

КАК ЧИТАТЬ МЕДИЦИНСКИЕ СТАТЬИ: ЧАСТЬ 5. ИСПЫТАНИЯ МЕТОДОВ ЛЕЧЕНИЯ И ПРОФИЛАКТИКИ ЗАБОЛЕВАНИЙ

В. В. Власов

Для клинического испытания методов лечения разработаны весьма полные стандарты. Их формирование началось еще в прошлом веке, но только в конце 40-х годов впервые стали проводиться клинические испытания, подобные современным. Сегодня проведение контролируемых клинических испытаний перспективных лекарственных и нелекарственных методов лечения стало уже не только наукой, но и особым медицинским бизнесом. Проводить их дорого, это требует специального управления и специальных навыков. Тем не менее значительная часть статей, представляющих результаты испытания методов лечения, содержит непригодную для использования информацию. Полноценная статья должна включать удовлетворительное описание 10 наиболее важных позиций (табл. 1). Эти критерии достоверности относятся к любому вмешательству — лечебному или профилактическому.

Таблица 1.

Методологические стандарты испытаний методов лечения (критерии достоверности). Публикуется по [5, 7, 10, 11] с изменениями.

Структура исследования

1. Наличие контрольной группы.
2. Ясные критерии отбора (включения и исключения) больных.
3. Включение больных в исследование до рандомизации по группам.
4. Случайный метод распределения больных по группам (рандомизация).
5. «Слепое» лечение.
6. «Слепая» оценка результатов лечения.

Представление результатов

7. Приведена информация об осложнениях и побочных эффектах лечения.
8. Приведена информация о числе больных, выбывших в ходе эксперимента.
9. Проведен адекватный статистический анализ, имеются ссылки на использованные статьи, программы и т.д.
10. Приведена информация о размере выявленного эффекта и статистической силе исследования.

Простейший критерий качества исследования — его структура: в исследовании должны быть «опытная» группа, получающая изучаемое лечение, и «контрольная» группа, получающая плацебо или иное

лечение (обычно наиболее приемлемое на момент клинического испытания). Если никакого контроля нет, оценить эффект изучаемого лечения невозможно и статью следует отбрасывать как бесполезную. К сожалению, такого рода неполноценных статей множество, и чем более экзотические средства лечения применяются (магнитные поля, лазерное облучение крови, воздействия на биологически активные точки, гипоксия при болезнях легких), тем реже клинические испытания правильно организованы и тем чаще приводятся блестящие результаты [1]. Оценивая структуру испытания, Вы должны ответить на вопрос: не является ли его результат следствием какой-нибудь ошибки? Отсутствие систематической ошибки в испытании называют обоснованностью (validity).

Если критерии исключения больных из испытания достаточно часто указываются в публикациях, то критерии включения — редко. Читатель не знает, как отбирались больные, где проводилось лечение, можно ли этих больных сопоставить с его больными. Между тем весьма существенно, включались ли, например, все больные, получавшие лечение при стенокардии, или только больные, лечившиеся в университетской клинике. Конечно, испытания на большом числе больных предпочтительнее, поскольку при этом снижается вероятность случайных эффектов.

Достаточно распространенный недостаток — выбывание больных по ходу испытания. В результате конечные цифры эффективности лечения могут быть существенно искажены. В статье, как минимум, должны быть приведены разнообразные сведения о больных, которых удалось наблюдать до конца испытания, и о выбывших больных. Точных критериев не существует, но если: а) выбывшие больные достоверно отличаются по каким-то признакам от наблюдавшихся до конца и б) если выбыло много (>20%) больных, то можно ожидать самых разнообразных искажений.

Для проверки надежности результата есть полезный, хотя и приблизительный метод. Предположим, что новое лечение лучше. Затем допустим, что у всех выбывших больных из группы с наилучшим исходом отмечалась неблагоприятный исход. Если пересчет оригинальных авторских данных выявляет тем не менее преимущества нового метода лечения, то можно поверить, что он действительно лучше.

Похожее влияние на результат испытания оказывает неисполнение плана лечения у отдельных

Адрес для корреспонденции:

Проф. В. В. Власов, Саратовский медицинский университет,
410601 Саратов, а/я 1528.

больных. Например, при сравнении терапевтического лечения с хирургическим часть больных, рандомизировано включенных в группу хирургического лечения, не будет прооперирована. У больных, не получающих предписанного лечения, состояние здоровья и прогноз обычно хуже. Поэтому в идеале должны быть представлены результаты как у больных, фактически получивших предписанное лечение, так и у всех больных, которым лечение было запланировано при рандомизации (*intention-to-treat analysis*, см. например, дискуссию [3])

Важнейшее условие — «слепой» метод проведения клинического испытания. Больной не должен знать, какое лечение он получает (плацебо или лекарственный препарат). Если о назначаемом препарате известно врачу, такая организация клинического испытания называется “простым слепым испытанием” (single-blind trial). Однако результат лечения зависит не только от изучаемого метода, но и от ряда факторов, обусловленных работой врача и других специалистов, поэтому желательно, чтобы все работающие с больными не знали, кто какое лечение получает. Это называется «двойным слепым испытанием» (double-blind trial). Такое клиническое испытание в особенности желательно там, где велика роль сопутствующего лечения. Например, в ходе проведения клинического испытания антибиотиков при раневых инфекциях необходимо, чтобы обработка и перевязка ран проводилась одинаково всем больным. Тут двойное «ослепление» абсолютно незаменимо.

Чем объективнее метод измерения эффекта, тем меньше необходимость в использовании слепого клинического испытания, но в любом случае оно не будет лишним. Даже при оценке эффективности физиотерапевтических методов или иглоукалывания возможно применение слепых методов. Именно такие испытания обычно выявляют весьма невысокую эффективность этих методов лечения [12, 15, 16].

Для проведения слепого клинического испытания необходим творческий подход. Например, если стандартный лекарственный препарат вводится посредством ингаляций, а новый — посредством инъекций, то стандартное лечение можно дополнить инъекциями плацебо, а новое лечение — ингаляциями плацебо. Тогда больной и персонал не смогут различать два способа лечения.

Важная деталь организации клинического испытания — прекратить лечение, назначенное больному до включения в испытание, и определить сроки, когда это нужно сделать. Высококачественное сообщение должно содержать информацию о том, как часто больные в группе лечения и в контрольной группе использовали дополнительное, не назначенное в ходе испытания лечение. Это бывает почти всегда и должно учитываться при анализе результатов.

Необходимо представить все применяемые методы статистического анализа, за исключением самых простых (средняя, среднеквадратичное отклонение, проценты). Реже всего встречается описание статистической силы исследования, т.е. возмож-

ности выявить с определенной степенью вероятности небольшие различия между сравниваемыми группами. Статистическая сила снижается при использовании малочисленных групп в «опыте» и «контроле». Это особенно важно в случае отрицательного результата исследования. Оценить достаточность числа больных для отрицательного результата исследования можно с помощью приводимых таблиц. В ячейке, соответствующей частоте событий в группе лечения и в контрольной группе представлено число больных в каждой группе, необходимое для выявления снижения частоты на 25% (табл. 2) или 50% (табл. 3). Если в рассматриваемой статье число больных меньше, значит эффект мог быть не обнаружен только из-за малого числа больных.

Обратите внимание на то, как в исследовании учтены сопутствующие факторы (confounding factors), например, курение, употребление алкоголя и т.д. Вполне возможно, что различия в результатах между группами обусловлены не изучаемым препаратором, а этими сопутствующими факторами. При статистическом анализе оценка различий между группами должна быть скорректирована с учетом таких факторов, потенциально влияющих на исход (прогноз).

Таблица 2.

Оценка достаточности числа больных в группах сравнения: величина выборки в каждой из двух групп, необходимая для отклонения гипотезы о 25% разности частоты определенного исхода в группах (по [6] с изменениями)

Частота исхода в контрольной группе	Частота исхода в группе лечения									
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05
0,95	14	68								
0,9	11	38	105							
				7						
0,8		18	78							
0,7			11	28	165					
0,6				15	43	402				
0,5					20	65	160			
								7		
0,4								11	25	98
0,3								14	30	152
0,2								13	34	261
0,1								13	34	589
0,05								17	74	1245

Примечание. Результат эксперимента отражается в частоте определенного исхода (выздоровления, смерти, исчезновения клинических симптомов) в группе лечения и в контрольной группе. При оценке результатов различие между группами недостоверно. Таблица показывает минимально необходимое число больных в каждой группе, чтобы быть уверенными в том, что клинически значимое различие в 25% ($p < 0,05$; односторонний тест) не может быть случайно отвергнуто. Таблица предназначена только для оценки результатов исследования, а не для планирования исследований!

Таблица 7.

Оценка достаточности числа больных в группах сравнения: величина выборки в каждой из двух групп, необходимая для отклонения гипотезы о 50% разности частоты определенного исхода в группах (по [6] с изменениями)

Частота исхода в контрольной группе	Частота исхода в группе лечения										
	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,08	0,06	0,04	0,02
0,98	14	50	580								
			3	Наблюдаемое снижение риска ≥ 50% и не может быть отвергнуто независимо от численности группы							
0,95	12	37	921								
0,9	14	26	236								
0,8		15	63								
0,7		29	268								
0,6			17	68							
0,5			11	30	260						
0,4			16	62							
0,3	Менее 10 больных в группе	10	26	205							
0,2			20	145							
0,1			17	102	254	2017					
				7							
0,08			14	66	131	453					
0,06			12	44	76	179	1313				
0,04			10	31	47	87	274				
0,02			22	30	47	97	561				

Результат эксперимента отражается в частоте определенного исхода (выздоровлений, смерти, исчезновения клинических симптомов) в группе лечения и в контрольной группе. При оценке результатов различие между группами недостоверно. Таблица показывает минимальное необходимое число больных в каждой группе, чтобы быть уверенным в том, что клинически значимое различие в 50% ($p < 0,05$; односторонний тест) не может быть случайно отвергнуто. Таблица предназначена только для оценки результатов исследования, а не для планирования исследований!

Ичерпывающее описание всех 10 позиций редактора встречается даже в статьях, опубликованных в лучших журналах. Это не означает, что высокое качество статей недостижимо, просто редакторы сформулировали достаточно ясные и общие для всех журналов требования относительно недавно [7]. Поэтому при рассмотрении статьи нужно подходить к обнаруживаемым недостаткам клинического испытания взвешенно и попытаться ответить на 6 основных вопросов.

Было ли распределение больных по группам рандомизированным? Не следует доверять статьям об исследованиях, где распределение больных по группам было нерандомизированным (неслучайным) или метод распределения был неудовлетворительным, или вообще нет его описания. Утверждение о том, что больных распределяли по группам случайным образом, не дает оснований для полного доверия. Иногда авторы используют неудовлетворительные

методы распределения по группам (например, по номерам историй болезни — «четные—нечетные» или по дням недели). Изредка приходится встречаться с «рандомизацией», которая на поверку оказывается распределением на две группы по полу! В международных журналах сообщения о рандомизации приводятся в 9/10 статей о клинических испытаниях, но только в 1/3 статей уточняется метод рандомизации [9]. Если возникает сомнение в рандомизации — отбрасывайте статью.

Одна из главных целей рандомизации — подобрать группы сравнения, одинаковые по всем признакам, как по известным заранее, так и по неизвестным. В статье об исследовании должно быть показано, что сформированные группы действительно мало отличаются по известным важным признакам. Различия при рандомизированном формировании групп случайны, но если они существуют, то при анализе результатов обязательно должна быть сделана поправка на эти различия.

Причин отказа от рандомизации в клинических исследованиях много. Главная из них — недостаточная образованность врачей. Но есть еще ряд «поведенческих» причин [4, 14]:

- существующее лечение может поддерживать (или отвергать) авторитетными учеными или сложившейся традицией;

- отношения врача и больного формируют у врача желание именно помогать, а не испытывать новый вид лечения;

- комплекс «олимпийца» — нежелание подвергать свой авторитет, свой способ действий испытаниям;

- боязнь того, что при случайному распределении лечения больной не получит оптимальной помощи или лекарство может оказаться вредным (это заблуждение тоже от необразованности, так как именно этими опасениями и определяется необходимость проведения рандомизированных испытаний);

- неблагоприятное влияние результатов предыдущих плохо выполненных испытаний, формирующих ошибочные исходные ожидания.

Именно из-за плохо выполненных первичных клинических испытаний множество больных получало лечение, которое позднее, при тщательно проведенной проверке, оказалось неэффективным (например, перевязка внутренней грудной артерии, профилактика атеросклероза клофифбратором, лечебная бронхография, лечение язвы желудка замораживанием и т.д.). Колossalные средства уже затрачены и продолжают тратиться на поддержание никогда не испытывавшейся по-настоящему системы «диспансеризации». Список лечебных и профилактических методов, которые при проведении качественных клинических испытаний оказались бесполезными, можно продолжать на многих страницах.

Если не удается найти клинического испытания, выполненного на высоком методическом уровне, по интересующей Вас проблеме «семь раз» убедитесь, что такие испытания действительно не проводились.

Только потом переходите к методически менее совершенным исследованиям. Их всегда больше, они всегда менее убедительны. Глядя на обилие исследований, не выдерживающих критики, помните о законе Gresham: объем некачественных данных всегда имеет тенденцию к превалированию над доброкачественными сведениями.

Читая статьи, постоянно оценивайте, нельзя ли приведенные результаты объяснить видимыми и возможными погрешностями, увеличением исследователей и т.п. Не стесняйтесь критического анализа — за Вами стоит интересы Ваших больных! Если Вам не удалось найти убедительного ответа на вопрос, не смущайтесь, неопределенности. Ведь Вы каждый день принимаете ответственные клинические решения в неопределенных ситуациях. При использовании результатов несовершенных исследований существует способ взвесить вероятность ошибки (анализ чувствительности к смещениям). Он не слишком сложен, но выходит за пределы нашего обсуждения [13].

Ориентируйтесь на основной критерий оценки эффективности лечения. Так, при диабете лечение не может быть оценено вполне по стойкости и величине гипогликемического эффекта; нужно знать, как оно влияет на уровень смертности. Самый блестящий гипогликемический эффект перечеркивается даже небольшим повышением смертности. Как бы ни снижалась сердечно-сосудистая смертность, но если уровень общей смертности повышается, значит, основной результат не достигнут. К сожалению, подмена «конечных» результатов (истинных критериев оценки — клинических исходов) «промежуточными» (косвенными критериями оценки вроде снижения уровня глюкозы или холестерина в крови, артериального давления) весьма распространена. Часто исследование действительно не выявляет различий между сравниваемыми методами лечения. Если это исследование отвечает критериям качества, то оно может быть полезным для нас уже тем, что позволит не вводить в свою практику новый метод лечения.

В качественном исследовании обязательно должны быть отслежены конечные результаты. Без них обоснование метода лечения является неполноценным. Из истории известен пример, когда американский врач Б. Раш лечил желтую лихорадку кровопусканием. Он лечил по 100 человек в день, и ни один не умер в процессе лечения. Однако отдаленных результатов никто не наблюдал, что позволило Б. Рашу длительно отстаивать свой метод лечения.

Значимы ли результаты исследования? Достоверны ли они статистически? Не представляет ли автор статистически достоверные, но небольшие различия как клинически значимые? Статистически значимо то, что действительно происходит с высокой вероятностью. Клинически значимо то, что своими размерами (например, величиной снижения смертности) убеждает врача в необходимости изменить свою практику в пользу нового образа действий. В клинических испытаниях, проводимых на большом числе больных, можно статистически достоверно выявить достаточно небольшие эффекты от приме-

нения изучаемых лекарственных препаратов. Особенно часто небольшие различия выявляются при мета-анализе, т.е. при объединении данных множества клинических испытаний. Мета-анализ очень важен для обобщения результатов таких испытаний. Рассматривая статью с мета-анализом, обращайте внимание не только на статистическую достоверность различий между группами, получавшими разное лечение, но и на размер эффекта. Малый размер эффекта означает, что лишь очень небольшая часть больных (например, 1%) получает пользу от такого лечения [8]. Размер эффекта может быть выражен различным образом. Обратим внимание на популярную сейчас меру эффективности вмешательства — число больных, которых надо пролечить, чтобы получить эффект у одного дополнительного больного (number needed to treat — NNT). Это, пожалуй, самая очевидная мера эффективности лечения при его сравнении с другим методом лечения.

Если различия в эффективности изучаемых методов лечения недостоверны, следует проверить, достаточно ли число включенных в исследование? Недостаточное число больных делает отрицательный результат (ожидаемого эффекта не найдено) неубедительным, недостаточным для негативной оценки лечения, а в случае положительного результата (подтвержден ожидаемый лечебный эффект) делает неопределенной величину полученного эффекта. Бесполезность исследований на малом числе больных делает их неэтичными [2].

Дополнительно оцените результаты с точки зрения больного. Ведь для многих часто не так уж важно небольшое снижение смертности, но очень привлекательно качество жизни, особенно в ближайший период. Лечение может быть травматичным, неприятным, болезненным, неэстетичным, противоречивым религиозным установкам и это может быть решающим основанием для отказа от него.

Доступен ли метод в Вашей практике? Для этого нужно убедиться в том, что метод хорошо описан и его можно воспроизвести. Уясните его биологические и клинические обоснования, механизм действия. Оцените еще раз, избежал ли исследователь неконтролируемого, неизвестного лечения у больных контрольной группы и побочного, дополнительного и неконтролируемого лечения у больных, получавших изучаемый препарат. Неконтролируемое лечение в контрольной группе почти неизбежно и приводит к уменьшению различий между сравниваемыми группами. Дополнительное лечение в опытной группе тоже почти неизбежно и может исказить размер эффекта. Главное средство избежать различий между группами — высокое качество организации «слепого» метода назначения лечения.

Может возникнуть впечатление, что приведенные критерии качества относятся только к одному виду клинических испытаний — собственно испытаниям методов лечения и не касаются многочисленных исследований, посвященных изучению побочных положительных и негативных эффектов, оптимизации лечения, комбинированного лечения и прочих детализирующих исследований. Это ложное впечатление. Все исследования эффектов ле-

чебного вмешательства должны проводиться по схеме контролируемого клинического испытания, независимо от того, считаются ли эти эффекты главными или второстепенными, считается ли лечение

основным или дополнительным. Отход от стандарта качества исследования неминуемо приводит к получению некачественных данных, которые Вам следует игнорировать.

Литература

1. *Бащинский С.Е.* Evidence-based medicine и Международный журнал медицинской практики. МЖМП 1996;1:6–11.
2. *Власов В.В.* Международные этические требования к медицинским исследованиям с участием человека. Тер арх 1996;1:58–64.
3. Применение селегилина при болезни Паркинсона (Письмо в редакцию с комментарием). МЖМП 1997;3:63–4.
4. *Chalmers T.C.* Ethical aspects of clinical trials. Am J Ophthalmol 1975;79:753–8.
5. *Der Simonian R., Charette L.J., McPeck B., Mosteller F.* Reporting on methods of clinical trials. N Engl J Med 1982;306:1332–7.
6. *Detsky A.S., Sackett D.L.* When was a “negative” clinical trial big enough? How many patients you needed depends on what you found. Arch Int Med 1985;145:709–12.
7. *CONSORT Group.* Improving the quality of reporting of randomised controlled trials: the CONSORT statement. JAMA 1996; 276:637–9.
8. *Feinstein A.R.* Meta-analysis: Statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71–9.
9. *Gotzsche P.C.* Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroid antiinflammatory drugs in rheumatoid arthritis. Contr Clin Trials 1989;10:31–56.
10. *Guyatt G.H., Sackett D.L., Cook D.J.* User's guides to the medical literature: II. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993;270:2598–601.
11. *Guyatt G.H., Sackett D.L., Cook D.J.* User's guides to the medical literature: II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients. JAMA 1994;271:59–63.
12. *Kleijnen J., ter Riet G., Knipschild P.* Acupuncture and asthma: a review of controlled trials. Thorax 1991;46:799–802.
13. *Rosenbaum P.R.* Discussing hidden bias in observational studies. Ann Int Med 1991;115:901–5.
14. *Spodick D.H.* Randomized controlled clinical trials: the behavioral case. JAMA 1982;247:2258–60.
15. *ter Riet G., Kleijnen J., Knipschild P.* Acupuncture and chronic pain: a criteria-based meta-analysis. J Clin Epidemiol 1990;43:1191–9.
16. *Vincent C., Richardson P.H.* The evaluation of therapeutic acupuncture. Pain 1986;24:1–13.