (“intention to treat”) позволяет оценить эффект вмешательства в условиях, приближенных к реальным.

Принцип “on treatment” *не позволяет оценить реальные клинические перспективы нового препарата либо вмешательства, связанные с возможными неудобствами при его приеме, побочными эффектами, снижением качества жизни и т.п.*; эти факторы поддаются учету только при использовании подхода “intention to treat”.



8. Международные документы, регламентирующие проведение РКИ

1. Consolidated Guideline for Good Clinical Practice (ICH E6)
2. Хельсинкская декларация ВМА / Declaration of Helsinki
3. Директива Комиссии 2005/28/ЕС от 8.04.2005 г.
4. Директива 2001/20/ЕС Европарламента и Совета от 4.04.2001 г.
5. Ethical Considerations for Clinical Trials on Medicinal Products Conducted with the Paediatric Population
6. Clinical Safety Data Management: Definitions and Standards for Expedited Reporting (ICH E2A).
7. Structure and Content of Clinical Study Reports (ICH E3)

В 2004 г. члены Международного комитета редакторов медицинских журналов (International Committee of Medical Journal Editors, ICMJE, <http://www.icmje.org/>) опубликовали редакционную статью, посвященную необходимости регистрации всех без исключения клинических испытаний. ICMJE – не единственная организация, которая преследует такие цели. На заседаниях ВОЗ в Нью-Йорке, Мехико и Женеве также говорилось о необходимости создания единых мировых стандартов, касающихся информации, которую обязаны предоставлять авторы контролируемых исследований. В настоящее время правительства разных стран начинают подводить под это положение правовую основу. Например, конгресс США разработал Акт об открытом доступе к результатам клинических испытаний – Fair Access to Clinical Trials (FACT), значительно ужесточающий требования к их

Существует несколько международных регистров РКИ, одним из наиболее популярных и авторитетных из них является регистр ВОЗ (т.н. Международная платформа ВОЗ для регистрации клинических испытаний, <http://www.who.int/ictrp/en/>).

Регистрации подлежат все испытания, основной задачей которых является изменение или подтверждение существующей клинической практики (т.н. III фаза исследований).

Другие известные международные регистры РКИ – Реестр клинических испытаний методов лечения инсульта (<http://www.strokecenter.org/trials/>), IFPMA Clinical Trials Portal (http://clinicaltrials.ifpma.org/clinicaltrials/no_cache/en/myportal/index.htm), Current Controlled Trials (<http://www.controlled-trials.com/>) и т.д. Кроме того, существуют национальные реестры клинических испытаний, наиболее известным из которых является национальный регистр США (<http://clinicaltrials.gov/>)