

КАК ОРГАНИЗОВАТЬ СОБСТВЕННЫЕ ИССЛЕДОВАНИЯ И НЕ ЗАПУТАТЬСЯ В ИНТЕРПРЕТАЦИИ ЧУЖИХ (часть 2)

Н.В. Жуков
ФГУ ФНКЦ
ДГОИ Росздрава

Исследования II фазы

В случае большого количества доступных пациентов (распространенные заболевания, нежесткие критерии отбора) и большого количества методик для тестирования можно использовать преимущества рандомизированных исследований II фазы [9]. Несмотря на то что исследования II фазы по определению не предусматривают сравнения с ранее существовавшими стандартами, возможно проведение рандомизированного сравнения между различными режимами с использованием исследуемого препарата (препаратов). Такой подход позволяет выбрать наиболее многообещающий режим для дальнейшего исследования III фазы. В табл. 2 показано необходимое количество пациентов в каждой лечебной группе для уверенности в том, что будет выявлен наилучший режим использования препарата, если его истинная эффективность действительно превышает режим сравнения на 10%. Уровень 10% выбран как наименьшее из клинически значимых различий в эффективности двух режимов лечения, одинаковых по цене и токсичности.

Таблица 2. *Необходимое число пациентов в каждой сравниваемой группе для выбора наиболее эффективного лечения в случае, если истинное различие в частоте ремиссий составляет 10%*

Исходная вероятность ремиссии*, %	85% вероятность выявления наилучшего режима	90% вероятность выявления наилучшего режима
5	20	29
10	28	42
15	35	53
20	41	62
30	49	75
40	54	82
50	54	82
60	49	75
70	41	62
80	28	53

* Определяется на основании предыдущих исследований II фазы с этим же препаратом, но в одном из исследуемых режимов.

Исследования II фазы комбинированных режимов

В настоящее время генерируется масса так называемых исследований II фазы комбинированных режимов. К сожалению, их целесообразность далеко не всегда ясна. Одной из оправданных целей может служить доказательство того, что комбинация является переносимой и приемлемой при проведении на базе нескольких клиник до того, как будет адаптирована для исследований III фазы. Достижение этой цели обычно не требует включения большого количества пациентов. Другая оправданная цель — определить, является ли режим достаточно эффективным, чтобы оправдать переход в III фазу. Достижение этой цели требует тщательного планирования. К сожалению, многие из исследований II фазы данного типа планируются и анализируются недостаточно адекватно для того, чтобы дать реальную научную или практическую информацию.

Исследователи (в том числе и зарубежные) часто не делают различий между исследованиями II фазы комбинации цитостатиков и аналогичных исследований одного нового агента. В результате протоколы часто пишутся для выявления разницы между минимальной активностью (т.е. $p_0=5-10\%$) и некоторым незначительным уровнем активности (т.е. $p_1=20-25\%$). Разумеется, если новый препарат в таком исследовании комбинируется с цитостатиками, уже показавшими активность при данном виде опухоли, то результаты такого исследования бессмысленны. Если целью исследования является частота ремиссий, то уровень отсутствия интереса p_0 должен быть, как минимум, не меньше уровня активности старого цитостатика (цитостатиков), включенного в режим, или наиболее активных комбинаций, используемых при данном заболевании. Это абсо-

лотно логично, так как новый режим заслуживает внимания, только если его предполагаемая активность превосходит ранее имевшиеся режимы или близка к ним. Казалось бы, дальше все просто, необходимо обратиться к описанным ранее таблицам и, зная p_0 и желаемый уровень p_1 , определить необходимое число больных. Однако одна существенная проблема сохраняется и при таком подходе. Как выбрать осмысленный уровень p_0 для использования в таких исследованиях? Учитывая, что заключение таких исследований II фазы (исследований комбинаций) является сравнительным, нужно, чтобы сравнение проводилось с прогностически одинаковой группой больных, получавших стандартное лечение. Следовательно, планирование таких исследований должно включать проспективную (т.е. проведенную заранее, до начала исследования) идентификацию группы сравнения. Проспективное выявление группы сравнения необходимо по двум условиям: определение уровня p_0 и предотвращение в дальнейшем неблагоприятных манипуляций (подбор группы сравнения под уже полученные данные проспективного исследования). Хотя такой исторический контроль и недостаточен для того, чтобы устранить потребность в исследованиях III фазы, однако он позволяет при тщательном исполнении получить наиболее достоверное заключение о том, какой из режимов заслуживает продолжения исследований в III фазе. Размеры выборки для сравнительных исследований по частоте ремиссий с использованием исторического контроля были предложены R. Makuch и R. Simon [10] и частично приведены в табл. 3. Данная таблица предусматривает достижение 80% силы исследования и односторонней статистической значимости $p < 0,05$.

Если режим, использовавшийся у 50 больных в группе исторического контроля, показал частоту ремиссий 30%, то для выявления режима, имеющего эффективность на 20% больше (целевой уровень ремиссий при использовании нового режима 50%), новым режимом должно быть пролечено 69 больных¹. В случае, если в группе исторического контроля имеется 100 подходящих больных, то проспективная группа должна состоять только из 48 пациентов. С другой стороны, если имеется только 20 больных в группе исторического контроля, то необходим

набор 882 пациентов в группу экспериментального лечения. В случае, если уровень достижения ремиссий при стандартном лечении неясен (в связи с ограниченностью числа пациентов исторического контроля), проведение таких исследований II фазы малоинформативно. В такой ситуации более результативным и эффективным является проведение рандомизированного исследования II—III фазы (см. далее) по сравнению со стандартным режимом.

Отсутствие контрольной группы или использование недостоверного контроля делает проведение исследований комбинированных режимов во II фазе бессмысленным, так как они не несут никакой полезной информации (не отвечают на главный вопрос исследований II фазы — перспективен ли режим для изучения в III фазе — или отвечают на этот вопрос «наугад»). В медицинской литературе существует множество примеров, когда по результатам исследований II фазы комбинированный режим «признавался» эффективным, однако в дальнейшем таковым не оказывался. Такие исследования зачастую не являются и предшественниками исследований III фазы, а проводятся лишь ради возможности опубликовать «обнадеживающие результаты исследования II фазы» (исследования ради исследования).

Описаны многочисленные примеры (особенно в некоторых нозологических группах — рак молочной железы, яичников и др.), когда комбинированные режимы, позволявшие достигать очень высокой частоты ремиссий в исследованиях II фазы, в дальнейшем не приводили к улучшению выживаемости. Это обусло-

Таблица 3. Число больных в основной группе, необходимое для выявления различий в непосредственной эффективности лечения

Частота ремиссий в группе исторического контроля, %	Необходимо выявить различие в эффективности в л%	Больных в группе исторического контроля			
		20	40	50	100
10	15	>1000	108	80	50
	20	116	40	35	27
	30	22	15	14	13
20	15	>1000	285	167	83
	20	385	67	55	40
	30	31	21	19	17
30	15	>1000	554	259	108
	20	882	87	69	48
	30	31	24	22	19
40	15	>1000	699	303	120
	20	913	92	74	52
	30	36	24	22	20
50	15	>1000	538	267	115
	20	455	83	68	50
	30	30	22	20	18
60	15	>1000	295	185	97
	20	179	63	55	42
	30	22	17	17	15

¹Для удобства строка выделена в таблице полужирным шрифтом.

вило интерес к исследованиям II фазы, ставящим в качестве основной цели оценку выживаемости. Такой подход получил распространение и при некоторых типах опухолей, оценка ремиссии при которых представляется затрудненной, так как большинство больных не имеет измеряемых очагов (например, опухоли головного мозга, яичников, рак желудка и предстательной железы). Таблицы для планирования исторически контролируемых исследований II фазы, предусматривающих сравнение выживаемости или времени до прогрессирования, разработали D. Dixon и R. Simon [11].

Исследования III фазы

Хорошее клиническое исследование задает важные вопросы и получает на них правильные ответы. Как было сказано выше, многие исследования II фазы не способны дать правильные ответы. К сожалению, некоторые исследования III фазы не задают важных вопросов: наиболее важные клинические исследования чаще всего наиболее трудны для организации. Такие исследования могут ставить под сомнение методы лечения, ставшие уже догматическими, «перераспределять» право лечить пациентов между специалистами (раньше это был «хлеб» хирургов, а теперь — химиотерапевтов). Кроме того, результаты таких исследований могут потенциально привести к стандартизации лечения среди врачей, каждый из которых уверен, что его путь лучший. Одной из основных задач исследований III фазы является создание прямых рекомендаций для практикующих врачей в отношении лечения конкретных пациентов (по результатам исследования, такой-то вид лечения признается более эффективным у данного контингента больных, т.е. рекомендуется для применения в практике).

Учитывая, что результаты данных исследований являются «последним шагом» препарата или методики перед внедрением в широкую клиническую практику, конечными целями исследований должны являться показатели, напрямую определяющие выигрыш пациента от проводимого лечения. Такими важнейшими для больного конечными целями лечения являются выживаемость и контроль симптомов. К сожалению, контроль симптомов редко может быть использован в качестве основных целей исследования, так как на этот показатель могут повлиять разнообразные факторы, включая сопутствующую сопроводительную терапию. Так, любимая многими отечественными исследователями непосредственная противоопухолевая эффективность (сокращение размеров опухоли) не является приемлемой целью исследования III фазы, поскольку может иметь небольшую связь с выигрышем пациента, а может и вообще ее не иметь. Так, в большом метаанализе V. Torgi и соавт. [12] был поставлен вопрос, насколько часто увеличение частоты ремиссий, достигаемое при применении новых режимов, «транслируется» в увеличение выживаемости при диссеминированном раке яичников.

Оказалось, что даже режимы, приводившие к значительному увеличению частоты ремиссий в рандомизированных исследованиях, практически не изменяли медиану выживаемости больных в исследованиях III фазы. Таким образом, использование частоты ремиссий в качестве конечной цели исследования приводило к назначению больным более токсичной, интенсивной и дорогостоящей терапии при небольшом (или отсутствующем) влиянии на выживаемость. Такую ситуацию легко представить: новое «более эффективное» лечение приводит к уменьшению асимптоматического метастаза в печень с 5 до 1 см, но не влияет на продолжительность жизни пациента. Приносит ли это пользу больному? К счастью, результаты отечественных исследований, организованных по принципам, описанным в начале статьи, редко переносятся в клиническую практику. Однако выводы западных исследований, ориентированных на частоту ремиссий (чаще всего II фазы), переносятся в клиническую практику врачей в РФ достаточно часто.

Крайне важной является также возможность переноса результатов исследований III фазы на общую популяцию больных «за пределами» исследовательских центров и рамок исследовательских протоколов. Во многом это обеспечивается проведением многоцентровых исследований, предусматривающих лечение на различных лечебных базах. Критерии отбора также во многом влияют на возможность переноса полученных выводов на общую популяцию больных. Исследования, использующие узкие (строгие) критерии отбора, обычно гораздо менее воспроизводимы в общей практике. Попытка переноса результатов исследования режима X, который использовался у абсолютно «сохранных» молодых больных в рамках клинического исследования с «узкими» критериями отбора, на 65-летнего больного с ИБС может закончиться плачевно.

Рандомизация

Если новый препарат (или вид лечения) позволяет излечить пациентов с заболеванием, которое до этого в короткие сроки приводило к гибели всех больных, то для осмысленного и достоверного вывода достаточно и исторического контроля.

Однако за исключением этих абсолютно однозначных ситуаций подбор адекватного нерандомизированного контроля для исследований остается крайне сложным. В исследованиях с нерандомизированным контролем чаще всего диагностические процедуры, стадирование, сопроводительное лечение, последующее наблюдение и многие другие факторы значительно различаются между контрольной и исследуемой группами. Выбор лечения в контрольной группе обычно основывается на предпочтениях врача, проводящего своеобразный отбор пациентов. При формировании основной (лечебной) группы в нее не включают больных, заведомо неспособных перенести ис-

следуемое лечение; более того, перед окончательным анализом зачастую из исследуемой группы исключают больных, не получивших исследуемое лечение в полном объеме, отказавшихся от терапии на одном из ее этапов или имеющих значительные отклонения от условий протокола. С другой стороны, группа исторического контроля многих таких пациентов как раз и включает. Все это приводит к различному распределению известных и неизвестных прогностических факторов в контрольной и исследуемой группах. К сожалению, в большинстве случаев информации для выявления этих различий недостаточно, так как даже доказанные прогностические факторы могут быть не оцениваемы у больных группы исторического контроля.

Перспективное формирование контрольной группы путем рандомизированного назначения лечения позволяет избежать большинства незапланированных влияний, перечисленных выше, так как отбор пациентов в обе группы происходит на одинаковых условиях. Разумеется, рандомизация не гарантирует, что исследование включит репрезентативную выборку всех пациентов с данным заболеванием, однако помогает быть уверенным в том, что различие в результатах лечения будет обусловлено именно лечением, а не другими влияющими факторами.

Иногда утверждается, что рандомизация может не использоваться, поскольку существует метод парного подбора больных для группы исторического контроля (*matched pair analysis*). Однако сопоставимость больных из лечебной и контрольной групп может быть достигнута только по ограниченному числу известных прогностических факторов [13]. Такой подбор практически невозможен даже для всех известных прогностических факторов и тем более не гарантирует от непропорционального распределения неизвестных (неоцениваемых) факторов. Нередко считается, что и рандомизация неэффективна в связи с отсутствием уверенности в том, что группы идентичны в отношении неизвестных прогностических факторов, за исключением ситуаций, когда количество пациентов крайне велико. Это правильно, однако отражает недопонимание целей рандомизации. Рандомизация не проводится с целью создать абсолютно одинаковые по прогнозу группы, однако она группирует неизвестные и известные влияющие факторы в соответствии с правилами случайного распределения. Именно такое распределение позволяет с успехом применять статистические тесты (определение статистической значимости и доверительного интервала), показывающие вероятность того, что именно лечение, а не распределение прогностических факторов влияет на исход болезни. Без рандомизированного распределения применение этих же тестов не позволяет судить о том, чем обусловлена разница в результатах: эффектом лечения или различием в распределении неизвестных прогностических факторов.

Нерандомизированное распределение (как менее «затратное») может быть использовано в более «ранних» исследованиях (II фаза) и для оценки эффективности лечения больных с однозначно плохим прогнозом. Для главных вопросов, связанных с выработкой тактики и стратегии здравоохранения, за исключением ситуаций, когда эффект исследуемого лечения огромен, необходимо проведение рандомизированных исследований.

Рандомизация должна проводиться после того, как пациент признан подходящим по критериям отбора, дал согласие на участие в исследовании (в том числе и на сам процесс рандомизации). Предпочтительной является истинная рандомизация, на ход которой не может повлиять лечащий врач или клиника (централизованная компьютерная рандомизация, конвертная рандомизация и т.д.). Рандомизация по номеру истории болезни, первой букве фамилии и т.д., т.е. по факторам, которые известны врачу (и соответственно могут быть использованы для того, чтобы повлиять на ход рандомизации) неприемлема. Конвертная рандомизация проста и легко может быть использована во «внутренних» исследованиях. Проведение адекватной рандомизации (в том числе и конвертной) всегда придает исследованию больший «вес» и увеличивает степень его доказательности (а также возможность опубликовать его результаты в солидном зарубежном журнале). Например, можно пролечить 100 больных с определенным диагнозом за год и потом искать группу исторического контроля. На выходе получится исследование, которое в лучшем случае будет принято в качестве абстракта на не очень престижной конференции. Можно же потратить на исследование 2 года и 200 конвертов, но в итоге получить исследование с гордым названием «рандомизированное». Статус такого исследования при попытке опубликовать его в зарубежной печати гораздо выше, и что самое главное, его результаты будут гораздо более значимыми, чем просто публикация. Их можно будет использовать для выработки осмысленной тактики лечения, не приводя бесконечные оправдания: требуется подтверждение исследования, возможно использование данных после ...

Стратификация

Если при заболевании, на материале которого планируется исследование, существуют важные прогностические факторы, необходимо проведение стратифицированной рандомизации (для создания баланса между группами по известным прогностическим факторам). Обычно это реализуется через создание отдельного листа рандомизации (или набора карточек в заклеенных конвертах) для каждой группы стратификации. Такой подход позволяет сбалансировать группы по наиболее важным прогностическим факторам, и после включения блока из 4—10 пациентов в лечебной и контрольной группах будет присутство-

вать одинаковая доля пациентов из каждого стратума. В пределах стратума назначение лечения является рандомизированным. Факторы, используемые для стратификации, должны быть известны для каждого пациента до момента рандомизации. Например, при раке молочной железы рецепторный статус имеет явное прогностическое влияние. Если рандомизировать больных на проведение определенного лечения без учета этого фактора, при окончательном анализе может оказаться, что в основной группе доля больных с рецепторотрицательными опухолями составит 50%, а в контрольной — 25%. Сопоставление результатов лечения таких групп весьма затруднительно. Многие клинические исследования в настоящее время используют адаптивный метод стратификации, позволяющий эффективно балансировать группы по многим прогностическим факторам. Однако такой метод стратификации обычно требует достаточно сложного компьютерного планирования. Проведение стратификации позволяет сбалансировать группы. Отдельно (в пределах одного рандомизационного листа или набора конвертов) будут рандомизироваться больные с рецепторположительными опухолями, отдельно — с рецепторотрицательными. В итоге в обеих группах (экспериментальной и контрольной) окажется одинаковая пропорция больных с рецепторотрицательными опухолями.

Предпочтительным является ограничение стратифицирующих факторов только теми, которые оказывают значимое и независимое влияние на исход лечения. Слишком большое «дробление» группы значительно увеличивает число больных, необходимое для завершения исследования. Если прогностические факторы имеют четкую корреляцию, для стратификации должен использоваться только один из них. Некоторые авторы считают, что проведение стратификации является ненужным осложнением, так как исправление дисбаланса известных факторов можно провести и при окончательном анализе. Однако это верно лишь для очень больших исследований. Стратификация также может оказаться весьма полезна при необходимости проведения промежуточного анализа, когда размер выборки еще невелик, однако необходимо принятие решения о продолжении/приостановке исследования. Кроме того, стратификация позволяет априори определить прогностические факторы, которые важны для пациентов, участвующих в исследовании. Последующие анализы «по подгруппам» могут быть проведены именно в пределах стратума.

ЛИТЕРАТУРА

9. Simon R., Wittes R.E., Ellenberg S.S. Randomized phase II clinical trials. *Cancer Treat Rep* 1985;69:1375.
10. Makuch R.W., Simon R. Sample size considerations for nonrandomized comparative studies. *J Chronic Dis* 1980;33:171.

11. Dixon D.O., Simon R. Sample size considerations for studies comparing survival curves using historical controls. *J Clin Epidemiol* 1988;41:1209.
12. Torri V., Simon R., Russek-Cohen E. et al. Relationship of response and survival in

advanced ovarian cancer patients treated with chemotherapy. *J Natl Cancer Inst* 1992;84:407.

13. Simon R. The importance of prognostic factors in cancer clinical trials. *Cancer Treat Rep* 1984;68:185.

Размер выборки

Протоколы исследования III фазы должны определять требуемое количество пациентов для набора и длительность периода наблюдения (follow-up) после закрытия набора больных. На основании этих параметров и определяется время, необходимое для проведения окончательного анализа. Методы планирования размера выборки базируются на предположении, что к окончанию периода наблюдения будет возможно проведение статистически значимого анализа, сравнивающего экспериментальное лечение с контрольным. Уровень статистической значимости в 0,05 имеет следующее значение: если между исследуемыми методами лечения нет истинного различия в эффективности, то вероятность увидеть различие в результатах такой величины, как наблюдается в данном исследовании, составляет 0,05. Проще говоря, если в исследовании выявлена разница в выживаемости в 20% при $p=0,05$, то шанс, что различие подобного уровня будет получено «случайно», а не в результате истинного различия между методами лечения, составляет 5%. Уровень статистической значимости в 0,05 не означает, что вероятность равенства эффективности исследуемого и контрольного метода лечения (вероятность верности нулевой гипотезы) равняется 0,05.

Если в исследование включено незначительное число пациентов, различие в наблюдаемых результатах должно быть крайне велико для того, чтобы получить статистическую достоверность. С другой стороны, вероятность получить статистически значимые результаты может быть крайне мала, даже если наблюдается истинное (но меньшее) различие в эффективности лечения. Вероятность получить статистически значимый результат в случае, если исследуемые методы лечения действительно различаются по эффективности, называется статистической силой исследования. Статистическая сила исследования возрастает при увеличении размера выборки и удлинении интервала наблюдения. Однако в наибольшей степени сила исследования зависит от величины истинного различия эффективности между двумя методами лечения. В основном при выборе размера выборки ориентируются на статистическую силу исследования в 0,80 или 0,90 при условии, что истинное различие в эффективности, которое необходимо обнаружить, является наименьшим из признанных важным по медицинским соображениям (клинически значимым).