

Планируем клиническое исследование.

Вопрос №1: Как определить необходимый объем выборки?

Г. П. Тихова

Карельский научный центр РАН, 185910, Петрозаводск

Planning clinical research. Question #1: How to calculate enough sample volume?

G. P. Tikhova

FSBI "Karelian Research Center of Russian Academy of Science", 185910, Petrozavodsk

В статье рассмотрены важнейшие вопросы планирования клинического исследования и эксперимента, в частности, расчет объема выборки, достаточного для проверки статистической достоверности различия на заданном уровне значимости с принятой мощностью критерия. Детально описаны методы и формулы расчета для различных типов данных. Приведены примеры расчета объема выборки для числовых и бинарных показателей, принятых в качестве конечных точек исследования. *Ключевые слова:* клиническое исследование, клиническое испытание, объем выборки, уровень значимости, мощность критерия.

The article is dedicated to the important questions of clinical study design in particular sample size calculation enough for statistical significance testing with determined significance level and power. Calculation techniques and formulas are described in details for different data types. The paper includes examples of sample size calculating for continuous and binary variable assigned as study endpoints. *Key words:* clinical study, clinical trial, sample size, significance level, test power.

Чтобы добиться содержательных результатов, клиническое исследование должно быть тщательно спланировано, а его протокол детально проработан. В особенности это касается клинических испытаний, которые проводятся для проверки эффективности каких-либо методов лечения и включают этап статистической обработки данных. В этом случае содержательные заключения об эффективности делаются на основе статистических выводов о достоверности межгруппового различия, поэтому достижение статистической значимости этого различия становится одной из основных целей проводимого исследования. Если оно не достигнуто в результате пилотного клинического исследования, которое обычно является наблюдательным, это еще допустимо, в конце концов, результатом в данном случае можно считать и обнаруженные тенденции. Эти тенденции однако ожидают объективного подтверждения или опровержения, для чего необходимо проводить клиническое испытание по всем канонам жанра, с рандомизацией, определением значения клинически значимой разницы и статистически значимым подтверждением межгруппового различия. Если дизайн такого испытания имеет изъяны, не позволяющие статистически достоверно подтвердить или опровергнуть предположение об эффективности (или различии), это уже

настоящий провал, попусту потраченное время и силы. Именно поэтому при планировании клинического испытания (эксперимента) необходимо уделять большое внимание подготовительной части: однозначно и четко описать конечные точки и точно спроектировать дизайн исследования. Собственно, эта пара в основном и определяет протокол, в соответствии с которым будет осуществляться сбор данных в ходе исследования. Дизайн исследования включает много аспектов. Приведем примерный список вопросов, на которые исследователю будет полезно найти максимально точные ответы при планировании своего исследования:

- Как исследовательская проблема может быть сформулирована в практической плоскости?
- Как (в каких величинах) будет измеряться исход исследования? Будет ли он объективным? Будет ли он воспроизводимым?
- Какой количественной величиной будут выражаться взаимосвязи (взаимозависимости)?
- Будет ли планируемое исследование экспериментом или наблюдением?
- Если оно будет носить экспериментальный характер, можно сделать его слепым? Будет ли проведена рандомизация? Как будет сформирована контрольная группа и какова ее природа (плацебо, активный/пассивный контроль)?

- Если исследование не является экспериментом, будет ли сбор данных проспективным или ретроспективным? Будет ли выборка формироваться естественным образом, в виде когорты или в формате «случай – контроль»? Как вы предполагаете бороться с вмешивающимися факторами?

- Какой минимальный объем выборки надо обеспечить, чтобы достичь поставленной цели исследования, в частности, добиться определенности статистического вывода о достоверности различия или эффекта?

Последний вопрос о необходимом объеме выборки на первый взгляд кажется самым важным, хотя на самом деле на него невозможно ответить однозначно, не получив конкретные ответы на все предыдущие вопросы. Действительно, чтобы успешно провести расчеты необходимого объема выборки, мы должны четко определить конечные точки исследования, гипотезы и статистические тесты. А для определения этих трех позиций мы должны в свою очередь разобраться с типами наших данных, количеством исследуемых групп и целью нашего исследования, т.е. хотим ли мы проверить эффективность вмешательства или изучить взаимосвязь каких-то показателей. От этого будет зависеть выбор статистического метода тестирования и/или описания наших результатов.

Начнем с выяснения типа исследуемого показателя или конечной точки.

По типу принимаемых значений исследуемые (регистрируемые) показатели делятся на 4 основных класса:

- количественные, или числовые (continuous, numeric);
- порядковые, или ранговые (ordinal);
- качественные (categorical);
- бинарные, или дихотомические (binary, trigger, dummy, dichotomous).

К количественным (числовым) величинам относятся все измеряемые показатели и параметры, значения которых выражаются в числах. Это, например, практически все показатели клинического или биохимического анализа крови, показатели центральной гемодинамики, такие как систолическое и диастолическое артериальное давление, ЧСС и т.п. Количественными могут считаться и некоторые рассчитываемые оценки риска, такие как APACHE, TISS и другие подобные шкалы. Но, например, шкала Глазго может считаться количественным показателем только в том случае, если во внимание принимаются собственно рассчитанные баллы, а не клиническая интерпретация по ним глубины комы. Иными словами, все показатели, значения которых заключены в некоторый

числовой интервал (не обязательно непрерывный) являются количественными (числовыми).

Если мы исследуем какой-то симптом или признак, по которому можем сравнить наших пациентов в терминах «более выраженный, менее выраженный» или «легкий, средний, тяжелый», то этот признак является порядковой величиной. Кардинальное отличие числового показателя от порядкового состоит в том, что для первого мы всегда можем точно указать в некоторых единицах, насколько его значение у одного пациента больше/меньше, чем у другого, и эта разность имеет смысл и равна определенному числу указанных единиц. В случае порядкового признака мы можем констатировать лишь сам факт «больше/меньше», но точную разность указать не можем, или же она вообще не имеет смысла. Для проведения статистического анализа градации порядкового признака, как правило, обозначают целыми числами от 0 (полное отсутствие признака) до целого, равного общему количеству градаций. Эти числа называют рангами, поэтому такие величины еще называются ранговыми, и методы, применяемые к ним, также называются ранговыми критериями. Например, преэклампсия может быть легкой, средней и тяжелой степени. Если обозначить эти степени 1, 2, 3, а отсутствие преэклампсии – 0, то это будет порядковая величина. Другой пример – ожирение от 1-й до 4-й степени, также порядковый показатель. Иными словами, в том случае, когда можно сказать, что у этого пациента данный симптом или заболевание более выражено или имеет более тяжелую степень, чем у другого, или вы можете упорядочить всех пациентов вашей выборки по степени выраженности этого признака (симптома, заболевания, осложнения), то данный признак является порядковым.

Качественный показатель немного похож на порядковый, но у него есть важнейшее отличие: о нем невозможно сказать «больше/меньше», у него нет степеней выраженности, по нему невозможно упорядочить пациентов в исследуемой выборке. Например, в анестезиологической карте врач отмечает сопутствующие патологии больного, выбирая из ограниченного списка: сердце, легкие, печень, почки и т.д. Если у одного больного в качестве сопутствующей патологии отмечено заболевание сердца, а у другого – легочная патология, то упорядочить их по этому признаку не получится, потому что эти патологии совершенно различны, и нельзя сказать, какая из них более патология, а какая менее, это просто бессмысленно. Качественные признаки очень часто являются факторами, по которым производят стратификацию выборки и анализ межгрупповых различий.

Наиболее очевидной и понятной является бинарная величина, т.е. показатель, который может принимать только два возможных значения – 0 или 1 в математической записи, а в содержательном смысле это могут быть любые два возможных варианта откликов: да/нет, есть/нет, легкий/тяжелый, норма/отклонение и т. д. Эти случайные величины имеют целый класс собственных методов статистического анализа, среди которых столь популярные в настоящее время относительный риск, отношение шансов, чувствительность/специфичность и т. п.

Есть еще один тип исследуемых данных – это так называемые кривые дожития. Эти данные по сути дела являются временными рядами и включают в себя помимо значения исследуемого показателя еще и время измерения (или регистрации) этого показателя относительно некоторого начала отсчета времени (нулевой точки временной шкалы). Они требуют более сложных методов обработки, и в данной публикации мы не будем их рассматривать.

До 90% всех клинических исследований имеют дело с числовыми данными и бинарными показателями, которые обычно выражаются в частоте появления какого-либо события (осложнения, симптома, исхода), поэтому на примерах этих двух типов данных мы и разберем, как проводить расчет требуемого объема выборки.

Возьмем для этого самый простой и наиболее частый случай, когда в ходе исследования набираются 2 группы пациентов и сравниваются между собой по некоторому статистическому параметру, например групповому среднему значению (для числовых показателей) или по частоте исхода (для бинарного признака).

Итак, если мы взялись за расчет требуемого объема выборки, нам обязательно надо иметь некоторые предположения о том, какой приблизительно результат мы ожидаем получить или какой результат мы будем считать клинически значимым (о значении терминов «статистическая и клиническая значимость» см. предыдущие публикации [1-3]). Например, мы хотим сравнить два вида лечения и решить, какой из них более эффективен. Для этого нам надо выбрать конечную точку, т. е. показатель или бинарный признак, который мы будем регистрировать у всех больных в одной и второй группе, а затем решить, какой статистический параметр мы выберем для сравнения этих двух групп. Обычно в роли такого параметра выступает среднее значение показателя – индикатора эффективности, но сейчас наряду с ним все чаще используется медиана или другие статистические величины, также отражающие средние тенденции

групповой выборки. Например, мы традиционно решили сравнить две группы по среднему значению некоторого показателя. Отлично, но что значит сравнить? Это значит, что мы рассчитаем среднее значение в одной и второй группе и посмотрим, насколько сильно они друг от друга отличаются. В этом месте нашего планирования мы должны задержаться и решить (до начала самого первого измерения!), какое различие мы можем считать клинически значимым, конкретно – какая разность между средними значениями регистрируемого показателя будет указывать нам на то, что эффективность двух исследуемых видов лечения различна. Эту разность в расчетах объема выборки принято называть разностью эффекта или дельтой эффекта и обозначать латинской буквой **d** (в некоторых публикациях ее обозначают греческой Δ). Это один вариант предположения. Второй обычно имеет место в том случае, если для сравнения эффективности двух методов уже проводилось пилотное исследование с небольшой выборкой (и возможно, без рандомизации), в результате которого была получена некоторая разность эффектов. Ее большое значение впечатлило исследователей и заставило предпочесть один метод другому, но изъяны обсервационного дизайна и малый объем выборки не позволяют им в полной мере аргументировать свое убеждение перед оппонентами. С целью доказать, что найденная разница в эффектах не случайна и не ошибочна, проводится повторное, более строго спланированное клиническое испытание, опирающееся на результаты предыдущего исследования. В этом случае мы знаем (или предполагаем), какую дельту эффекта мы ожидаем получить. Различие предыстории этих двух вариантов никак не влияет на формулу расчета объема выборки, она одна и та же в обоих случаях. Из всех величин, определяющих объем выборки, необходимый для исследования, разность эффекта является наиболее критичной. Как мы увидим далее из формулы, снижение дельты эффекта, например, вдвое заставляет увеличить объем выборки в 4 раза, а снижение втрое приводит к требованию набрать в 9 раз больше больных в каждую группу. Иными словами, необходимый объем выборки возрастает вовсе не пропорционально уменьшению дельты эффекта, он растет гораздо стремительнее.

Однако не только этот параметр определяет количество больных в группе исследования и сравнения. Двумя другими условиями являются принятые нами (также заранее, до начала первого измерения) ошибки I и II рода, с которыми мы готовы смириться в нашем исследовании. Их природу и смысл мы обсуждали в предыдущей публикации

нашего раздела [4]. Эти статистические ошибки сопровождают проверку любой статистической гипотезы, в том числе и проверку статистической значимости различия эффектов двух методов лечения. Их предельные уровни всегда должны быть заданы на этапе планирования исследования, до его начала. Ошибка I рода обозначается греческой буквой α и называется уровнем значимости, а ошибка II рода – буквой β и разность $1-\beta$ называется мощностью критерия. Уровень значимости обычно устанавливается на 5% (0,05 в терминах теории вероятности). Как правило, редакции зарубежных научных журналов не принимают статьи, где заключения сделаны при более слабом уровне значимости (выше 5%), поэтому лучше остановиться на этом числе. Мощность критерия, равная 80% (0,80), является вполне приемлемой и в большинстве исследований имеет место именно это значение, но если расчет объема выборки дает результаты, которые очень трудно или невозможно реализовать на практике, то мощность можно опустить вплоть до 70%, чтобы иметь возможность все-таки провести испытание. Чем выше мощность критерия, тем больше пациентов необходимо набрать в каждую группу для исследования, причем увеличение это не пропорционально, а экспоненциально, т.е. каждая следующая прибавка 1% мощности дается все большим количеством добавляемых больных. Следует очень внимательно оценить баланс между этими двумя факторами, чтобы с одной стороны иметь объективную возможность опираться на результаты исследования, а с другой – не утонуть в огромном потоке данных и их обработке. Обычно уровень значимости, равный 5%, и мощность, равная 80%, являются таким сбалансированным сочетанием, которое присутствует в большинстве исследований.

И наконец, еще одна, четвертая, характеристика наших данных, от которой напрямую зависит необходимый объем выборки, это вариабельность. Интуитивно должно быть понятно, что чем компактнее располагаются данные вокруг своего среднего (или медианы), тем легче обнаружить даже небольшое различие между группами. Если измерения в двух группах имеют большое рассеяние, то для доказательства даже достаточного выраженного различия необходимы большие объемы выборок в обеих группах. Параметром, отражающим вариабельность выборки, является ее среднеквадратическое отклонение (SD). О нем мы подробно говорили в предыдущих публикациях [3, 4]. Таким образом, если перевести все в термины статистики, то чем больше значение среднеквадратического отклонения изучаемого показателя, тем больший

объем выборки потребуется для достижения статистической значимости различия.

Мы рассмотрели все четыре фактора, влияющие на величину объема выборки и однозначно его определяющие. Эти четыре параметра комбинируются в формулу, которую мы не будем здесь выводить и доказывать (это задача для математиков), а просто примем к использованию в своих исследовательских целях. Если мы имеем дело с числовыми данными, то расчет объема выборки производится по следующей формуле:

$$N = 2 * (Z_{\alpha/2} + Z_{\beta})^2 / (d/SD)^2, \quad (1)$$

где

- N – рассчитываемый объем выборки,
- $Z_{\alpha/2}$ и Z_{β} – значения нормального распределения при вероятности $\alpha/2$ и β соответственно,
- d – клинически значимая разность групповых средних значений,
- SD – среднеквадратическое отклонение.

Если мы зададим стандартные условия для уровня значимости (0,05) и мощности критерия (0,80), то $Z_{\alpha/2} + Z_{\beta} = 1,96 + 0,84 = 2,8$. Это значение необходимо возвести во вторую степень и умножить на 2, в результате чего получим 15,68. Если мы не будем менять наши условия относительно ошибок первого и второго рода, то это значение можно оставить в качестве постоянной, которую будем делить на различные значения выражения $(d/SD)^2$. Эти два параметра можно задавать по отдельности, но можно поступать проще, задавая сразу значения их частного. В этом случае мы находим объем выборки, который необходим для подтверждения достоверности различия двух средних при условии, что их разность составляет некоторую часть от их вариабельности. Можно даже составить таблицу для некоторых значений этого частного, тогда даже неважно, какие показатели и в каких единицах мы собираемся исследовать. Такой таблицей можно будет пользоваться при расчете объема выборки, необходимого для проверки значимости различия двух средних, составляющего заданную часть от их общего среднеквадратического отклонения. (табл.1). При этом, правда, нужно помнить, что эта таблица верна только для сочетания уровня значимости, равного 0,05, и мощности критерия 0,80. Для других значений двух указанных порогов надо будет рассчитать другую таблицу аналогичным образом.

Обратите внимание, как стремительно растет необходимый объем выборки, когда дельта эффекта, которую мы собираемся статистически доказывать, становится меньше половины стандартного отклонения (величины вариабельности)

Таблица 1. Пример расчетной таблицы объемов выборки, необходимых для проверки статистической достоверности различия двух средних на уровне значимости 0,05 и при мощности критерия 0,80 для различных значений d/SD

d/SD	Расчетный объем выборки для одной группы	Реальный объем выборки для одной группы (N)	Суммарный объем выборки для 2 групп исследования (без учета выбывания)	Суммарный расчетный объем выборки для 2 групп исследования (с учетом 10% выбывания)	Реальный суммарный объем выборки для 2 групп исследования (с учетом 10% выбывания)
0,1	1568	1568	3136	3449,6	3450
0,2	392	392	784	862,4	864
0,3	174,2	174	348	382,8	384
0,4	98	98	196	215,6	216
0,5	62,72	63	126	138,6	140
0,6	43,6	44	88	96,8	98
0,7	32	32	64	70,4	72
0,8	24,5	25	50	55	56
0,9	19,4	20	40	44	44
1	15,68	16	32	35,2	36
1,1	13,0	13	26	28,6	30
1,2	10,9	11	22	24,2	26
1,3	9,3	10	20	22	22
1,4	8	8	16	17,6	18
1,5	7,0	7	14	15,4	16

показателя, принятого за индикатор эффективности в нашем исследовании. Если для доказательства достоверности различия эффектов, равного 1 среднеквадратическому отклонению нам достаточно по 16 больных на группу (без учета возможного выбывания их в ходе исследования), то для дельты эффекта, равной 0,5, т. е. половине среднеквадратического отклонения, необходимо уже по 63 пациента в каждой группе. А если важно доказать даже небольшое различие, размером всего 0,3 SD, то объем выборки в каждой группе должен быть не менее 174 пациентов, т. е. 348 больных в целом для исследования. Необходимо также отметить, что в ходе больших или длительных исследований, например когортных или мультицентровых, часть пациентов выбывает из наблюдения по разным причинам. Процент таких потерь составляет по разным данным от 10 до 30, а для постмаркетинговых клинических исследований эффективности новых препаратов иногда даже более 50. Об этом следует помнить при планировании продолжительного или объемного исследования и по возможности закладывать 10–15% на случайную потерю данных, чтобы к концу исследования не оказаться у разбитого корыта, не имея возможности достичь достоверности различия заданного эффекта из-за недостаточного объема выборки.

Существует также формула для приблизительной оценки необходимого объема выборки при стандартно заданных условиях проверки

достоверности различия (уровень значимости равен 0,05 и мощность критерия 0,80):

$$N = 16/(d/SD)^2 \quad (2)$$

Она дает вполне приемлемые результаты, достаточно близкие к точным расчетам, особенно при больших значениях разности эффекта. При маленьких значениях дельты, она немного завышает необходимый объем выборки, что не сказывается на достоверности выводов, но может лишь несколько усложнить формирование групп.

Пример 1.

На примере данных одной из статей, опубликованных в нашем журнале, покажем, как рассчитать необходимый объем выборки для получения статистически значимых результатов сравнения двух групп.

В исследовании, посвященное влиянию физической реабилитации на характер послеоперационной боли у родильниц, перенесших кесарево сечение [5], было включено 30 пациенток, разбитых на две группы по 15 пациенток в каждой. В 1-й группе применялись средства физической реабилитации, во 2-й – реабилитация пациенток не проводилась. Через 12 ч у пациенток 1-й группы интенсивность боли по ВАШ в покое составила в среднем 10,8 балла, во 2-й группе этот же параметр был равен 12,5. Стандартное отклонение составило 1,5 в 1-й группе и 3,1 – во 2-й. Статистическая достоверность различия средних двух групп, полученных

через 12 ч после операции, не была достигнута, хотя дельта, равная примерно 2 баллам, имела тот же знак, что и последующие разности двух средних, полученных через 24 и 48 ч, достоверность которых была подтверждена статистически на уровне значимости 0,05. Возможно, 15 пациенток в группе просто недостаточно, чтобы подтвердить влияние физической реабилитации на характер послеоперационной боли уже через 12 часов после кесарева сечения. Рассчитаем минимальный объем выборки, который будет достаточным, чтобы разница средней интенсивности боли по ВАШ в 2 балла была статистически достоверной на уровне значимости 0,05 при мощности критерия 0,80.

Во-первых, вычислим общее среднеквадратическое отклонение для двух групп:

$$SD_0 = ((SD_1^2 * N_1 + SD_2^2 * N_2) / (N_1 + N_2))^{1/2}, \text{ где}$$

SD_0 – среднеквадратическое отклонение для двух групп вместе,

SD_1 – среднеквадратическое отклонение, полученное в 1-й группе,

SD_2 – среднеквадратическое отклонение, полученное во 2-й группе,

N_1 – количество пациенток в 1-й группе,

N_2 – количество пациенток во 2-й группе.

$$SD_0 = ((1,5^2 * 15 + 3,1^2 * 15) / (15 + 15))^{1/2} = 2,44$$

Получаем $SD_0 = 2,44$. Дельта эффекта (d), как мы уже выяснили, составляет 2 балла, значит отношение $d/SD_0 = 0,82$ или приблизительно 0,8. Находим в таблице 1 строку с соответствующим значением отношения d/SD и получаем значения объема выборки, равное 25 пациенткам в каждой из двух групп. Итак, для подтверждения статистической значимости различия средних оценок интенсивности боли по ВАШ в каждую группу необходимо набрать по 25 рожениц.

Как было отмечено ранее, второй тип данных, достаточно широко представленных в клинических исследованиях, – это бинарные отклики. Как правило, это частота какого-либо осложнения, распространенность заболевания, частота симптома при данной патологии и т. п. В этом случае все четыре статистических параметра, которые мы уже подробно рассмотрели, также определяют количество пациентов в каждой группе, и формула по сути остается такой же, но поскольку мы имеем дело в данном случае с частотой, то появляется возможность упростить процедуру расчета, опираясь на некоторые особенности частот и бинарных показателей. Дело в том, что дисперсия (квадрат среднеквадратического отклонения) частоты, т.е. ее вариабельность, вычисляется по формуле:

$p^*(1-p)$ и таким образом формулу расчета требуемого объема выборки для сравнения двух частот можно записать следующим образом:

$$N = (2 * p^*(1 - p)^* (Z_{\alpha/2} + Z_{\beta})^2) / d^2, \quad (3)$$

где все обозначения имеют тот же смысл, что и в формуле 1, а p' рассчитывается как среднее значение двух сравниваемых частот: $p' = (p_1 + p_2) / 2$. Если клинически значимую разность частот, статистическую достоверность которой мы собираемся доказать в ходе нашего исследования, переписать в виде разности этих частот, то формула примет вид:

$$N = (2 * p^*(1 - p)^* (Z_{\alpha/2} + Z_{\beta})^2) / (p_1 - p_2)^2$$

и станет понятно, что для определения требуемого объема выборки нам необходимо предположить или получить в пилотном наблюдении значения двух сравниваемых частот. В этом случае также можно составить таблицу для разных комбинаций сравниваемых частот при фиксированных условиях критерия сравнения. Мы приводим пример такой таблицы с шагом частот в 10% при стандартных значениях уровня значимости и мощности критерия (табл. 2). Эту таблицу можно использовать для предварительной оценки объема выборки, который будет необходим для доказательства значимости различия частот какого-либо события в двух группах. Понятно, что частоты скорее всего не будут точно различаться на 10%, различие может составлять и 3%, и 14%, но, прежде чем начинать точный расчет, можно просто прикинуть по таблице 2, как много пациентов потребуется набрать в исследование: пару десятков, несколько десятков, больше сотни или близко к тысяче. Такая приближительная оценка позволит оперативно принять решение об изменении дизайна исследования или конечных точек, чтобы было реально его провести в условиях конкретной клиники.

Пример 2.

Для демонстрации вычисления требуемого объема выборки в случае сравнения частот обратимся к одной из статей, опубликованных в нашем журнале [6]. В этом исследовании сравнивались различные методы профилактики интраоперационной тошноты и рвоты (ИОТР) при спинномозговой анестезии (СМА) во время операции кесарева сечения. Хотя авторы сравнивали более чем два метода, для нашего примера мы рассмотрим сравнение только двух групп. В 1-й премедикация проводилась путем внутривенного введения 10 мг метоклопрамида за 10 мин до начала СМА и частота ИОТР в этой группе составила 18,2% + 2,6%. Во второй группе всем беременным за 10 мин до СМА было введено внутривенно 8 мг дексаметазона, и частота ИОТР в этой группе составила 11,1% + 1,6%.

Таблица 2. Объемы выборки, необходимые для успешной проверки статистической достоверности различия двух частот на уровне значимости 0,05 и с мощностью критерия 0,80

Частоты	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	200	63	33	21	15	11	8	7
0,2		251	69	33	20	14	10	8
0,3			304	79	36	21	14	10
0,4				351	89	40	23	15
0,5					383	96	43	24
0,6						392	98	44
0,7							374	94
0,8								327

Достоверность различия частот не была статистически подтверждена в данном исследовании, так как совокупный объем выборки в двух группах составил 71 пациентку. Какой объем выборки необходим для получения статистической значимости межгруппового различия указанных частот ИОТР?

Рассчитаем вначале среднюю частоту p' (будем писать частоты в терминах теории вероятностей, а не в процентах):

$$p' = (p_1 + p_2) / 2 = (0,182 + 0,111) = 0,147$$

Далее по формуле (2) найдем значения N , объем выборки, необходимый для статистического подтверждения гипотезы о различии частот в указанных двух группах на уровне значимости, равном 0,05 с мощностью критерия 0,80.

$$N = 15,68 * 0,147 * (1 - 0,147) / (0,182 - 0,111)^2 = 390,03$$

Таким образом, для статистического подтверждения значимости на заданном уровне в каждую из двух групп необходимо набрать по 390 пациенток, всего для исследования 780. Как видим, в проведенном пилотном исследовании в каждой группе было почти в 10 раз меньше пациенток, чем требуется для достижения статистической значимости. Если мы снизим мощность критерия, например до 0,70, мы, конечно, получим меньшее число, но тем самым мы увеличиваем вероятность пропустить различие этих двух частот (т.е. ошибочно отвергнуть различие двух методов профилактики ИОТР), тогда как оно в действительности существует.

Как уже отмечалось выше, баланс между риском принять неправильное решение и реальностью проведения исследования определяется специалистом. Только ему известно, насколько важны те или иные выводы, и каковы возможности сформировать выборку необходимого объема в условиях той клиники, где предполагается проводить исследование.

Литература

1. Тихова Г. П. Четырехпольная таблица частот – Бритва Оккама в мире статистики. Часть 1. Как рассчитывать относительный риск и другие параметры из четырехпольной частотной таблицы. Регионарная анестезия и лечение острой боли. 2012; 3: 69–75
2. Тихова Г. П. Корреляционный анализ данных: зонд в глубины скрытых механизмов взаимодействий. Регионарная анестезия и лечение острой боли. 2013; 1: 48–53
3. Тихова Г. П. Значение и интерпретация ошибки среднего в клиническом исследовании и эксперименте. Регионарная анестезия и лечение острой боли. 2013; 3: 50–53
4. Тихова Г. П. Проверка гипотезы: что необходимо знать о достоверности различия. Регионарная анестезия и лечение острой боли. 2014; 1: 61–66
5. Антипина Н. П., Антипин Э. Э., Совершаева С. Л., Нестеренко С. Е. Влияние физической реабилитации на характер болевого синдрома у родильниц, перенесших операцию кесарева сечения. Регионарная анестезия и лечение острой боли. 2009; 3: 32–35
6. Погодин А. М., Шифман Е. М. Профилактика тошноты и рвоты при спинномозговой анестезии во время операции кесарева сечения. Регионарная анестезия и лечение острой боли. 2009; 1: 11–14

References

1. Tikhova G. P. Fourfold frequency table – Occam's Razor in the world of statistics. Part 1. Calculating relative risk and other parameters from fourfold frequency table. Regionarnaya anesteziya i lechenie ostroy boli. 2012; 3: 69–75
2. Tikhova G. P. Correlation analysis: exploring hidden mechanics of relationships. Regionarnaya anesteziya i lechenie ostroy boli. 2013; 1: 48–53
3. Tikhova G. P. Importance and interpretation of standard error of mean in clinical study and trial. Regionarnaya anesteziya i lechenie ostroy boli. 2013; 3: 50–53
4. Tikhova G. P. Testing of hypothesis: what is necessary to know about statistically significant difference. Regionarnaya anesteziya i lechenie ostroy boli. 2014; 1: 61–66
5. Antipina N. P., Antipin E. E., Sovershaeva S. L., Nesterenko S. E. Effect of physical rehabilitation on course of pain syndrome in women underwent cesarean section. Regionarnaya anesteziya i lechenie ostroy boli. 2009; 3: 32–35
6. Pogodin A. M., Shifman E. M. Prevention of nausea and vomiting during cesarean section under spinal anesthesia. Regionarnaya anesteziya i lechenie ostroy boli. 2009; 1: 11–14