

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
«Центр экспертизы и контроля качества медицинской помощи»
Министерства здравоохранения Российской Федерации
(ФГБУ «ЦЭКМП» Минздрава России)»**

Утверждено приказом
ФГБУ «ЦЭКМП» Минздрава России

от «29» декабря 2017 г. № 181-од

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОЦЕНКЕ КАЧЕСТВА
СТАТИСТИЧЕСКОГО АНАЛИЗА В КЛИНИЧЕСКИХ
ИССЛЕДОВАНИЯХ**

Москва 2017 г.

Авторский коллектив:

Генеральный директор ФГБУ «ЦЭККМП» Омеляновский В.В.
Минздрава России

Советник руководителя ФГБУ «ЦЭККМП» Авксентьева М.В.
Минздрава России

Начальник отдела методологического Сура М.В.
обеспечения проведения комплексной
оценки технологий в здравоохранении
ФГБУ «ЦЭККМП» Минздрава России

Заместитель начальника отдела Хачатрян Г.Р.
методологического обеспечения
проведения комплексной оценки
технологий в здравоохранении ФГБУ
«ЦЭККМП» Минздрава России

Главный специалист отдела Федяева В.К.
методологического обеспечения
проведения комплексной оценки
технологий в здравоохранении ФГБУ
«ЦЭККМП» Минздрава России

Содержание

1. Введение.....	4
2. Основные термины и определения	5
3. Рекомендации по представлению информации о статистическом анализе в публикациях о клинических исследованиях	9
3.1. Рекомендации по представлению информации о плане статистического анализа в разделе «Материалы и методы».....	9
3.2 Рекомендации по представлению информации о результатах статистического анализа в разделе «Результаты»	14
3.3 Рекомендации по представлению информации в разделе «Выводы»	19
4. Рекомендации по оценке риска некорректности статистического анализа (риска возникновения случайных ошибок) в клинических исследованиях.....	21
Литература	31
Приложение 1. Перечень терминов на английском языке.....	32
Приложение 2. Методика дополнительных расчетов для проверки заключений/выводов в клинических исследованиях.....	33

1. Введение

В качестве правовой нормы в Российской Федерации введено определение «комплексной оценки лекарственного препарата», в рамках которой проводится оценка зарегистрированного лекарственного препарата, включающая в себя анализ информации о сравнительной клинической эффективности и безопасности лекарственного препарата, оценку экономических последствий его применения, изучение дополнительных последствий применения лекарственного препарата в целях принятия решений о возможности включения лекарственного препарата в перечень жизненно необходимых и важнейших лекарственных препаратов, нормативные правовые акты и иные документы, определяющие порядок оказания медицинской помощи, или исключения его из указанных перечня, актов и документов [1].

Одним из компонентов оценки клинической эффективности и безопасности лекарственного препарата является оценка методологического качества клинических исследований, изучающих клиническую эффективность и безопасность лекарственного препарата. В ходе оценки методологического качества клинических исследований необходима также оценка качества статистического анализа (СА), проведенного в клинических исследованиях.

Дизайном клинических исследований, являющимся «золотым стандартом» клинических исследований, являются рандомизированные контролируемые испытания (РКИ). РКИ предусматривают наиболее сложно организованный дизайн, в том числе для СА, при этом данные, полученные в РКИ, обладают самым высоким уровнем достоверности по результатам комплексной оценки лекарственных препаратов [2]. Фокусом настоящих методических рекомендаций является оценка качества статистического анализа в РКИ.

Целью настоящих методических рекомендаций является формирование единого подхода к оценке качества СА в клинических исследованиях лекарственных препаратов при проведении их комплексной оценки.

Настоящие методические рекомендации могут рассматриваться также как рекомендации по планированию, выполнению и представлению результатов СА в клинических исследованиях лекарственных препаратов.

Настоящие методические рекомендации предназначены для представителей экспертных организаций и лиц, осуществляющих проведение комплексной оценки лекарственных препаратов, а также всех заинтересованных лиц, участвующих в комплексной оценке лекарственных препаратов, а также для лиц, занимающихся подготовкой публикаций о клинических исследованиях лекарственных препаратов.

2. Основные термины и определения

Абсолютная частота признака — количество объектов исследования в группе, имеющих данную характеристику (значение признака) [3].

Альтернативная гипотеза – статистическая гипотеза о существовании различий между группами. Является противоположностью нулевой гипотезе [3].

Альфа-ошибка (альфа-уровень, ошибка первого рода, вероятность гипердиагностики) – пороговый уровень статистической значимости; вероятность ошибочно отклонить нулевую гипотезу. Чем меньше альфа-уровень, тем меньше риск совершения этой ошибки. Устанавливается исследователем произвольно (обычно принимается равным 0,05, 0,01, или 0,001) [3].

Бинарные признаки – данные, выраженные только двумя допустимыми альтернативными значениями (например, «есть» - «нет», «здоров» - «болен») [3].

Безопасность лекарственного средства – характеристика лекарственного средства, основанная на сравнительном анализе его эффективности и риска причинения вреда здоровью [1].

Гипотеза научная – утверждение, которое можно подтвердить или опровергнуть на основании результатов исследования [3].

Доверительный интервал (ДИ) – интервал значений признака, рассчитанный для какого-либо параметра распределения (например, среднего значения признака) по выборке и с определенной вероятностью (например, 95% для 95% ДИ) включающий истинное значение этого параметра во всей популяции [3].

Интерквартильный размах – интервал значений признака, содержащий центральные 50% наблюдений выборки, т.е. интервал между 25-м и 75-м процентилями. Используется вместе с медианой для описания распределения данных как параметр вариабельности [3].

Исход – любой возможный результат, возникающий от воздействия причинного фактора, профилактического или терапевтического вмешательства, все установленные изменения состояния здоровья, возникающие как следствие вмешательства [4].

Качественные данные – номинальные и порядковые данные, которые отражают условные коды неизмеряемых категорий или условную степень выраженности признака [3].

Ковариата – переменная, которая, возможно, предсказывает изучаемый исход. Ковариата может быть изучаемой, вмешивающейся переменной или модификатором эффекта [4].

Количественные данные – данные, измеряемые в шкале, поделенной на равные интервалы. Количественные данные могут быть непрерывными и дискретными [3].

Комплексная оценка лекарственного препарата – оценка зарегистрированного лекарственного препарата, включающая в себя анализ информации о сравнительной клинической эффективности и безопасности лекарственного препарата, оценку

экономических последствий его применения, изучение дополнительных последствий применения лекарственного препарата в целях принятия решений о возможности включения лекарственного препарата в перечень жизненно необходимых и важнейших лекарственных препаратов, нормативные правовые акты и иные документы, определяющие порядок оказания медицинской помощи, или исключения его из указанных перечня, актов и документов [1].

Лекарственные препараты (ЛП) – лекарственные средства в виде лекарственных форм, применяемые для профилактики, диагностики, лечения заболевания, реабилитации, для сохранения, предотвращения или прерывания беременности [1].

Медиана – значение признака, распределяющее пополам распределение наблюдений на интервале значений признака, т.е. половина значений в выборке меньше медианы, а вторая половина – больше ее. Используется для описания центральной тенденции распределений количественных и порядковых признаков независимо от закона распределения [3].

Непараметрические методы статистики – класс статистических методов, которые используются главным образом для анализа количественных данных, не образующих нормальное распределение, а также для анализа качественных данных [3].

Непрерывные данные – количественные данные, принимающие значения на непрерывной шкале значений [3].

Номинальные данные – качественные данные, которые отражают условные коды неизмеряемых категорий, не подлежащих упорядочиванию. Альтернативой служат качественные порядковые данные, которые могут быть ранжированы в убывающем или возрастающем порядке [3].

Нормальное распределение (гауссово распределение) – симметричное, колоколообразное распределение, у которого среднее, мода и медиана совпадают. Пологость кривой (т.е. вариация данных) может быть различной [3].

Нулевая гипотеза – статистическая гипотеза об отсутствии различий между признаками или отсутствии взаимосвязи признаков. Проверяется с помощью статистических критериев [3].

Относительная частота признака — отношение числа объектов с каким-либо значением признака к общему числу объектов [3].

Параметрические методы статистики – класс статистических методов, используемых для анализа данных, которые образуют известное распределение (обычно нормальное). Названы так потому, что опираются на анализ параметров нормального распределения) [3].

Пороговый уровень статистической значимости – допускаемая исследователем величина альфа-ошибки, т.е. максимально допускаемая исследователем вероятность ошибочного отклонения нулевой гипотезы (об отсутствии различия между группами, об отсутствии

взаимосвязи признаков и т.д.) при применении статистического теста. Как правило за величину уровня статистической значимости принимаются значения 0,05, 0,01 или 0,001 [3].

Порядковые (ранговые) данные – качественные данные, значения которых могут быть ранжированы по какому-либо принципу, но интервал между значениями таких данных не может быть измерен количественно. Обычно качественно отражают условную степень выраженности какого-либо признака. Порядковые признаки являются качественными (иногда их называют “полуколичественными”) оценками какой-либо характеристики [3].

Проблема множественных сравнений – проблема завышения статистической значимости результатов статистических тестов, возникающая при проверке большого числа статистических гипотез на одной выборке [3].

Протокол клинического исследования – план или комплекс действий, которому необходимо следовать при проведении исследования или программы вмешательства [4].

Рандомизированное контролируемое испытание (РКИ) – эксперимент, в котором субъекты рандомизируются в группы, обычно называемыми группами (исследуемого) вмешательства и контроля, для того чтобы им проводились либо не проводились экспериментальные профилактические или терапевтические вмешательства [4].

Распределение признака в выборке – набор частот упорядоченных значений признака в выборке. Обычно представляется в виде графика (по оси абсцисс – интервалы значений признака, по оси ординат – частоты упорядоченных значений признака) [3].

Случайная ошибка – часть изменчивости результатов измерения, не связанная заметно с другим измерением или переменной. Обычно рассматривается как возникающая случайно [4].

Среднее значение – описательная статистика (параметр), являющийся мерой центральной тенденции для приближенно нормального распределения. Если распределение не соответствует закону нормального распределения, то для характеристики центральной тенденции следует использовать медиану [3].

Среднеквадратическое отклонение (СКО) – описательная статистика (параметр), являющаяся мерой рассеяния (мерой вариабельности) для приближенно нормально распределенных данных. Если распределение не соответствует закону нормального распределения, то для характеристики рассеяния следует использовать какой-либо интерпроцентильный интервал, например, интерквартильный размах [3].

Стандартная ошибка среднего – мера точности оценки среднего значения в генеральной совокупности на основании данных исследуемой выборки. В отличие от СКО не является характеристикой, описывающей рассеяние наблюдений (мерой вариабельности) [3].

Статистическая гипотеза – представление научной гипотезы в форме, приемлемой для проверки методами статистического анализа данных [3].

Статистическая значимость – максимально допускаемая исследователем вероятность ошибочного отклонения нулевой гипотезы (об отсутствии различия между группами, об отсутствии взаимосвязи признаков и т.п.) статистического теста [3].

Статистическая мощность – вероятность обнаружить определенное различие, когда оно на самом деле существует, вероятность правильно отклонить нулевую гипотезу. Как правило принимается равной 0,8 или 0,9 [3].

3. Рекомендации по представлению информации о статистическом анализе в публикациях о клинических исследованиях

Основная информация, позволяющая судить о качестве проведенного СА (риска возникновения случайных ошибок), должна быть представлена в разделах «Материалы и методы» и «Результаты» публикаций о клинических исследованиях.

3.1. Рекомендации по представлению информации о плане статистического анализа в разделе «Материалы и методы».

Раздел «Материалы и методы» является основным разделом для приведения информации о СА и отражает корректность планирования СА в исследовании. Корректное планирование СА снижает риск получения недостоверных результатов исследования, хотя и не гарантирует их надежности.

Основным критерием качественного представления информации о СА является возможность его повторения любым исследователем, имеющим соответствующие навыки в выполнении СА (при наличии доступа к первичным данным клинического исследования).

Ниже приведены основные необходимые компоненты планирования СА, информацию о которых рекомендуется отразить в протоколе клинического исследования, а также при подготовке текста публикации о проведенном клиническом исследовании в разделе «Материалы и методы».

3.1.1 Основная проверяемая научная гипотеза.

В публикации о клиническом исследовании рекомендуется указывать одну из трех научных гипотез при исследовании вмешательств:

1) **Гипотеза превосходства**¹ предполагает, что эффективность и/или безопасность исследуемой технологии выше, чем технологии сравнения, по выбранному критерию (критериям). Как правило, данная гипотеза используется в клинических исследованиях при сравнении с плацебо или стандартной терапией (чаще) или при прямых сравнительных исследованиях конкурентных препаратов (реже);

2) **Гипотеза неуступающей эффективности** предполагает, что исследуемая технология не хуже технологии сравнения по выбранному критерию (критериям) эффективности и/или безопасности. Данная гипотеза наиболее часто используется в прямых сравнительных исследованиях конкурирующих лекарственных препаратов или в исследованиях генерических лекарственных препаратов;

¹ Поскольку большое количество публикаций о клинических исследованиях написаны на английском языке, перевод некоторых наиболее употребляемых терминов на английский язык представлен в Приложении 1.

3) **Гипотеза эквивалентности** предполагает, что эффективность и/или безопасность сравниваемых технологий является эквивалентной по выбранному критерию (критериям) эффективности и/или безопасности.

3.1.2 Расчет объема выборки согласно проверяемой гипотезе исследования.

В публикации о клиническом исследовании рекомендуется приводить информацию о расчете размера выборки.

Расчет размера выборки рекомендуется осуществлять по первичному исходу, т.е. исходу, по которому делается основной вывод исследования. Данный исход, как правило, является клиническим (т.н. «конечным»), подробнее см. [2]) и чаще всего бинарным (наличие или отсутствие ремиссии, выздоровления, частота объективных событий, таких как смерть, прерывание беременности или др.) или количественным (например, общая выживаемость, продолжительность госпитализации или др.).

Расчет объема выборки рекомендуется осуществлять с использованием сертифицированного программного обеспечения – статистического пакета (предпочтительно) или онлайн-калькулятора с использованием следующей информации: статистическая мощность исследования (чаще всего используется статистическая мощность 80% или 90%, или 0,8 или 0,9), ожидаемый размер эффекта в каждой из групп сравнения (в процентах, в абсолютных единицах), уровень альфа-ошибки (ошибки первого рода). Рекомендуется, чтобы информация о данных показателях была приведена в тексте публикации о клиническом исследовании. Кроме того, рекомендуется приведение информации о критерии клинической эффективности или безопасности, по которому производился расчет размера выборки, а также используемом программном обеспечении (в случае, если используемое для расчета размера выборки программное обеспечение отличалось от программного обеспечения, использованного для проведения непосредственно СА).

3.1.3 Информация об анализируемом наборе данных.

В публикации о клиническом исследовании рекомендуется приводить информацию об анализируемом наборе данных – наборе данных по намерению лечить или наборе данных по протоколу.

В ходе длительного наблюдения пациентов (в динамических исследованиях) неизбежно возникновение пропусков в данных из-за выбывания пациентов, технических ошибок и прочих причин. В таких ситуациях возможны два подхода к формированию набора данных в зависимости от вида проверяемой гипотезы:

1) для проверки гипотезы превосходства рекомендуется анализировать набор данных по намерению лечить, т.е. данные всех пациентов с назначенным лечением или получивших хотя бы одну дозу лекарственного препарата. Анализ по намерению лечить в

данном случае является предпочтительным, т.к. позволяет оценить эффект более консервативно.

2) при проверке гипотезы неуступающей эффективности или эквивалентности рекомендуется анализировать набор данных всех пациентов, закончивших наблюдение по протоколу.

3.1.4 Метод заполнения пропусков по выбывшим пациентам (в случае анализа по намерению лечить).

В публикации о клиническом исследовании рекомендуется описывать способ заполнения пропусков по выбывшим пациентам, если применен анализ по намерению лечить.

В случае анализа по протоколу заполнения пропусков по выбывшим пациентам не осуществляется.

Наиболее часто для заполнения пропусков по выбывшим пациентам используются данные последнего наблюдения, при этом заполнение пропусков может осуществляться по оптимистическому (успех) или пессимистическому (неуспех) сценарию.

Оптимистический сценарий подразумевает, что пропуски данных заполняются в соответствии с предположением о том, что у выбывших пациентов вмешательство продемонстрировало положительный эффект, пессимистический сценарий – в соответствии с предположением о том, что у выбывших пациентов не наблюдался эффект вмешательства. Предпочтительным вариантом для клинических исследований вмешательств является пессимистический сценарий, т.к. в данном случае оценка результатов лечения происходит более консервативно.

В случае, если количество выбывших из исследования пациентов составило более 20%, рекомендуется указать, не утрачена ли сопоставимость групп изучаемого вмешательства и контроля в результате выбывания пациентов (рекомендуется повторная проверка статистических гипотез о сопоставимости групп по всем или наиболее значимым исходным характеристикам) [3].

3.1.5 В случае анализа по протоколу: сопоставление частот и времени событий для выбывших, исключенных, некомплаентных пациентов, пропусков в данных и прочих.

В публикации о клиническом исследовании рекомендуется приводить информацию о частотах выбываний в каждой из групп сравнения, а также описать и привести результаты сопоставления причин выбываний и их временные промежутки.

Проведение сопоставления частот и причин выбывания рекомендуется с целью выявления равномерности выбывания и анализа причин выбывания. Также, если количество выбывших из исследования пациентов составило более 20%, рекомендуется указать, не

утрачена ли сопоставимость групп изучаемого вмешательства и контроля в результате выбывания пациентов (на основании повторной проверки статистических гипотез о сопоставимости групп по всем или наиболее значимым исходным характеристикам пациентов) [3].

3.1.6 Описание используемых статистических методов.

В публикации о клиническом исследовании рекомендуется перечислять все использующиеся методы СА (статистические критерии), а также типы данных (признаков), для которых были использованы выбранные методы.

Выбор статистического критерия является одним из ключевых моментов проведения СА, поскольку использование нерелевантных методов проверки гипотез приводит к искажению результатов и формулировке неверных выводов [5].

Выбор метода СА определяется типом анализируемых признаков, для количественных признаков при выборе метода СА учитывается также характер распределения признака.

В табл. 1 приведены основные типы признаков и наиболее распространенные соответствующие им статистические критерии.

Таблица 1. Рекомендуемые статистические критерии в зависимости от типа признака и вида его распределения [5]

Тип признака и распределения	Две независимые группы	2 зависимые группы	3 независимые группы и более
Количественный признак: нормальные распределения	Критерий Стьюдента для независимых групп (t-критерий для независимых групп)	Критерий Стьюдента для зависимых групп (t-критерий для зависимых групп)	ANOVA по Пирсону
Количественный признак: любые распределения	Критерий Манна-Уитни (U-критерий)	Критерий Вилкоксона	ANOVA по Крускалу-Уоллису
Качественный порядковый признак	Критерий Манна-Уитни (при числе значений признака более 5), хи-квадрат	Критерий Вилкоксона	ANOVA по Крускалу-Уоллису (при числе значений признака более 5), хи-квадрат
Качественный номинальный признак	Хи-квадрат	Критерий Кокрана	Хи-квадрат

Бинарный признак	Точный критерий Фишера (ТКФ)	Критерий МакНемара	Хи-квадрат
------------------	------------------------------	--------------------	------------

3.1.7 Указание использованных статистических пакетов.

В публикации о клиническом исследовании рекомендуется указывать название использованного при анализе данных пакета программ, его версии и фирмы-производителя.

Использование собственных разработок (аналогичных коммерческим статистическим пакетам) не рекомендуется. Это связано с тем, что использование разных алгоритмов при программировании одних и тех же статистических методов может приводить к несколько разным результатам, правильность алгоритмов же коммерческих пакетов программ проверена, и их высокое качество общеизвестно [3]. В настоящее время в России наиболее распространены следующие коммерческие пакеты:

- STATISTICA (StatSoft, USA);
- SPSS (SPSS Inc., USA);
- SAS (SAS Inst., USA).

3.1.8 Пороговый уровень статистической значимости (максимально допустимая вероятность альфа-ошибки – ошибки первого рода).

В публикации о клиническом исследовании рекомендуется указывать пороговый уровень статистической значимости (P), на основании рассчитанного значения которого происходит принятие или отвержение нулевой статистической гипотезы.

Общепринятый и наиболее часто используемый в публикациях о клинических исследованиях пороговый уровень статистической значимости составляет 0,05 (при $P < 0,05$ нулевая статистическая гипотеза об отсутствии статистически значимых различий отвергается и принимается альтернативная гипотеза о наличии статистически значимых различий между группами сравнения; при $P \geq 0,05$ принимается нулевая статистическая гипотеза об отсутствии статистически значимых различий между группами сравнения).

В случае проведения анализа подгрупп (проверки множества статистических гипотез в одной выборке) возникает проблема множественных сравнений, которая связана с повышением вероятности получить ошибочные результаты по причине выполнения множества действий. В связи с существованием проблемы множественных сравнений, при проведении анализа подгрупп рекомендуется устанавливать порог статистической значимости на уровне хотя бы 0,01 (чем меньше пороговый уровень статистической значимости, тем более достоверны выявленные различия). Для преодоления проблемы

множественных сравнений также применяются специальные методы, например, поправка Бонферрони (подробнее см. [3]).

3.2 Рекомендации по представлению информации о результатах статистического анализа в разделе «Результаты»

3.2.1 Описательная статистика и сопоставление исходных характеристик групп.

В публикации о клиническом исследовании рекомендуется приводить информацию об исходных характеристиках сравниваемых групп, а также результаты сопоставления этих характеристик между группами с использованием соответствующих статистических методов (с целью выявления дисбаланса групп).

Приведение описательной статистики характеристик групп исследования и их сопоставления рекомендуется даже при использовании рандомизации, поскольку рандомизация не гарантирует исходной сопоставимости групп. В случае, если выявлен исходный дисбаланс групп, рекомендуется проведение анализа соответствующих ковариат или анализ в подгруппах.

Следует отметить, что приведение в тексте исследования или под таблицей с исходными характеристиками групп информации о том, что группы были сопоставимы (не различались по исходным характеристикам, были сбалансированы, преимущественно сбалансированы или другие синонимы), не является обоснованием сопоставимости групп. Обоснованием сопоставимости групп являются только рассчитанные уровни статистической значимости для каждой из характеристик групп с указанием используемого статистического метода.

Пример корректного представления исходных характеристик групп приведен в табл. 2, некорректного – в табл. 3. Как видно в табл. 3, исходные характеристики групп не сопоставлены (не проверены статистические гипотезы об отсутствии различий между группами по выделенным характеристикам, не приведены рассчитанные уровни статистической значимости для каждой из характеристик пациентов), а также неясно, какие были использованы меры центральной тенденции и меры их вариабельности (см. п. 3.2.2 настоящего раздела).

В табл. 2 приведена корректная информация о сопоставлении групп по исходным характеристикам – правый столбец таблицы содержит информацию о рассчитанных уровнях статистической значимости и используемом статистическом критерии (U-тест).

Таблица 2. Корректное представление исходных характеристик групп

Клинические параметры	Новорожденные, получавшие	Новорожденные, получавшие	p, U-test
----------------------------------	--------------------------------------	--------------------------------------	------------------

	препарат X	препарат Y	
	Me [LQ; UQ] (min-max)	Me [LQ; UQ] (min-max)	
Гестационный возраст, нед.	30 [28; 36] (25-40)	29 [27; 31] (25; 40)	0,087
Масса тела, г	1382 [1120; 2080] (660-4300)	1347 [997; 1890] (450-3950)	0,210
Длина тела, см	41 [34; 46] (28-55)	38 [35; 41] (25-54)	0,123
Окружность головы, см	28 [27; 31] (23-27)	28 [27; 33] (21-37)	0,299

Таблица 3. Некорректное представление исходных характеристик групп (см. разъяснение в тексте)

Показатель	Новорожденные, получавшие препарат X	Новорожденные, получавшие препарат Y
Гестационный возраст, нед.	31± 0,7 (25-40)	30 ± 1,4 (25; 40)
Масса тела, г	1380 ± 186,1 (660-4300)	1345 ± 150,2 (450-3950)
Длина тела, см	40 ± 1,9 (28-55)	37 ± 1,4 (25-54)
Окружность головы, см	28 ± 0,4 (23-27)	28 ± 0,3 (21-37)

3.2.2 Представление описательной статистики количественных и порядковых признаков.

В публикации о клиническом исследовании в таблице с исходными характеристиками групп при описании распределений количественных признаков рекомендуется приводить четкое указание на то, какие были использованы меры центральной тенденции и вариабельности.

Только для нормально распределенных количественных признаков возможно использование среднего и СКО (условное обозначение $M \pm s$ или M, s). Для любых распределений количественных признаков и для порядковых признаков (условно количественных – с количеством возможных значений признака не менее 5) могут

использоваться только медиана и интерквартильный размах (условное обозначение $Me [Q1; Q3]$).

Представление описательной статистики количественных и условно количественных (порядковых) признаков должно отвечать следующим критериям:

- СКО должно быть не больше половины среднего – в случае, если признак имеет только положительные значения. В противном случае это свидетельствует о том, что распределение признака отличается от нормального, и среднее и СКО не могут быть использованы для описания распределения признака (см. п.2 настоящего раздела);
- Стандартная ошибка среднего (принятое условное обозначение m) не используется как мера вариабельности среднего (см. раздел «Основные термины»);
- Для порядковых признаков (условно количественных – с количеством возможных значений признаков не менее 5) не вычисляются средние и СКО, корректной описательной статистикой для таких признаков являются медиана и интерквартильный размах ($Me [Q1; Q3]$).

Примером корректного описания распределений количественных признаков также может служить табл. 2, некорректного – табл. 3. В табл. 2 содержится четкое указание на то, что для описания мер центральной тенденции количественных признаков и их вариабельности использованы медиана и интерквартильный размах. Из табл. 3 остается неясным, какие использованы меры центральной тенденции количественных признаков и меры вариабельности.

3.2.3 Описание качественных признаков.

В публикации о клиническом исследовании в таблице с исходными характеристиками групп при описании качественных признаков рекомендуется приводить абсолютные (n и N , где n – количество пациентов с наблюдаемым значением признака, N – общее количество пациентов в группе) и относительные (выраженные в процентах) частоты значений признаков.

Всегда должно быть ясно, что принято за 100%. Необходимо, чтобы при суммировании наблюдаемых частот по всей структуре явления итоговая сумма относительных частот равнялась 100%, а итоговая сумма абсолютных частот соответствовала общему количеству пациентов в группе.

Пример: описывается признак «Функциональный статус по шкале ECOG», при этом в критериях включения пациентов в исследование указано, что включались пациенты с ECOG 0-2. В таком случае, в таблице с исходными характеристиками групп должна быть представлена информация о количестве пациентов с каждым (0, 1, 2) функциональным

статусом по шкале ECOG, при этом при проверке (суммировании абсолютных и относительных частот) должно получаться общее количество пациентов в каждой из групп и 100% соответственно.

3.2.4 Выявление различий между группами.

В публикации о клиническом исследовании для обоснования выявленных статистически значимых различий (или отсутствия статистически значимых различий) между группами сравнения рекомендуется приводить все рассчитанные точные уровни статистической значимости, полученные в результате проверки статистических гипотез, а также ДИ для каждого результата, полученного в клиническом исследовании в каждой из групп сравнения.

Для каждого рассчитанного уровня статистической значимости рекомендуется указать использованный статистический критерий. В случае, если при каждом рассчитанном уровне статистической значимости не приведена информация о статистическом критерии, следует выяснить, указан ли в перечне статистических критериев метод, адекватный данному анализу.

Рассчитанный уровень статистической значимости рекомендуется приводить до третьего знака после запятой, допустимо – до второго знака после запятой, не рекомендуется приводить рассчитанный уровень статистической значимости в виде неравенства. Исключение составляют случаи, когда рассчитанный уровень статистической значимости $P < 0,001$ – в случае корректно приводить уровень статистической значимости в виде неравенства.

Для каждого ДИ рекомендуется указание вероятности, с которой истинное значение критерия во всей популяции включено в данный интервал (например, 95% для 95% ДИ).

3.2.5 Приведение в таблицах и на рисунках числа наблюдений для каждого признака в каждой точке наблюдения.

В публикации о клиническом исследовании на всех рисунках и во всех таблицах рекомендуется приводить общее количество пациентов в каждой из групп сравнения и точное количество пациентов (или долю от общего количества пациентов) с описываемым исходом или признаком в каждой временной точке наблюдения, которая запланирована в протоколе клинического исследования.

Пример корректного представления рисунков приведено на рис. 1, некорректного представления рисунков – на рис. 2. Как видно на рис. 2, ни на одном столбце гистограммы не приведено количество наблюдений (доли (%)) пациентов, у которых наблюдался изучаемый исход), а также не приведено общее количество пациентов в каждой из групп. Общее количество пациентов рекомендуется приводить на каждом рисунке, поскольку зачастую количество анализируемых пациентов может различаться для разных исходов в

клинических исследования: например, для основного исхода использован набор данных по намерению лечить, а для вторичного исхода – набор данных по протоколу.

3.2.6 Обоснование применимости параметрических методов СА (в случае их использования).

В случае использования параметрических методов СА в публикации о клиническом исследовании рекомендуется приводить обоснования их применимости.

Информация об обоснованности применения параметрических методов СА может быть приведена в разделе «Материалы и методы» или разделе «Результаты».

Обоснованием применимости параметрических методов СА является информация о том, что распределение признаков, анализируемых с помощью параметрических методов, является нормальным в каждой из групп сравнения. Рекомендуется приведение рассчитанных уровней статистической значимости и статистических критериев, на основании которых осуществлялась данная проверка.

К параметрическим методам СА относятся такие методы как t-критерий Стьюдента, корреляционный анализ по Пирсону, ANOVA по Пирсону и другие. Данная рекомендация касается как как сравнения групп по исходным характеристикам, так и основного СА, по результатам которого формулируются выводы исследования).

В случае применения непараметрических методов СА проверки вида распределения признаков не требуется.

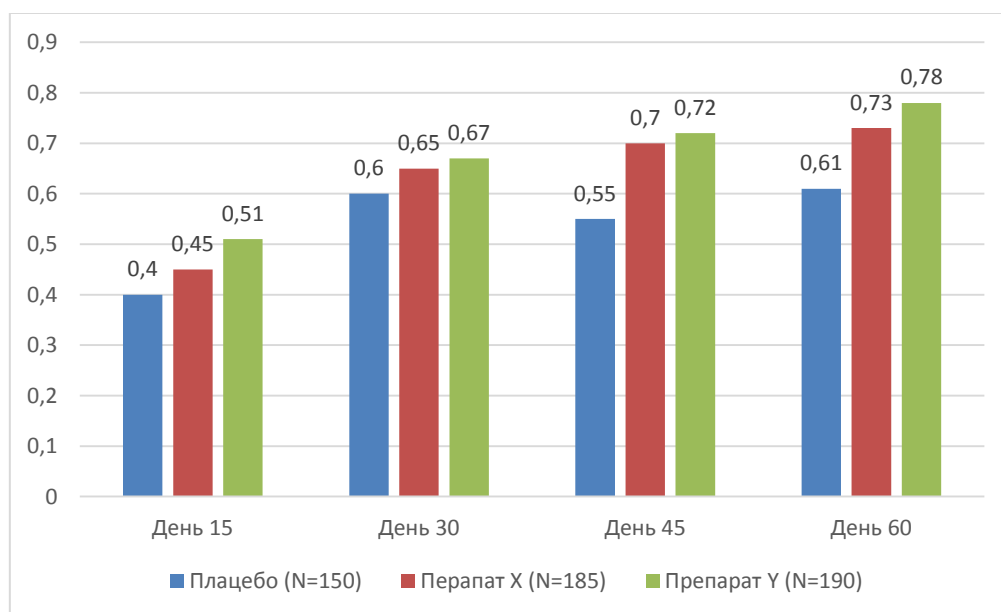


Рисунок 1. Доля пациентов с клиническим ответом: корректное представление данных на рисунке

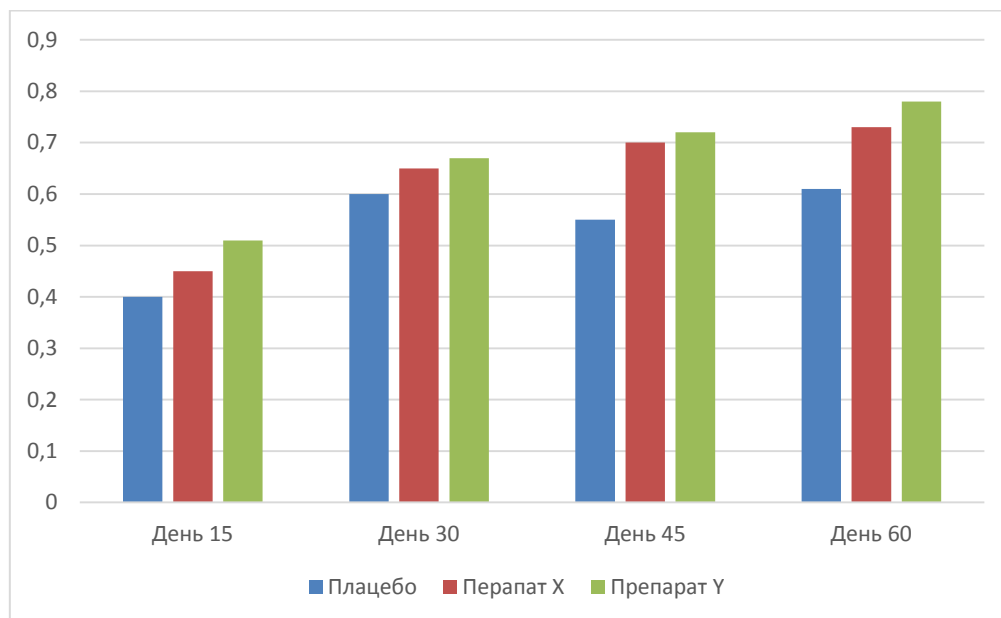


Рисунок 2. Доля пациентов с клиническим ответом: некорректное представление данных на рисунке (см. разъяснение в тексте)

3.3 Рекомендации по представлению информации в разделе «Выводы»

3.3.1 Интерпретация результатов.

При интерпретации результатов статистического анализа в публикации о клиническом исследовании рекомендуется учитывать следующие обстоятельства:

- 1) Невыявление эффекта в конкретном клиническом исследовании не означает его отсутствия в генеральной совокупности.
- 2) Методологические особенности клинического исследования (недостатки дизайна), приводящие к ограничению репрезентативности результатов.

3.3.2 Формулировка выводов.

При формулировке выводов клинического исследования рекомендуется учитывать следующие обстоятельства:

- 1) Основные выводы клинического исследования рекомендуется делать на основании результатов, полученных по первичному исходу, заявленному в разделе «Материалы и методы». Например, если в качестве первичного исхода исследования заявлена частота достижения клинического ответа, а среди вторичных исходов – изменение некоторого лабораторного показателя, то при отсутствии статистически значимых различий по первичному исходу и наличии статистически значимых различий по вторичному исходу не рекомендуется делать основной вывод исследования на основании результатов, полученных по вторичному исходу.
- 2) Выводы клинического исследования должны быть сформулированы однозначно согласно полученным в клиническом исследовании результатам: вывод не может противоречить информации, приведенной в разделе «Результаты».

Например, если частота достижения клинического ответа на терапию не различалась между группами А и В: ОР=1,9 95% ДИ [0,89; 2,03], Р=0,087 (ТКФ) – т.е. статистически значимые различия не выявлены, однако направление эффекта в пользу группы А, т.к. точечное значение ОР>1, то:

Пример корректной формулировки вывода: различия в частоте достижения клинического ответа на терапию между группами А и В не выявлены.

Пример некорректной формулировки вывода: клинический ответ на терапию наблюдался чаще в группе А, чем в группе В.

3) Выводы клинического исследования должны быть сформулированы осторожно: при формулировании выводов следует учитывать ограничения, связанные с характеристиками включенных в клиническое исследование пациентов и их потенциальные отличия от генеральной совокупности, методологические ограничения исследования и прочие ограничения. Кроме того, рекомендуется принимать во внимание при формулировании выводов, что невыявление эффекта в конкретном клиническом исследовании не означает его отсутствия в генеральной совокупности.

4) Выводы, основанные на анализе вторичных исходов и анализе подгрупп, не могут являться основными выводами клинического исследования.

4. Рекомендации по оценке риска некорректности статистического анализа (риска возникновения случайных ошибок) в клинических исследованиях

При оценке риска некорректности статистического анализа рекомендуется использовать опросник оценки риска некорректности статистического анализа.

В табл. 4 приведен вопросник оценки качества СА в клинических исследованиях, в табл. 5 – форма для заполнения вопросника и вычисления итогового риска возникновения случайных ошибок. Данный инструмент предназначен для оценки риска случайных ошибок преимущественно в РКИ (инструмент первоначально был разработан именно для оценки риска случайных ошибок в РКИ). Проспективные исследования нерандомизированного дизайна могут быть оценены на основании данного инструмента, однако с существенными ограничениями (применимо меньшее количество доменов). Поскольку результаты, полученные в исследованиях нерандомизированного дизайна, исходно имеют меньшую надежность, оценка качества проведенного в них СА не является важной задачей. Применение данного инструмента для исследований несравнительного и ретроспективного дизайна не предусмотрено.

Инструмент оценки риска случайных ошибок в клинических исследованиях включает в себя 22 домена. Домены условно распределены по разделам публикации, в которых рекомендуется представление информация, на основании которой оценивается соответствующий домен: «Материалы и методы», «Результаты», «Заключение»/«Выводы».

Каждый домен представлен в формате вопроса, который предусматривает три варианта ответа: «Да», «Нет» и «Неприменимо». В случае, если информация корректно представлена в публикации в явном виде, исследователь, оценивающий риск случайных ошибок, делает положительное заключение по домену (ответ «Да»). В случае, если информация не представлена в публикации в явном виде или представлена в явном виде, но представление не является корректным, исследователь, оценивающий риск случайных ошибок, делает отрицательное заключение по домену (ответ «Нет»). В случае, если домен не является применимым, исследователем выбирается ответ «Неприменимо».

Итоговый риск случайных ошибок оценивается следующим образом:

- 1) вычисляется доля ответов «да» среди всех ответов «да» и «нет»;
- 2) используется следующая стратификация риска случайных ошибок:
 - a. если доля ответов «Да» составляет 67-100%, делается заключение о том, что в публикации об исследовании выявлен низкий риск возникновения случайных ошибок;
 - b. если доля ответов «Да» составляет 34-66%, делается заключение о том, что в публикации об исследовании выявлен средний риск возникновения случайных ошибок;

- с. если доля ответов «Да» составляет 0-33%, делается заключение о том, что в публикации об исследовании выявлен высокий риск возникновения случайных ошибок.

Таблица 4. Вопросник для оценки качества СА (риска возникновения случайных ошибок)

№	Вопросы	Ответ
Раздел «Материалы и методы»		
1.	Указана ли основная проверяемая гипотеза и критерий ее принятия?	
	Выполнение хотя одного из следующих критериев: <ul style="list-style-type: none"> • Приведена информация в явном виде об используемой гипотезе (например, превосходства, неуступающей эффективности или эквивалентности). • Для гипотезы превосходства: в исследовании вмешательство сравнивалось с плацебо. • Для гипотезы превосходства: в исследовании был выполнен расчет объема выборки, при этом ожидаемый размер эффекта в группе изучаемого вмешательства превосходил ожидаемый размер эффекта в группе сравнения. 	Да
	Отсутствует информация об используемой гипотезе или информации, приведенной в публикации, недостаточно для заключения о виде гипотезы исследования	Нет
2.	Выполнялся ли расчет необходимого объема выборки по основному критерию оценки эффективности или безопасности с учетом вида проверяемой гипотезы?	
	Приведена информация о расчете объема выборки: ожидаемый размер эффекта в каждой из групп (или разница в ожидаемом размере эффекта между группами сравнения), допустимый уровень альфа-ошибки и статистическая мощность	Да
	Недостаточно информации или информация о расчете объема выборки не приведена	Нет
3.	В гипотезе превосходства: анализировался ли набор данных по намерению лечить (все пациенты с назначенным)? В гипотезе неуступающей эффективности/эквивалентности: анализировался ли набор данных для пациентов, закончивших наблюдение (набор данных по протоколу)?	
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • В тексте публикации явно указано, что анализировался набор данных по намерению лечить или все пациенты, получившие хотя бы одну дозу вмешательства (для гипотезы превосходства). • Количество анализируемых пациентов по первичному исходу совпадало с количеством пациентов, распределенных по группам (для гипотезы превосходства). • Анализировался набор данных для пациентов, закончивших наблюдение (для гипотез неуступающей эффективности или эквивалентности). • Количество анализируемых пациентов было меньше, чем количество пациентов, распределенных по группам. • Первичным исходом является анализ времени до события. 	Да

	<p>Выполнение хотя бы одного из следующих условий:</p> <ul style="list-style-type: none"> • В гипотезе превосходства анализировался набор данных для пациентов, закончивших наблюдение. • В гипотезах неуступающей эффективности или эквивалентности анализировался набор данных по намерению лечить. • Нет информации об использованном наборе данных. 	Нет
4.	Описан ли метод заполнения пропусков по выбывшим пациентам в случае анализа набора данных по намерению лечить?	
	<p>В исследовании применен один из следующих вариантов заполнения пропусков:</p> <ul style="list-style-type: none"> • данные последнего наблюдения; • оптимистический сценарий (у выбывших пациентов лечение считалось успешным); • пессимистический сценарий (у выбывших пациентов лечение считалось неуспешным); • заполнение пропусков медианами или средними значениями признака по всей выборке (в зависимости от вида распределения признака). 	Да
	Не приведена информация о способе заполнения пропусков	Нет
	Использован набор данных для пациентов, закончивших наблюдение	Неприменимо
5.	В случае анализа набора данных для пациентов, закончивших наблюдение: сопоставлены ли частоты и времена выбывания для выбывших, исключенных, низкокомплаентных, пропусков и проч.?	
	Частоты и времена выбывания в группах сравнения сопоставлены и/или описаны	Да
	Частоты и времена выбывания в группах сравнения не сопоставлены и не описаны	Нет
	Использован набор данных по намерению лечить	Неприменимо
6.	Перечислены ли применяемые статистические методы и адекватны ли они дизайну исследования?	
	Применяемые статистические методы перечислены и адекватны дизайну исследования	Да
	<p>Выполнение хотя бы одного условия:</p> <ul style="list-style-type: none"> • Применяемые статистические методы не перечислены. • Применяемые статистические методы перечислены, но неадекватны дизайну исследования. 	Нет
7.	Указан ли пороговый уровень статистической значимости?	
	Пороговый уровень статистической значимости указан в разделе «Материалы и методы»	Да
	Пороговый уровень статистической значимости не указан в разделе «Материалы и методы»	Нет
8.	Применены ли меры по преодолению проблемы множественных сравнений (для анализа подгрупп или при проверке нескольких статистических гипотез в одной выборке)?	
	<p>Применен один из следующих методов преодоления проблемы множественных сравнений:</p> <ul style="list-style-type: none"> • поправка Бонферрони; • post hoc тесты; • другие валидизированные статистические инструменты преодоления проблемы множественных сравнений; • уменьшение порогового уровня статистической значимости. 	Да
	Не применены никакие способы преодоления проблемы	Нет

	множественных сравнений	
	В исследовании не проводился анализ подгрупп или не проверялось несколько статистических гипотез в одной выборке	Неприменимо
9.	Использован ли общеизвестный статистический пакет?	
	Использован общеизвестный статистический пакет (Statistica, SPSS, SAS или др.), MS Excel или онлайн-калькуляторы (со ссылкой на источник)	Да
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Использованы невалидизированные статистические пакеты (например, собственные разработки программного обеспечения). • Не использованы статистические пакеты или онлайн-калькуляторы. • Не приведена информация об используемом программном обеспечении. 	Нет
Раздел «Результаты»		
10.	Приведена ли описательная статистика и сопоставление исходных характеристик групп в анализируемой публикации (или других публикациях о данном исследовании)?	
	Приведена описательная статистика исходных характеристик групп и сопоставление исходных характеристик групп (в виде рассчитанных уровней статистической значимости) по всем или хотя бы основным характеристикам групп	Да
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Не приведена описательная статистика исходных характеристик групп и сопоставление исходных характеристик групп (в виде рассчитанных уровней статистической значимости). • Приведена описательная статистика исходных характеристик групп, однако группы не сопоставлены ни по одной из характеристик. • Приведена описательная статистика исходных характеристик групп, группы сопоставлены по всем или хотя бы по основным характеристикам групп, однако результаты данного сопоставления не приведены в виде рассчитанных уровней статистической значимости. 	Нет
11.	Если выявлены различия хотя бы по одной из исходных характеристик групп (дисбаланс групп), проведен ли анализ соответствующих ковариат или анализ в соответствующих подгруппах?	
	Анализ соответствующих ковариат или анализ в соответствующих подгруппах проведен	Да
	Анализ соответствующих ковариат или анализ в соответствующих подгруппах не проведен	Нет
	Выполнение хотя бы одного из условий: <ul style="list-style-type: none"> • Дисбаланс групп в результате проверки статистических гипотез не выявлен. • Проверка статистических гипотез об отсутствии различий в группах сравнения по исходным характеристикам не проводилась. • Описательная статистика исходных характеристик групп не представлена ни в анализируемой публикации, ни в других публикациях о данном исследовании. 	Неприменимо
12.	Ясно ли, какие описательные статистики использованы при представлении	

	распределений количественных признаков?	
	Информация об использованных описательных статистиках (мерах центральной тенденции и мерах их рассеяния) приведена в публикации в явном виде	Да
	Информация об использованных описательных статистиках (мерах центральной тенденции и мерах их рассеяния) не приведена в публикации в явном виде	Нет
	Отсутствие количественных признаков среди исходных характеристик групп	Неприменимо
13.	Отсутствуют ли ошибки в представлении описательной статистики количественных признаков?	
	Выполнение обоих следующих условий: <ul style="list-style-type: none"> • Для нормально распределенных данных использованы среднее и среднеквадратическое отклонение ($M (SD)$), при этом нормальность распределения доказана. • Для данных с распределением, отличающимся от нормального, или неизвестным видом распределения использованы медианы и интерквартильный размах ($Me [Q1; Q2]$). 	Да
	Выполнение хотя бы одного из условий: <ul style="list-style-type: none"> • Для признаков с распределением, отличным от нормального, или признаков с неизвестным видом распределения использованы среднее и среднеквадратическое отклонение $M (SD)$. • Среднеквадратическое отклонение больше половины среднего хотя бы в одном случае. • В качестве меры рассеяния среднего использована стандартная ошибка среднего (m). 	Нет
	Отсутствие количественных признаков среди исходных характеристик групп	Неприменимо
14.	Отсутствуют ли ошибки в описании качественных признаков?	
	Выполнение обоих следующих условий: <ul style="list-style-type: none"> • Всегда ясно, что принято за 100%. • Сумма долей по всей структуре явления составляет 100%. 	Да
	Невыполнение хотя бы одного из обозначенных выше условий	Нет
	Отсутствие качественных признаков среди исходных характеристик групп	Неприменимо
15.	Отсутствуют ли ошибки в сопоставлении относительных частот²?	
	Ошибки в сопоставлении относительных частот отсутствуют (по результатам выборочной проверка хотя бы по первичным критериям эффективности и/или безопасности)	Да
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Выявлены ошибки в сопоставлении относительных частот по результатам выборочной проверки. • Сопоставление относительных частот предусмотрено дизайном исследования, однако относительные частоты не сопоставлены. 	Нет
	Сопоставление относительных частот не проводилось, т.к. не предусмотрено дизайном исследования	Неприменимо
16.	Приведено ли в таблицах и на рисунках число наблюдений для каждого признака в каждой точке наблюдения?	

² Способы проверки результатов СА по данным, представленным в тексте публикации о клиническом исследовании, приведены в Приложении 2.

	Во всех таблицах и на всех рисунках приведено число наблюдений для каждого признака в каждой временной точке наблюдения (или в основных критически важных временных точках)	Да
	Не во всех таблицах и не на всех рисунках приведено число наблюдений для каждого признака в каждой временной точке наблюдения (или в основных критически важных временных точках)	Нет
	В публикации об исследовании отсутствуют таблицы и рисунки	Неприменимо
17.	Приведено ли точное значение рассчитанных уровней статистической значимости P (два-три десятичных разряда), а не неравенство?	
	Приведены точные значения всех рассчитанных уровней статистической значимости P (кроме случаев, когда рассчитанный уровень статистической значимости $P < 0,001$)	Да
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Не приведены точные значения P для всех рассчитанных уровней статистической значимости. • В публикации об исследовании не приведены рассчитанные уровни статистической значимости. 	Нет
18.	Ясно ли, какими методами были получены приведенные уровни статистической значимости?	
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Для каждого из рассчитанных уровней статистической значимости указан метод, которым получен данный уровень статистической значимости. • В перечне использованных методов статистического анализа присутствует метод, адекватный данному анализу. 	Да
	Невыполнение хотя бы одного из обозначенных выше условий	Нет
19.	В случае применения параметрических методов приведены ли обоснования их применимости?	
	Приведены обоснования применения параметрических методов статистического анализа (хотя бы один из вариантов): <ul style="list-style-type: none"> • Проверены гипотезы о нормальности распределения признаков, для которых использованы параметрические методы статистического анализа и приведены результаты данной проверки (рассчитанные уровни статистической значимости). • Приведено графическое обоснование нормальности распределений. • Доказано, что среднее, медиана и мода для данных признаков совпадают. 	Да
	Не приведены обоснования применения параметрических методов статистического анализа	Нет
	Параметрические методы статистического анализа не применялись	Неприменимо
20.	Приведены ли доверительные интервалы для основных результатов исследования?	
	Доверительные интервалы приведены для всех основных результатов исследования (первичных исходов)	Да
	Доверительные интервалы не приведены хотя бы для одного основного результата исследования	Нет
Разделы «Заключение» и/или «Выводы»		
21.	Если цель исследования не достигнута (согласно научной гипотезе исследования), достаточна ли мощность исследования?	
	Цель исследования не достигнута, и статистическая мощность исследования составляет 80% или более (по результатам расчета)	Да

	размера выборки, приведенным в разделе «Материалы и методы»)	
	Выполнение хотя бы одного из следующих условий: <ul style="list-style-type: none"> • Цель исследования не достигнута, но статистическая мощность исследования составляет менее 80%. • Цель исследования не достигнута, но оценить статистическую мощность невозможно. 	Нет
	Цель исследования достигнута	Неприменимо
22.	Если цель исследования не достигнута (согласно научной гипотезе исследования), корректна ли интерпретация результатов? Корректно ли сформулированы выводы?	
	Цель исследования не достигнута, при этом выводы не противоречат результатам исследования	Да
	Цель исследования не достигнута, при этом выводы противоречат результатам исследования	Нет
	Цель исследования достигнута	Неприменимо

Таблица 5. Оценка качества проведения статистического анализа в клинических исследованиях и представления его результатов в публикации о клиническом исследовании

№	Вопросы	Да	Нет	Неприменимо
Раздел «Материалы и методы»				
23.	Указана ли основная проверяемая гипотеза и критерий ее принятия?			
24.	Выполнялся ли расчет необходимого объема выборки по основному критерию оценки эффективности или безопасности с учетом вида проверяемой гипотезы?			
25.	<u>В гипотезе превосходства:</u> анализировался ли набор данных по намерению лечить? <u>В гипотезе неуступающей эффективности/эквивалентности:</u> анализировался ли набор данных для пациентов, закончивших наблюдение?			
26.	Описан ли метод заполнения пропусков по выбывшим пациентам в случае анализа данных по намерению лечить?			
27.	В случае анализа набора данных для пациентов, закончивших наблюдение: сопоставлены ли частоты и времена выбывания для выбывших, исключенных, низкокомплаентных, пропусков и проч.?			
28.	Перечислены ли применяемые статистические методы и адекватны ли они дизайну исследования?			
29.	Указан ли пороговый уровень статистической значимости?			
30.	Применены ли меры по преодолению проблемы множественных			

	сравнений (для анализа подгрупп или при проверке нескольких статистических гипотез в одной выборке)?			
31.	Использован ли общеизвестный статистический пакет?			
Раздел «Результаты»				
32.	Приведены ли описательная статистика и сопоставление исходных характеристик групп в анализируемой публикации (или других публикациях о данном исследовании)?			
33.	Если выявлены различия хотя бы по одной из исходных характеристик групп (дисбаланс групп), проведен ли анализ соответствующих ковариат или анализ в соответствующих подгруппах?			
34.	Ясно ли, какие описательные статистики использованы при представлении распределений количественных признаков?			
35.	Отсутствуют ли ошибки в представлении описательной статистики количественных признаков?			
36.	Отсутствуют ли ошибки в описании качественных признаков?			
37.	Отсутствуют ли ошибки в сопоставлении относительных частот?*			
38.	Приведено ли в таблицах и на рисунках число наблюдений для каждого признака в каждой точке наблюдения?			
39.	Приведено ли точное значение рассчитанных уровней статистической значимости P (два-три десятичных разряда), а не неравенство?			
40.	Ясно ли, какими методами могут быть получены приведенные			

	уровни статистической значимости?			
41.	В случае применения параметрических методов приведены ли обоснования их применимости?			
42.	Приведены ли доверительные интервалы для основных результатов исследования?			
Разделы «Заключение» и/или «Выводы»				
43.	Если цель исследования не достигнута (согласно научной гипотезе исследования), достаточна ли мощность исследования?***			
44.	Если цель исследования не достигнута (согласно научной гипотезе исследования), корректна ли интерпретация результатов? Корректно ли сформулированы выводы?			
ИТОГО				
% (ДА/(ДА+НЕТ))				
Заключение об уровне риска случайных ошибок***				

*в случае, если проверка невозможна, домен оценивается как «Нет»

**оценивается по фактическому объему групп и данных в разделе «Материалы и методы» о рассчитанной мощности

***стратификация риска случайных ошибок вычисляется по доле ответов «Да» среди всех ответов «Да» и «Нет»:

- 67-100% - низкий риск возникновения случайных ошибок;
- 34-66% - средний риск возникновения случайных ошибок;
- 0-33% - высокий риск возникновения случайных ошибок.

Литература

1. Федеральный закон от 12.04.2010 N 61-ФЗ (ред. от 29.12.2015) «Об обращении лекарственных средств».
2. Методические рекомендации по оценке сравнительной клинической эффективности безопасности лекарственного препарата. ФГБУ «ЦЭКМП» МЗ РФ [Электронный ресурс]. Режим доступа: <http://rosmedex.ru/wp-content/uploads/2016/12/MR-E%60B-23.12.2016.pdf>
3. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA //М.: МедиаСфера. – 2002. – Т. 305.
4. Ласт Д. Эпидемиологический словарь/под ред. //Джона М. Ласта. М.: Глобус. – 2009.
5. Реброва О. Ю. Описание статистического анализа данных в оригинальных статьях. Типичные ошибки //Медицинские технологии. Оценка и выбор. – 2011. – №. 4

Приложение 1. Перечень терминов на английском языке

Гипотеза неуступающей эффективности – non-inferiority

Гипотеза превосходства – superiority

Гипотеза эквивалентности – equivalence

Данные последнего наблюдения – last observation carried forward (LOCF)

Набор данных всех пациентов с назначенным лечением, набор данных по намерению лечить –full analysis set, intention-to-treat (ITT)

Набор данных по протоколу - per protocol (PP), valid cases, efficacy sample, evaluable subjects

Относительный риск – relative risk

Отношение угроз – hazard ratio

Отношение шансов – odds ratio

Первичный исход – primary outcome/end point/variable

Разность средних – mean difference

Исходные характеристики групп – baseline characteristics

Приложение 2. Методика дополнительных расчетов для проверки заключений/выводов в клинических исследованиях

При анализе публикации о клиническом исследовании и оценке риска случайных ошибок рекомендуется выполнить проверку основных результатов исследования (хотя бы по первичному исходу, на основании которого делаются основные выводы исследования). Проверка осуществляется с использованием данных, доступных в полных текстах публикаций о клинических исследованиях и/или дополнительных материалах к данным публикациям. Существует два основных варианта проверки результатов:

- 1) Расчет ДИ (для любых исходов);
- 2) Расчет уровней значимости (для бинарных исходов).

В случае, если для основных результатов, приведенных в публикации о клиническом исследовании не приведены ДИ, рекомендуется выполнить расчет ДИ для таких исходов (критериев эффективности и/или безопасности исследуемой технологии). ДИ отражает диапазон, в который с определенной вероятностью попадает истинное значение показателя, для которого рассчитан данный ДИ (как правило, рассчитывается 95% ДИ, который показывает, что с вероятностью 95% в ДИ попадает истинное значение анализируемого показателя в популяции).

Для бинарных исходов: как правило из полного текста публикации возможно извлечь информацию о количестве пациентов с наблюдаемым исходом (и/или долю (%) пациентов с наблюдаемым исходом) и общем количестве пациентов в каждой из групп. С целью проверки для бинарных исходов рассчитываются ДИ для относительных рисков (ОР), отношения шансов (ОШ) или разности абсолютных рисков (РАР) и сами эти показатели, а также осуществляется проверка статистической гипотезы об отсутствии различий между группами (рассчитывается уровень статистической значимости с применением ТКФ).

С целью расчета данных показателей можно воспользоваться любым доступным официальным программным обеспечением или онлайн-калькуляторами, размещенными в сети Интернет, например, 2-way Contingency Table Analysis (<http://statpages.info/ctab2x2.html>) или GraphPad (<http://www.graphpad.com/quickcalcs/contingency1/>). Для расчета ОР, ОШ, РАР, их ДИ и уровней значимости в интерфейсе выбранного онлайн-калькулятора (оба указанных нами калькулятора имеют интерфейс на английском языке) следует заполнить ячейки четырехпольной таблицы абсолютными частотами пациентов, имеющих изучаемый исход, и пациентов без изучаемого исхода в каждой из групп сравнения, а затем выполнить расчеты, нажав кнопку «Calculate» («Вычислить»).

Заключение о наличии статистически значимых различий делается преимущественно на основании рассчитанного уровня статистической значимости (как правило, в случае, если рассчитанный уровень статистической значимости меньше 0,05, статистически значимые

различия считаются выявленными, если не делалась поправка на проблему множественных сравнений), а также ДИ (если ДИ для ОР или ОШ содержит 1 или ДИ для РАР содержит 0, статистически значимые различия между группами не считаются выявленными).

В случае, если результаты расчета уровня статистической значимости и ДИ позволяют сделать согласованные выводы о наличии статистически значимых различий между группами сравнения, делается заключение о том, что ошибки в сопоставлении относительных частот не выявлены.

Для количественных исходов также, как правило, возможна проверка. Для проверки результатов, полученных по количественным исходам, рекомендуется рассчитать ДИ для средних значений исследуемых признаков в каждой из групп сравнения согласно следующей формуле: $[M-1,96m; M+1,96m]$, где m – стандартная ошибка среднего. При отсутствии в тексте публикации о клиническом исследовании значений m , данное значение вычисляется следующим образом:

$$m = \text{СКО}/\sqrt{n},$$

где n - количество пациентов в группе, для которой рассчитывается среднее.

Поскольку расчет уровней значимости для количественных переменных на основании данных, представленных в публикации о клиническом исследовании, не является возможным, заключение о наличии статистически значимых различий между группами по определенному критерию может быть сделано только на основании ДИ: в случае, если ДИ пересекаются, статистически значимые различия не считаются выявленными.

Для остальных параметров (таких как отношение угроз, разность средних и др.) проверка расчетов на основании данных, представленных в тексте публикации о клиническом исследовании, не является возможной, поскольку для такой проверки необходимы первичные данные исследования и/или использование статистического пакета.

Отметим, что заключение о наличии статистически значимых различий между группами на основании расчета ДИ и на основании расчета уровней значимости несколько отличаются по интерпретации:

- проверка статистических гипотез позволяет оценить вероятность неверного отклонения нулевой гипотезы об отсутствии статистически значимых различий между группами сравнения (однако при этом ничего не говорит о вероятности альтернативной гипотезы);
- ДИ позволяет оценить размер эффекта и его точность [5].

Таким образом, два способа выявления статистически значимых различий между группами сравнения дополняют друг друга, и их совместное использование повышает степень уверенности в полученных результатах.