

ГЛАВА 5

СТАТИСТИЧЕСКИЕ ПРИНЦИПЫ ПРОВЕДЕНИЯ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

*СОСТАВИТЕЛИ: к. б. н. А.Н. Васильев; Е.В. Гавришина; д. м. н. Д.В. Горячев;
д. м. н., профессор А.Н. Миронов; к. м. н. Р.Р. Ниязов;
Е.Ю. Парфенова; к. фарм. н. И.В. Сакаева*

5.1. ВВЕДЕНИЕ

5.1.1. Предпосылки и цель

Эффективность и безопасность лекарственных препаратов должны быть подтверждены результатами клинических исследований, проведенных согласно требованиям Национального стандарта Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика». В этом документе указано, что статистика играет ключевую роль при планировании и анализе клинических исследований. Широкое внедрение статистических исследований в область клинической разработки наряду с ключевой ролью клинических исследований для процесса государственной регистрации лекарственных препаратов и здравоохранения в целом требует емкого документа по статистическим вопросам клинических исследований. Основной целью настоящей главы Руководства является гармония принципов статистической методологии, используемые при проведении регистрационных клинических исследований в Европе, Японии и США.

Отправной точкой при создании настоящего документа являлось Руководство Комитета по лекарственным препаратам для медицинского применения по биостатистической методологии в регистрационных клинических исследованиях лекарственных препаратов (декабрь 1994 г.). На этот процесс также повлияли Руководство по статистическому анализу клинических исследований (март 1992 г.) Министерства здравоохранения и благополучия Японии и Руководство по формату и содержанию клинических и статистических разделов регистрационного досье оригинальных лекарственных препаратов США (июль 1988 г.). Некоторые вопросы статистической методологии также представлены в других методических рекомендациях и публикациях, в частности, указанных ниже.

Глава 6 настоящего издания «Структура и содержание отчетов о клинических исследованиях».

Глава 3 настоящего издания «Общие принципы проведения клинических исследований».

Методические рекомендации по проведению клинических исследований в особых группах пациентов (в разработке).

Национальный стандарт Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика».

Глава 4 настоящего издания «Составление протокола контролируемого клинического исследования лекарственного препарата (выбор контрольной группы)».

Глава 1 настоящего издания «Доклинические исследования безопасности с целью проведения клинических исследований и государственной регистрации лекарственных препаратов».

Вопросы, освещенные в других документах, будут отдельно обозначены в тексте.

Целью настоящей главы является представление спонсорам руководства по дизайну, проведению, анализу и оценке клинических исследований лекарственных препаратов с позиций их целостной клинической разработки. Документ также призван помочь специалистам в подготовке резюме по регистрационному досье и оценке доказательств эффективности и безопасности, преимущественно по результатам клинических исследований поздних фаз разработки.

5.1.2. Сфера применения и направление

Предметом настоящей главы являются статистические принципы. В них не рассматривается использование отдельных статистических процедур и методов. Обязанностью спонсора является обосновать, что использованные статистические процедуры надлежащим образом удовлетворяют указанным принципам. В настоящей главе обсуждается интеграция данных, полученных по результатам клинических исследований, однако эта тема основной не является. Ряд принципов и процедур, относящихся к управлению данными и мониторингу клинических исследований, освещен в других методических рекомендациях и в настоящих не рассматривается.

Настоящий документ предназначен для специалистов различного профиля. Тем не менее, согласно Национальному стандарту Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика» обязанность статистической обработки клинических исследований возлагается на квалифицированного и опытного статистика. Обязанностью статистика исследования (см. Глоссарий) является обеспечение совместно с другими участниками исследования соблюдения статистических принципов при проведении клинических исследований при разработке лекарственного препарата. Таким образом, статистик исследования должен обладать достаточными знаниями/умениями и опытом для реализации принципов, обозначенных в настоящих методических рекомендациях.

Принципы, изложенные в настоящей главе, в первую очередь применимы к проведению клинических исследований поздних фаз разработки, большинство из которых является исследованиями, подтверждающими эффективность. В дополнение к эффективности, основной переменной подтверждающих исследований могут служить переменные безопасности (например, нежелательное явление, лабораторный или инструментальный показатель), фармакодинамические или фармакокинетические переменные (например, в подтверждающем исследовании биоэквивалентности). Более того, некоторые подтверждающие данные могут быть результатом интеграции данных различных исследований; ряд принципов обращения с такими ситуациями рассматривается в настоящем документе. Наконец, несмотря на то что ранние фазы разработки лекарственного препарата состоят преимущественно из поисковых исследований, к ним также применимы статистические принципы. В этой связи положения настоящего документа должны, по возможности, распространяться на все фазы клинической разработки.

Многие из описанных в настоящей главе принципов направлены на минимизацию систематических ошибок (см. Глоссарий) и максимизацию точности. В настоящем документе под систематической ошибкой понимается систематическая тенденция какого-либо аспекта дизайна, проведения, анализа или интерпретации результатов клинического исследования отклонять оценку свойств лекарственного препарата (см. Глоссарий) от ее истинного значения. Важно как можно полнее выявить потенциальные источники систематических ошибок с целью последующей их минимизации. Наличие систематических ошибок может значительно снизить способность получения надежных выводов по результатам проведенных клинических исследований.

Некоторые источники систематических ошибок проистекают из дизайна исследования, например, такое распределение субъектов по группам исследования, когда часть пациентов с низким риском систематически попадает в группу исследуемого лекарственного препарата. Другими источниками систематических ошибок служат тактика проведения и анализа результатов клинического исследования. Например, нарушения протокола и исключение субъектов из анализа, основанное на знании исходов, является потенциальным источником систематических ошибок, которые могут повлиять на точность оценки эффекта лечения. Ввиду того, что систематические ошибки могут быть трудно обнаружимыми или возникать по неизвестным причинам, а их влияние сложно поддается прямому измерению, важно оценить надежность результатов и первичные выводы исследования. Надежность (устойчивость, *robustness*) — понятие, связанное с чувствительностью общих выводов к различным ограничениям по полученным данным, допущениям и подходам к анализу результатов. Надежность подразумевает, что эффект лечения и первичные выводы исследования незначительно изменяются при анализе данных на основании альтернативных допущений или с помощью иных аналитических подходов. При интерпретации статистической неопределенности эффекта лечения и сравниваемых методов лечения необходимо рассмотреть потенциальный вклад систематических ошибок в величину р-значения (вероятности отклонения нулевой гипотезы), доверительный интервал и выводы.

Ввиду преимущественного использования частотных статистических методов при дизайне и анализе клинических исследований, при обсуждении тестирования гипотезы и доверительных интервалов в настоящем документе рассматриваются преимущественно частотные методы (см. Глоссарий), что не следует рассматривать как единственно возможный подход: допускается использовать байесовский (см. Глоссарий) и другие подходы, если на то имеются достаточные основания, а полученные выводы вполне надежны.

5.2. ОБЩИЕ ВОПРОСЫ КЛИНИЧЕСКОЙ РАЗРАБОТКИ

5.2.1. Контекст исследования

5.2.1.1. План разработки

Наиболее общей целью процесса клинической разработки нового лекарственного препарата является выяснение диапазона доз и схемы лечения, при которых он может быть эффективным и безопасным в такой степени, что отношение ожидаемой пользы к возможному риску применения будет считаться благоприятным. Также необходимо определить непосредственную группу пациентов, которая получит пользу от лечения, и соответствующие показания к применению.

Для удовлетворения указанным целям, как правило, необходимо реализовать программу клинических исследований, каждое из которых предназначено для ответа на определенный вопрос (см. главу 3 настоящего издания «Общие принципы проведения клинических исследований»). Все это необходимо предусмотреть в плане клинической разработки или наборе таких планов с указанием этапов принятия решений и возможностью их изменения при получении новых данных. При государственной регистрации лекарственного препарата необходимо четко описать основное содержание таких планов и вклад каждого исследования в их реализацию. Интерпретация и оценка доказательств, полученных после реализации всей программы клинических исследований, включает синтез результатов, полученных на основании отдельных исследований (см. раздел 6.7.2). Этому способствует принятие единых стандартов по ряду свойств отдельных клинических исследований, как то словари медицинских терминов, определения и сроки проведения ключевых измерений, об-

ращения с отклонениями от протокола и т. д. Если один и тот же вопрос рассматривается в нескольких исследованиях, то могут быть информативны статистическое резюме, обзор или мета-анализ (см. Глоссарий). По возможности такого рода анализы необходимо предусмотреть в плане разработки, чтобы заранее определить значимые клинические исследования и все необходимые общие элементы дизайна. В этом плане также необходимо предусмотреть решение и других основных статистических вопросов (при их наличии), которые могут повлиять на ряд исследований, включенных в такой план.

5.2.1.2. Подтверждающее исследование

Подтверждающее исследование — это надлежащим образом контролируемое исследование, в котором заранее выдвигаются гипотезы с последующей оценкой последних. Подтверждающие исследования проводят, как правило, для получения надежных доказательств эффективности и безопасности. В таких исследованиях ключевая гипотеза напрямую вытекает из основной цели исследования, ее выдвигают заранее, ее проверка осуществляется непосредственно по завершении исследования. По результатам подтверждающих исследований одинаково важно с высокой точностью оценить величину исследуемого эффекта лечения и ее клиническую значимость.

Подтверждающие исследования направлены на обеспечение надежных доказательств выдвинутых гипотез, и поэтому приверженность протоколу и стандартным операционным процедурам является особенно важной. Необходимо объяснить и документировать неизбежные изменения и изучить их влияние. В протоколе необходимо представить обоснование дизайна каждого такого исследования, а также прочие важные статистические аспекты, как то основные параметры планируемого анализа. В каждом исследовании необходимо рассматривать лишь ограниченный круг вопросов.

Для надежного подтверждения выдвинутых гипотез необходимо, чтобы результаты подтверждающих клинических исследований свидетельствовали о наличии клинической пользы исследуемого лекарственного препарата. В связи с чем результаты подтверждающих клинических исследований должны быть достаточными для ответа на каждый поставленный ключевой клинический вопрос, необходимый для ясного и однозначного доказательства эффективности и безопасности. К тому же, важно, чтобы основания обобщаемости (см. Глоссарий) результатов исследования в отношении рассматриваемой популяции пациентов были понятны и объяснимы, это может также повлиять на количество и вид необходимых центров (например, специализированные или общей практики) и (или) исследований. Результаты подтверждающих исследований должны быть надежными. В некоторых случаях достаточно проведения одного подтверждающего исследования.

5.2.1.3. Поискное исследование

Основания и дизайн подтверждающих клинических исследований почти всегда проистекают из ранее проведенных поисковых клинических исследований. Как и в отношении иных клинических исследований, поисковые исследования должны иметь ясные и точные цели. Однако в отличие от подтверждающих исследований их целью не всегда является «простое» тестирование заранее выработанной гипотезы. К тому же поисковые исследования иногда требуют более гибкого дизайна, чтобы позволить изменять его по результатам накапливающихся данных. Их анализ может заключаться в изучении данных, в принципе, может быть проведено тестирование гипотезы, но ее выбор зависит от получаемых данных. Такого рода исследования не могут служить формальным основанием подтверждения эффективности, хотя они могут вносить определенный вклад в общий массив данных по эффективности.

Любое исследование может состоять из поисковых и подтверждающих аспектов. Например, результаты большинства подтверждающих исследований рассматриваются с позиций поискового анализа обнаруженных данных и дальнейшей выработки новых гипотез для последующих исследований. В протоколе необходимо четко обозначить подтверждающие аспекты и аспекты, которые подлежат поисковому анализу.

5.2.2. Параметры исследований

5.2.2.1. Популяция

На ранних этапах разработки лекарственного препарата на выбор субъектов клинического исследования может существенно повлиять намерение повысить вероятность выявления искомым клинических эффектов, вследствие чего изучению будет подвергаться узкая подгруппа пациентов, по сравнению с общей популяцией, которой будет предназначено лечение. Однако ко времени проведения подтверждающих исследований их субъекты должны в большей степени отражать целевую популяцию. В этой связи для таких исследований целесообразно максимально ослабить критерии включения и невключения в отношении целевой популяции, поддерживая при этом достаточную однородность с целью точной оценки терапевтического эффекта. В силу возможного географического местоположения, времени проведения, центров и подходов исследователей к лечению и т. д. ни одно отдельно взятое клиническое исследование не способно полностью отразить характеристики всех будущих пациентов. Тем не менее, по возможности, необходимо снизить влияние указанных факторов, а также проанализировать их при интерпретации результатов исследования.

5.2.2.2. Первичные и вторичные переменные

Первичная переменная («целевая» переменная, первичная конечная точка) — это переменная, способная обеспечить наибольшую клиническую значимость и убедительное доказательство, напрямую относящееся к основной цели исследования. Рекомендуется использовать только одну первичную переменную. Обычно это переменная эффективности, так как основной целью большинства подтверждающих исследований является обеспечение строгих научных доказательств эффективности. В некоторых случаях первичной переменной могут служить безопасность/переносимость, которые всегда являются важной составляющей. Определение качества жизни и фармакоэкономической целесообразности в дальнейшем также могут стать первичными переменными. Выбор первичной переменной должен отражать общепринятые нормы и стандарты в соответствующей области исследования. Рекомендуется использовать надежную, валидированную переменную, в отношении которой получен опыт по ранее проведенным клиническим исследованиям или которая описана в опубликованной литературе. Необходимо обосновать, что первичная переменная способна надежно и с высокой степенью воспроизводимости измерять некоторую важную, клинически значимую терапевтическую пользу у группы пациентов, охарактеризованную критериями включения и невключения. Первичную переменную, как правило, используют для определения размера выборки.

Во многих случаях, подход, основанный на оценке исходов субъектов, может оказаться неоднозначным и потребовать тщательного описания. Например, недопустимо без достаточных разъяснений выбирать смертность (летальность) в качестве первичной переменной, так как ее можно выразить, сравнивая долю живых в определенный момент времени или сопоставляя общее распределение выживаемости за некоторый период. В качестве другого примера можно рассмотреть рецидивирующее явление: так мерой терапевтического эффекта могут быть простая переменная с двумя исходами (возникновение явления или его отсутствие в течение определенного промежутка времени), время до первого возникновения явления, частота его возник-

новения (количество явлений за единицу времени наблюдения) и т. д. При изучении хронических заболеваний необходимость оценки функционального статуса во времени может также затруднить выбор первичной переменной. Существует ряд подходов: сравнение оценок в начале и в конце наблюдения; сравнение углов наклона, рассчитанных на основании всех оценок, полученных за период наблюдения; сравнение долей субъектов, достигающих или не достигающих некоторого порога; сравнение, основанное на методах, разработанных для множественных измерений и т. д. Во избежание проблем с множественными сравнениями, возникающих в силу определения первичной переменной по полученным результатам, в протоколе необходимо дать точное определение первичной переменной, которая будет использоваться в статистическом анализе. К тому же в протоколе необходимо описать и обосновать клиническую значимость выбранной первичной переменной и подтвердить пригодность соответствующих процедур измерения.

В протоколе необходимо указать первичную переменную, а также обосновать ее выбор. За редким исключением, пересмотр первичной переменной после снятия ослепления недопустим, так как возникающие вследствие этого систематические ошибки трудно поддаются оценке. Если клинический эффект, указанный в основной цели будет измеряться разными способами, в протоколе на основании клинической значимости, важности, объективности и (или) других значимых характеристик необходимо один из них указать в качестве первичной переменной, если такой выбор правомочен.

Вторичными переменными являются вспомогательные способы измерения достижения первичной цели или способы измерения, относящиеся к вторичным целям. В протоколе необходимо заранее их обозначить, а также обосновать их относительную значимость и роль в интерпретации результатов исследований. Количество вторичных переменных должно быть ограничено, они должны относиться к ограниченному числу вопросов, рассматриваемых в исследовании.

5.2.2.3. Составные переменные

Если первичную переменную невозможно выбрать из множества способов измерений достижения основной цели, в качестве альтернативы допускается интегрировать или совместить несколько способов измерения в одну «составную» переменную на основании заранее определенного алгоритма. В действительности, в некоторых случаях первичные переменные представляют собой комбинацию нескольких способов измерений (например, оценочные шкалы, используемые при артрите, психических расстройствах и др.). Такой подход позволяет решить проблему множественных измерений без введения поправки на ошибку I рода. В протоколе необходимо представить метод совмещения нескольких способов измерения, а также интерпретацию полученной в результате шкалы с точки зрения величины клинически значимой пользы. Если составная переменная используется в качестве первичной, в некоторых случаях допускается отдельный анализ составляющих такой переменной, если это клинически правомочно и целесообразно. Если в качестве первичной переменной используется рейтинговая шкала, особенно важно рассмотреть такие факторы, как содержательная валидность (см. Глоссарий), межэкспертная и внутриэкспертная надежность (см. Глоссарий) и способность реагировать на изменения в зависимости от тяжести заболевания.

5.2.2.4. Переменные общей (global) оценки

В некоторых случаях для измерения совокупной безопасности, совокупной эффективности и (или) иной совокупной пользы разрабатывают переменные «общей оценки» (см. Глоссарий). Этот вид переменных включает в себя объективные переменные и общее впечатление исследователя о состоянии или изменении состояния

субъекта, для чего обычно используются порядковые или категориальные шкалы. В некоторых областях, таких как неврология и психиатрия, методология осуществления общей оценки совокупной эффективности хорошо разработана.

Обычно переменные общей оценки включают субъективный компонент. При использовании переменной общей оценки в качестве первичной или вторичной, в протоколе исследования необходимо более подробно описать такую шкалу, а именно:

значимость шкалы при оценке первичной цели исследования;

данные, подтверждающие валидность и надежность шкалы;

способ обращения с данными, полученными от отдельного субъекта, с целью присвоения ему определенной категории по такой шкале;

способ присвоения субъектам определенной категории по шкале с недостающими данными или иной способ их оценки.

Если при осуществлении общей оценки исследователем рассматриваются объективные переменные, то такие переменные необходимо рассматривать как дополнительные первичные или, как минимум, важные вторичные переменные.

Общая оценка пользы включает в себя как пользу, так и риск и отражает процесс принятия решений лечащего врача, который должен сопоставить пользу и риск при принятии решения о назначении лекарственного препарата. Недостаток переменных общей оценки пользы заключается в том, что в некоторых случаях их использование может привести к выводу об эквивалентности двух сравниваемых лекарственных препаратов, обладающих резко отличающимися профилями эффективности и безопасности. Например, суждение на основании общей оценки, что лекарственный препарат эквивалентен или превосходит альтернативу, может маскировать низкую эффективность (или ее отсутствие) первого при небольшом количестве нежелательных явлений. Поэтому в качестве первичной переменной не рекомендуется использовать переменную общей пользы. Если общая польза принята за первичную переменную, в качестве дополнительных первичных переменных отдельно необходимо предусмотреть изучение исходов по эффективности и безопасности.

5.2.2.5. Множественные первичные переменные

В некоторых случаях желательно использовать несколько первичных переменных, каждая из которых (или их подмножество) может быть достаточной для изучения ряда эффектов лекарственных препаратов. Необходимо тщательно оговорить запланированный способ интерпретации такого рода доказательств. Необходимо четко обозначить, что будет считаться достижением целей исследования: изменение одной, некоторого минимума или всех переменных. Необходимо точно обозначить первичную гипотезу (гипотезы) и рассматриваемые параметры (например, среднее, доля, распределение) с точки зрения выбранных первичных переменных и представленного статистического подхода. Ввиду потенциального эффекта множественности (см. раздел 6.5.6) необходимо разъяснить влияние ошибки I рода, способ контроля над ней необходимо описать в протоколе. При оценке влияния ошибки I рода необходимо описать степень взаимной корреляции между предложенными первичными переменными. Если целью исследования является подтверждение влияния на все предусмотренные первичные переменные, тогда поправку на ошибку I рода вводить не требуется, однако необходимо подробно описать влияние ошибки II рода и размера выборки на эти показатели.

5.2.2.6. Суррогатные переменные

Если напрямую оценить клиническую пользу для субъекта путем наблюдения за реальной клинической эффективностью невозможно, допускается использовать не прямые критерии — суррогатные переменные (см. Глоссарий). При изучении ряда показаний используются общепринятые суррогатные переменные, которые, как счи-

тается, способны надежно прогнозировать клиническую пользу. Существуют два основных опасения использования суррогатных переменных. Во-первых, они могут не являться истинным предиктором исследуемого клинического исхода. Например, можно изучать активность терапии на основании определенного фармакологического механизма, который может не давать полной информации о всевозможных эффектах и ключевых показателях лечения, независимо от того, положительные они или отрицательные. Существует множество примеров, когда лекарственный препарат, оказывая выраженное благотворное влияние на предложенную суррогатную точку, приводил к неблагоприятному клиническому исходу у субъекта и, наоборот, в некоторых случаях лекарственный препарат приносил клиническую пользу без значимого влияния на предложенные суррогатные точки. Во-вторых, предложенные суррогатные переменные могут не давать количественной оценки клинической пользы, которую можно было бы напрямую сопоставить с нежелательными явлениями. Предложены статистические критерии валидации суррогатных переменных, но опыт их использования достаточно ограничен. На практике доказательная сила суррогатных переменных зависит от (1) биологического правдоподобия взаимосвязи [между суррогатной и конечной переменными], (2) подтверждения по результатам эпидемиологических исследований прогностической ценности суррогатных переменных в отношении клинического исхода и (3) доказательства по результатам клинических исследований, что влияние лекарственного препарата на суррогатные переменные соотносится с влиянием на клинический исход. Наличие взаимосвязи между клиническими и суррогатными переменными для одного лекарственного препарата необязательно является справедливым для лекарственного препарата с другим механизмом действия, применяющегося для лечения того же заболевания.

5.2.2.7. Категориальные переменные

В некоторых случаях желательно осуществлять дихотомическое деление или иное распределение по категориям непрерывных и порядковых (ранговых) переменных. Хорошо известным примером дихотомий являются критерии «успех» и «ответ», которые при этом требуют точного описания, например, минимальная доля улучшений (по сравнению с исходным состоянием) для непрерывной переменной или ранговая категория, соответствующая или превышающая определенный порог (например, «хорошо») для порядковой рейтинговой шкалы.

Общеизвестным примером дихотомического деления служит снижение диастолического давления ниже 90 мм рт. ст. Распределение по категориям наиболее ценно, когда такие категории имеют клиническую значимость. В протоколе необходимо описать и объяснить критерии распределения по категориям, так как осведомленность о результатах исследования может легко внести систематические ошибки при выборе таких критериев. В силу того, что распределение по категориям приводит к потере данных, его следствием является снижение мощности анализа, что необходимо учитывать при расчете размера выборки.

5.2.3. Подходы к дизайну с целью недопущения систематических ошибок

Наиболее важными элементами дизайна, служащими для минимизации систематических ошибок в клинических исследованиях, являются ослепление и рандомизация, которые должны стать стандартом для большинства контролируемых клинических исследований, результаты которых будут включены в регистрационное досье. В большинстве таких исследований используется метод двойного ослепления, при которой лекарственные препараты заранее упаковываются по соответствующей схеме рандомизации и доставляются в центр(ы), маркированные только номером субъекта и периодом лечения таким образом, чтобы ни один из участников исследования не был осведомлен о распределении субъектов по сравниваемым группам (в том чис-

ле в закодированном виде). Такой подход описан в разделе 5.2.3.1 и более подробно в разделе 5.2.3.2, исключения из правил приводятся в конце раздела.

Систематические ошибки можно минимизировать на этапе планирования исследования путем внедрения в протокол процедур, направленных на минимизацию любых возможных отклонений в проведении исследований, которые могут сказаться на осуществлении правильного анализа, включая различные нарушения протокола, выбывания из исследования и недостающие данные. В протоколе необходимо предусмотреть способы снижения частоты таких ситуаций и методы обращения с такими ситуациями, возникшими в ходе анализа данных.

5.2.3.1. Ослепление

Ослепление или маскировка направлено на ограничение возникновения умышленных или неумышленных систематических ошибок при проведении клинического исследования и интерпретации его результатов, возникающих вследствие влияния, которое может оказать осведомленность о назначенном лекарственном препарате на набор и распределение субъектов, последующий уход за ними, отношение субъектов к лекарственным препаратам, оценку конечных точек, способ обращения с выбыванием, исключение данных из анализа и т. д. Ключевая цель — предотвращение идентификации лекарственных препаратов до момента, пока не минуют все возможности для возникновения систематических ошибок.

Двойное слепое исследование — это такое исследование, при котором ни субъект, ни один из исследователей или представителей спонсора, вовлеченных в проведение или клиническую оценку субъектов, не знают о получаемом субъектом лекарственном препарате. К ним, помимо прочего, относятся все лица, определяющие годность субъектов, оценивающие конечные точки или проверяющие приверженность протоколу. Такая степень ослепления поддерживается на протяжении всего хода исследования, снятие ослепления осуществляется, когда количество данных достигает приемлемого уровня. Если с представителями спонсора, не вовлеченных в лечение или клиническую оценку субъектов, необходимо снять ослепление (например, биоаналитики, аудиторы, лица, вовлеченные в оповещение о серьезных нежелательных явлениях), спонсор должен предусмотреть соответствующую стандартную операционную процедуру для защиты от ненадлежащего раскрытия кодов лечения. В простом слепом исследовании исследователь и (или) представители спонсора, но не субъект, осведомлены о лекарственном препарате, или наоборот. В открытом исследовании о применяемом лекарственном препарате осведомлены все участники. Оптимальным подходом является двойное слепое исследование. Для этого необходимо, чтобы лекарственные препараты, применяемые в исследовании, были неотличимы (по внешнему виду, вкусу и др.) либо до, либо во время введения, а само ослепление поддерживалось на протяжении всего исследования.

На пути достижения двойного ослепления могут возникнуть препятствия: сравниваемые методы лечения могут существенно различаться по своей природе, например, хирургия и фармакотерапия; два лекарственных препарата могут отличаться по лекарственной форме, и хотя их можно сделать неотличимыми друг от друга путем помещения в капсулы, смена лекарственной формы может послужить причиной изменения фармакокинетических и (или) фармакодинамических свойств и потребовать подтверждения биоэквивалентности таких лекарственных форм; отличия также могут заключаться в суточном режиме дозирования двух лекарственных препаратов. В таких случаях для достижения двойного ослепления можно использовать метод «двойной пустышки» (см. Глоссарий). В некоторых случаях этот метод вынуждает использовать достаточно нестандартный режим дозирования, что может неблагоприятно сказаться на мотивации и приверженности субъектов. Также могут возникнуть этические затруднения, например, в связи с проведением ложных опе-

раций (операций-пустышек). Тем не менее, для преодоления указанных трудностей необходимо приложить максимум усилий.

Двойное ослепление в некоторых клинических исследованиях может быть частично снято ввиду явных терапевтических эффектов. В таких случаях ослепление можно частично повысить путем введения скрытия результатов некоторых исследований (например, определенных лабораторных показателей) для исследователей и представителей спонсора. Подобные подходы (см. ниже) для минимизации систематических ошибок в открытых исследованиях необходимо предусмотреть, где уникальные или редкие терапевтические эффекты могут привести к снятию ослепления с отдельных пациентов.

Если двойное ослепление неприменимо, необходимо воспользоваться простым ослеплением. В некоторых случаях с практических и этических позиций возможны только открытые исследования. Простые слепые и открытые исследования предоставляют дополнительную гибкость, однако важно, чтобы осведомленность исследователя о предполагаемом лечении не повлияла на решение о включении субъекта в исследование; такое решение должно предшествовать осведомленности о назначаемой терапии. Для таких исследований необходимо предусмотреть централизованную процедуру рандомизации, например, телефонную рандомизацию, по группам сравнения. В дополнение необходимо, чтобы клиническую оценку осуществлял медицинский персонал, не вовлеченный в лечение пациентов, который остается ослепленным к выбранному методу лечения. При проведении простых слепых и открытых исследований необходимо всеми возможными способами минимизировать источники систематических ошибок, а первичные переменные должны быть максимально объективными. Предполагаемую степень ослепления, а также другие способы для минимизации систематических ошибок необходимо описать в протоколе. Например, спонсор обязан предусмотреть стандартную операционную процедуру, которая обеспечивала бы ограничение доступа к рандомизационным кодам во время подготовки базы данных к анализу полученных результатов.

Снятие ослепления (с отдельного субъекта) допускается только тогда, когда знание лечащего врача о принимаемом субъектом лекарственном препарате необходимо в интересах последнего. О любом намеренном или случайном снятии ослепления необходимо сообщить и объяснить его по завершении исследования, независимо от его причин. Необходимо документировать процедуру и сроки вскрытия рандомизационных кодов.

В настоящем документе под проверкой данных без снятия ослепления (см. Глоссарий) подразумевается рассмотрение результатов исследования в период между его завершением (последнее наблюдение (визит) за последним субъектом) и снятием ослепления.

5.2.3.2. Рандомизация

Рандомизация намеренно вводит элемент случайности при распределении субъектов клинического исследования по группам сравнения. В ходе последующего анализа результатов исследования это обеспечивает надежную статистическую основу количественной оценки наличия эффектов лекарственного препарата. Это также способствует выравниванию сравниваемых групп по установленным и не установленным прогностическим факторам. Вместе с ослеплением рандомизация позволяет избежать возможных систематических ошибок при выборе и распределении субъектов, которые могли бы возникнуть в случае осведомленности о получаемом пациентом лекарственном препарате.

В схеме рандомизации клинического исследования отражают случайное распределение субъектов по сравниваемым группам. В наиболее простом случае — это упорядоченный перечень сравниваемых лекарственных препаратов (или групп срав-

нения в перекрестном исследовании) или кодов и соответствующих им номеров пациентов. Логистика некоторых исследований, например, в которых предусмотрена фаза скрининга, может усложнить рандомизацию, однако все это не должно отражаться на запланированном распределении субъектов по группам сравнения. Разнообразие процедур, генерирующих схему рандомизации, зависит от дизайна исследования. Схема рандомизации должна быть воспроизводимой (при необходимости).

Несмотря на то что неограниченная рандомизация допустима, осуществление блоковой рандомизации имеет некоторые преимущества. Это позволяет повысить сопоставимость сравниваемых групп, особенно если характеристики субъектов могут изменяться во времени, например, как результат изменения политики набора субъектов в исследование. Она также обеспечивает почти одинаковый размер сравниваемых групп. В перекрестных исследованиях блоковая рандомизация позволяет сбалансировать дизайн, повышая его эффективность и обеспечивая более простую интерпретацию результатов. Необходимо тщательно подбирать размер блоков: он должен быть достаточно маленьким, чтобы предотвратить потенциальную несбалансированность, и достаточно большим, чтобы избежать предсказуемости при завершении заполнения блока. Исследователи и другие заинтересованные стороны исследования не должны знать о размере блока (должны быть ослеплены), этого можно достичь, используя блоки различного размера и выбирая их в случайном порядке (предсказуемость в двойном слепом исследовании, теоретически, не является затруднением, однако фармакологические эффекты могут послужить причиной догадок о схеме рандомизации).

В многоцентровых исследованиях (см. Глоссарий) необходимо предусмотреть централизованную процедуру рандомизации. Рекомендуется использовать отдельную рандомизационную схему для каждого центра, то есть осуществить стратификацию по центру или распределить несколько полных блоков в каждый центр. Обобщая: стратификация по заранее установленным важным прогностическим факторам (например, тяжесть заболевания, возраст, пол и др.) может оказаться ценным способом повышения сбалансированности распределения в пределах страты; она более полезна в небольших исследованиях. Редко приходится использовать более 2–3 факторов стратификации, что, к тому же, затрудняет достижение сбалансированности и трудноосуществимо логистически. Использование процедуры динамического распределения (см. ниже) может позволить сбалансировать несколько используемых факторов стратификации, остальные используемые в исследовании процедуры можно адаптировать, чтобы приспособить дизайн к такому подходу. В ходе последующего анализа необходимо учесть факторы, по которым была проведена стратификация.

Каждого последующего субъекта, подлежащего рандомизации, необходимо распределять в группу, соответствующую очередному свободному номеру, предусмотренному в рандомизационной схеме (в соответствующую страту, если предусмотрена стратификация). Соответствующий номер присваивается субъекту, и его распределяют в соответствующую группу только после получения подтверждения, что он включен в рандомизированную часть исследования. В протоколе не должно содержаться подробного описания схемы рандомизации, которое повышало бы ее предсказуемость (например, размер блоков). Спонсор или независимая третья сторона должна хранить схему рандомизации в закрытом доступе, обеспечивая надлежащее ослепление на протяжении всего исследования. При этом необходимо предусмотреть возможность экстренного доступа к схеме рандомизации, чтобы обеспечить возможность снятия ослепления с любого субъекта. В протоколе необходимо описать процедуру снятия ослепления, сопроводительную документацию, оказанное впоследствии лечение и технику оценки исходов субъекта.

Альтернативой является динамическое распределение, при котором распределение субъектов по сравниваемым группам зависит от текущей сбалансированности

распределения, а при наличии стратификации по стратам и внутри них. Необходимо избегать детерминированного динамического распределения, а при его осуществлении необходимо использовать надлежащую степень рандомизации. Необходимо приложить все возможные усилия для поддержания двойного ослепления исследования. Например, осведомленность о схеме рандомизации должна быть ограничена центральным офисом исследования, из которого осуществляется контроль над динамическим распределением, как правило, путем телефонного контакта. Это, в свою очередь, позволяет ввести дополнительный контроль над включением в исследование, что крайне желательно для некоторых разновидностей многоцентровых исследований. Допускается использовать обычную для двойного слепого исследования схему предварительной упаковки и маркировки исследуемых лекарственных препаратов, но порядок их включения в исследование перестает быть последовательным. Рекомендуется использовать соответствующие компьютерные алгоритмы, чтобы сотрудники центрального офиса исследования оставались ослепленными в отношении схемы рандомизации. При планировании динамического распределения необходимо тщательно продумать сложность логистики и оценить его потенциальное влияние на анализ результатов.

5.3. ВОПРОСЫ ДИЗАЙНА ИССЛЕДОВАНИЯ

5.3.1. Конфигурация дизайна

5.3.1.1. Параллельный дизайн

Наиболее частым дизайном контролируемых клинических исследований является параллельный, при котором субъекты распределяются в одну из двух или более групп сравнения, при этом каждая из сравниваемых групп получает различные методы лечения. К таким различиям относятся назначение исследуемого лекарственного препарата в одной или более дозах и в одной и более сравниваемых группах, как то: плацебо и (или) активный контроль. Предпосылки, лежащие в основе такого дизайна, по сравнению с другими, более простые. Однако, как и при других дизайнах, некоторые дополнительные свойства исследования могут затруднить анализ и интерпретацию результатов (например, ковариаты, повторные измерения, взаимодействие различных факторов дизайна, нарушения протокола, выбывание (см. Глоссарий) и исключение из исследования).

5.3.1.2. Перекрестный дизайн

При перекрестном дизайне каждого субъекта рандомизируют в последовательность из двух или более методов лечения, таким образом, он служит контролем самому себе по сравниваемым методам лечения. Этот простой маневр привлекателен тем, что позволяет снизить (в некоторых случаях существенно) размер выборки и количество измерений, необходимых для достижения заданной мощности. В наиболее простом случае при перекресте 2×2 каждый субъект в случайном порядке получает оба сравниваемых лекарственных препарата в рамках двухэтапного исследования, которое часто разделено отмывочным периодом. В общем виде при перекрестном дизайне сравниваются n ($n > 2$) методов лечения в ходе n этапов, при этом каждый субъект получает все n методов лечения. Существуют различные вариации дизайна, при которых каждому субъекту назначают подмножество множества n методов лечения, или каждому субъекту несколько раз назначают одни и те же лекарственные препараты.

У перекрестных исследований есть несколько недостатков, которые могут повлиять на их результаты. Основным является перенос — остаточное влияние предыдущего лекарственного препарата на последующие. В аддитивной модели эффект неравномерного переноса вносит систематическую ошибку при прямом сравнении

исследуемых лекарственных препаратов. В дизайне 2×2 эффект переноса невозможно статистически отличить от влияния препарата и этапа, поскольку тесты обладают достаточной мощностью лишь для обнаружения влияния последних двух факторов на субъектов. Эта проблема не стоит так остро при использовании дизайнов более высоких порядков, но и в этом случае ее нельзя полностью исключить.

При использовании перекрестного дизайна необходимо стараться исключить перенос. Наилучшим образом это достигается тщательным выбором дизайна, руководствуясь всеми доступными знаниями о заболевании и новом лекарственном препарате. Заболевание должно иметь хроническое стабильное течение. Искомые эффекты лекарственного препарата должны полностью проявиться в течение периода лечения. Длительность отмывочного периода должна быть достаточной, чтобы эффекты лекарственного препарата полностью прекратились. Необходимо заранее обеспечить соблюдение указанных условий путем представления соответствующих данных.

Существуют и другие затруднения, требующие повышенного внимания к перекрестным дизайнам. Наиболее значимым является усложнение проведения анализа и интерпретации его результатов вследствие выбывания субъектов. К тому же вероятность переноса затрудняет отнесение нежелательных явлений, которые возникают на последующих этапах, к одному из сравниваемых лекарственных препаратов. Эти и другие вопросы описаны в ICH E4. Перекрестный дизайн необходимо использовать только в тех случаях, если ожидаемое количество выбываний из исследования небольшое.

Наиболее часто и в целом с удовлетворительными результатами перекрестный дизайн 2×2 используют для подтверждения биоэквивалентности двух лекарственных препаратов. В указанном частном случае влияние эффекта переноса на исследуемые фармакокинетические параметры у здоровых добровольцев маловероятно, если длительность отмывочного периода между этапами исследования достаточна. Тем не менее, важно проверять указанное допущение в ходе анализа, в частности, путем подтверждения, что на момент начала каждого из этапов исследуемые лекарственные препараты в крови не обнаруживаются.

5.3.1.3. Факторные дизайны

Использование факторного дизайна направлено на изучение двух и более лекарственных препаратов, которые назначаются одновременно в различных комбинациях. Наиболее простым является факторный дизайн 2×2 , когда субъектов в случайном порядке распределяют в одну из четырех групп двух лекарственных препаратов, например А и В: только А; только В; и А, и В; ни А, ни В. Во многих случаях этот дизайн направлен на выявление взаимодействия между А и В. Статистический тест на наличие взаимодействия может быть недостаточно мощным для его выявления, если размер выборки рассчитывался для оценки основных эффектов. Это является важным обстоятельством, если исследование направлено на установление совокупных эффектов А и В, в частности, при высокой вероятности комбинированной терапии.

Другим важным применением факторного дизайна является выявление взаимосвязи доза–эффект при одновременном применении лекарственных препаратов С и D, особенно, если ранее была установлена эффективность каждого из них в монотерапии. Выбираются m доз С, как правило, включая нулевую дозу (плацебо) и точно так же n доз D. Исследование состоит из $m \times n$ групп сравнения, каждая из которых получает комбинацию С и D в различных дозах. Затем результирующую оценку эффекта (поверхности отклика, response surface) используют для установления оптимальной комбинации С и D для клинического применения (см. методические рекомендации по проведению клинических исследований с целью подбора дозы).

В некоторых случаях дизайн 2×2 используют для «экономии» субъектов исследования при изучении эффективности двух лекарственных препаратов на одной и той же выборке, размер которой потребовался бы для установления каждого из них. Эта стратегия подтвердила свою ценность для крупных исследований по изучению смертности. Эффективность и надежность такого подхода зависит от отсутствия такого взаимодействия между лекарственными препаратами А и В, когда влияние А и В на первичные переменные эффективности соответствует аддитивной модели и, таким образом, эффект А фактически не зависит от дополнительного применения В. Так же, как и для перекрестного исследования на основании имеющихся данных, необходимо заранее удостовериться в соблюдении указанного условия.

5.3.2. Многоцентровые исследования

Многоцентровые исследования проводят по двум основным причинам. Во-первых, многоцентровое исследование — приемлемый способ более рационального изучения нового лекарственного препарата; в некоторых случаях это единственно возможный путь набора достаточного для достижения поставленных целей количества субъектов в течение приемлемого срока. Такого рода многоцентровые исследования допускается проводить на любом этапе клинической разработки. Центров может быть несколько с большим количеством субъектов в каждом или, для редких заболеваний, множество центров с несколькими субъектами в каждом.

Во-вторых, исследование планируется в качестве многоцентрового (с множеством исследователей) в основном для более надежной последующей обобщаемости полученных результатов. Этому способствует возможность набора субъектов из общей популяции и применения лекарственного препарата в различных клинических условиях, таким образом, экспериментально моделируется ситуация, которая больше соответствует будущим условиям применения. Вовлечение большего количества исследователей также способствует более широкой клинической оценке пользы лекарственного препарата. Такого рода исследование является подтверждающим и проводится на поздних этапах разработки лекарственного препарата, оно часто включает в себя множество исследователей и центров. С целью еще большего повышения обобщаемости (см. Глоссарий) исследование в некоторых случаях проводится в разных странах.

Для правильной интерпретации и экстраполяции результатов многоцентрового исследования его протокол должен однозначно трактоваться, а исследование должно проводиться единообразно во всех центрах. Более того, стандартный расчет размера выборки и вычисление мощности основываются на допущении, что различия между сравниваемыми лекарственными препаратами в центрах — равные несмещенные количественные характеристики. Необходимо создать общий протокол и проводить исследование с учетом этого. Все процедуры необходимо максимально стандартизировать. Разночтения в оценке критериев и схемах можно снизить, организовав встречи исследователей, заранее обучая персонал и осуществляя надлежащий мониторинг исследования. Целью надлежащего дизайна в целом должно служить достижение равномерного распределения субъектов по группам сравнения внутри центра, указанная цель должна поддерживаться эффективным управлением. В исследованиях, проводимых с участием относительно большого числа крупных центров с незначительными отличиями по числу субъектов, есть существенное преимущество: возникает возможность учета гетерогенности между исследовательскими центрами, так как в таких случаях уменьшаются различия между разными взвешенными оценками терапевтического эффекта (это указание не относится к исследованиям с множеством небольших центров, в которых свойство «центр» не включается в анализ). В сложных случаях неспособность соблюсти указанные предосторожности вместе с сомнениями относительно однородности результатов может снизить ценность мно-

гоцентрового исследования в такой степени, что результаты могут не подтвердить выдвигаемые гипотезы спонсора.

В простом многоцентровом исследовании каждый исследователь несет ответственность за субъектов, набранных в одном лечебно-профилактическом учреждении, таким образом, одному центру соответствует один исследователь или одно лечебно-профилактическое учреждение. Однако во многих случаях наблюдается более сложная картина. Один исследователь может набирать субъектов из нескольких лечебно-профилактических учреждений, он также может представлять команду клиницистов (соисследователей), набирающих субъектов из своих практик в рамках одного лечебно-профилактического учреждения или нескольких связанных учреждений. Во всех случаях, когда в статистической модели определение «центр» может трактоваться двояко, в статистическом разделе протокола (см. раздел 5.5.1) необходимо дать точное определение (например, по исследователю, расположению или региону) в отношении рассматриваемого исследования. В большинстве случаев центр определяется наличием исследователя, а в Национальном стандарте Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика» на этот счет представлены необходимые разъяснения. В сомнительных случаях цель правильного определения понятия «центр» заключается в достижении однородности важных факторов, влияющих на измерение первичных переменных и оценку эффектов сравниваемых лекарственных препаратов. В протоколе, по возможности, необходимо заранее описать и обосновать правила объединения центров в один анализ. Как бы то ни было, решение о принятии такого подхода необходимо принимать до снятия ослепления, например, в ходе проверки данных без снятия ослепления.

В протоколе необходимо описать статистическую модель, которая будет использоваться для оценки и тестирования (проверки) эффектов лекарственных препаратов. Вначале основной лекарственный эффект допускается изучать, используя модель, позволяющую учесть различия между центрами, при этом она может не определять взаимодействие факторов лекарственный препарат–центр. Если терапевтический эффект во всех центрах проявляется однородно, то шаблонное включение определения взаимодействия в статистическую модель снижает способность выявлять основные эффекты. При наличии истинной гетерогенности в терапевтических эффектах интерпретация основных из них является неоднозначной.

В некоторых исследованиях, например, крупных исследованиях по изучению смертности с небольшим количеством субъектов на центр, как правило, нет причин ожидать влияния центра на первичные и вторичные переменные, так как возможное влияние, скорее всего, не имеет клинической значимости. В других случаях заранее может быть известно, что ввиду ограниченного количества субъектов в центре включение эффекта «центр» в статистическую модель нецелесообразно. В этих случаях включать категорию «центр» в модель не правильно, а проводить стратификацию по центрам не требуется.

Если положительные терапевтические эффекты отмечаются у значительного количества пациентов в одном центре, как правило, необходимо изучить гетерогенность терапевтических эффектов между центрами, так как она может повлиять на обобщаемость выводов. Существенную гетерогенность можно выявить с помощью графического представления результатов по отдельным центрам или с помощью определенных методов анализа, например, с помощью критерия значимости по взаимодействию лекарственный препарат–центр. При использовании такого критерия значимости необходимо понимать, что он имеет низкую чувствительность в исследовании, направленном на выявление основных эффектов лекарственного препарата.

Если обнаруживается гетерогенность терапевтических эффектов, ее необходимо подробно изучить и осуществить тщательный поиск причин ее возникновения,

зависящих от управления исследованием или характеристик субъектов. Такое объяснение, как правило, представляет собой дальнейший анализ и интерпретацию результатов. В отсутствие объяснения выявленная гетерогенность эффектов лечения, подтверждающаяся, например, выраженными количественными взаимодействиями (см. Глоссарий), свидетельствует о необходимости использования альтернативной оценки терапевтических эффектов, в том числе присвоения различных весов центрам с целью повышения достоверности оценки терапевтических эффектов. Еще более важно понять причины гетерогенности, характеризующейся выраженными количественными взаимодействиями (см. Глоссарий), а неспособность найти ей объяснение может потребовать проведения новых клинических исследований с целью надежного обоснования терапевтических эффектов.

До настоящего момента описывались многоцентровые исследования, основанные на использовании моделей фиксированных эффектов. Для изучения гетерогенности терапевтического эффекта также используют смешанные модели. В таких моделях эффекты центра и лекарственного препарата на центр принимаются за случайные, что особенно значимо при высоком количестве исследовательских центров.

5.3.3. Виды сравнений

5.3.3.1. Исследования превосходства

С научной точки зрения наличие эффективности убедительно подтверждается превосходством исследуемого лекарственного препарата над плацебо в плацебо-контролируемом исследовании, превосходством над активным контролем или наличием взаимосвязи доза-эффект. Эти исследования относятся к исследованиям «превосходства» (см. Глоссарий). В настоящей главе в основном рассматриваются исследования превосходства, если отдельно не указывается иное.

В сложных случаях, когда существует лекарственный препарат, подтвердивший свою эффективность в исследованиях превосходства, проведение плацебо-контролируемых исследований может быть неэффективно. В этих случаях необходимо предусмотреть проведение научно обоснованного исследования с активным контролем. Выбор между плацебо и активным контролем необходимо осуществлять в индивидуальном порядке.

5.3.3.2. Исследования эквивалентности и не меньшей эффективности (безопасности)

В некоторых случаях исследуемый лекарственный препарат сопоставляют с лекарственным препаратом сравнения, не стремясь подтвердить превосходство. Такие исследования, в соответствии с поставленной целью, делятся на две основные категории: первое — исследование «эквивалентности» (см. Глоссарий), второе — исследование не меньшей эффективности (безопасности) (см. Глоссарий).

Исследование биоэквивалентности подпадает под первую категорию. В некоторых случаях исследования клинической эквивалентности проводятся по иным основаниям, как то подтверждение клинической эквивалентности воспроизведенного лекарственного препарата ранее зарегистрированному лекарственному препарату, когда действующее вещество не абсорбируется и не попадает в кровоток.

Многие исследования с активным контролем направлены на подтверждение не меньшей эффективности исследуемого лекарственного препарата активному контролю, таким образом, они попадают во вторую категорию. Другим примером служит исследование, в котором различные дозы исследуемого лекарственного препарата сравнивают с рекомендуемой или множеством доз лекарственного препарата сравнения. Цель такого дизайна — единовременное подтверждение взаимосвязи доза-эффект исследуемого лекарственного препарата и сравнение его с активным контролем.

Исследования эквивалентности и не меньшей эффективности с активным контролем могут включать использование плацебо, что объединяет несколько целей в одном исследовании. Например, с помощью такого исследования можно подтвердить превосходство над плацебо и таким образом валидировать дизайн и одновременно оценить, насколько эффективность и безопасность исследуемого лекарственного препарата соответствует таковым активного контроля. Недостатки исследований эквивалентности (и не меньшей эффективности) с активным контролем, в которых не используется плацебо или не рассматривается взаимосвязь доза–эффект, хорошо известны. Они обусловлены характерным для таких исследований отсутствием внутренней валидности (в отличие от исследований превосходства), что требует проведения внешней валидации. Исследование эквивалентности (и не меньшей эффективности) не является консервативным по своей природе, поэтому многие недочеты при планировании и проведении исследования способствуют возникновению систематических ошибок, которые склоняют выводы в сторону наличия эквивалентности. В связи с этим к особенностям дизайна таких исследований необходимо предъявлять повышенное внимание, а их проведение следует осуществлять с особой тщательностью. Например, особо важно минимизировать частоту несоблюдения критериев отбора, отсутствия приверженности, исключений, невозможности последующего наблюдения, недостающих данных и других отклонений от протокола, а также максимально ограничить их влияние на последующий анализ.

Необходимо тщательно подбирать активный контроль. Примером подходящего активного контроля может служить широко применяемый метод лечения, чья эффективность по рассматриваемому показанию была надежно подтверждена и измерена количественно в хорошо спланированных и надлежащим образом документированных исследованиях превосходства, от которого можно с надежностью ожидать проявления сходной эффективности в намеченном исследовании с активным контролем. В связи с этим планируемое исследование должно иметь те же ключевые элементы дизайна (первичные переменные, доза активного контроля, критерии отбора и т. д.), что и ранее проведенные исследования превосходства, в которых активный контроль достоверно проявил клинически значимую эффективность, принимая во внимание достижения в медицине и статистике, значимые для нового исследования.

Крайне важно, чтобы в протоколе исследования, направленного на подтверждение эквивалентности или не меньшей эффективности, содержалось четкое указание, что это его основная цель. В протоколе необходимо представить границу эквивалентности. Эта граница — есть наибольшее различие, которое считается клинически допустимым; она должна быть меньше, чем различия, отмечавшиеся в исследованиях превосходства с активным контролем. Для исследования эквивалентности с активным контролем необходимы как верхняя, так и нижняя границы, тогда как для исследования не меньшей эффективности — только нижняя. Выбор границы эквивалентности необходимо обосновать с клинических позиций.

Статистический анализ, как правило, основывается на использовании доверительных интервалов (см. раздел 5.5.5). Для исследований эквивалентности необходимо использовать двусторонний доверительный интервал. Эквивалентность считается доказанной, если доверительный интервал полностью уместился в границы эквивалентности (не вышел за границы эквивалентности). С технической точки зрения, это эквивалентно одновременному использованию двух односторонних тестов для проверки (составной) нулевой гипотезы о том, что различия между группами сравнения выходят за границы эквивалентности, против (составной) альтернативной гипотезы о том, что различия между группами находятся в пределах указанных границ. Ввиду того что две нулевые гипотезы независимы, ошибка I рода находится под контролем. Для исследований не меньшей эффективности используют односторонний доверительный интервал. Метод доверительных интервалов имеет альтер-

нативу — одностороннюю проверку гипотезы для тестирования нулевой гипотезы о том, что различия между лекарственными препаратами (исследуемый лекарственный препарат минус контроль) не больше нижней границы эквивалентности против альтернативы — различия между лекарственными препаратами больше, чем нижняя граница эквивалентности. Выбор ошибки I рода не должен зависеть от того, одно- или двусторонний тест будет применяться в анализе. На этих методах основан расчет размера выборки (см. раздел 5.3.4).

Заключение о подтверждении эквивалентности или не меньшей эффективности, основанное на незначимых результатах тестирования нулевой гипотезы о том, что разница между исследуемым лекарственным препаратом и активным контролем отсутствует, недопустимо.

Имеется ряд особенностей при выборе совокупностей, подлежащих анализу. Выбывшие или исключенные субъекты из исследуемой или контрольной группы будут склонны к отсутствию ответа на лечение, в связи с чем результаты анализа полной совокупности, подлежащей анализу (см. Глоссарий), могут вводить систематическую ошибку в пользу подтверждения эквивалентности (см. раздел 5.5.2.3).

5.3.3.3. Исследования взаимосвязи доза–эффект

Ответ на вопрос о том, как на эффекты исследуемого лекарственного препарата влияет его доза, можно получить на всех этапах разработки с помощью различных подходов (см. методические рекомендации по проведению клинических исследований с целью подбора дозы лекарственных препаратов). Исследования доза–эффект могут преследовать различные цели: подтверждение эффективности, изучение формы и расположения (location) кривой доза–эффект, определение оптимальной начальной дозы, установление оптимального способа подбора дозы у отдельного пациента, выявление максимальной дозы, превышение которой не приносит дополнительной пользы. Эти цели необходимо рассматривать, используя полученные данные при назначении исследуемых доз, включая при необходимости плацебо (нулевая доза). Использование процедур оценки взаимосвязи между дозой и ответом, включая построение доверительных интервалов и использование графических методов, для этих целей так же важно, как и статистических тестов. Может понадобиться приспособить используемые методы тестирования гипотезы к естественному порядку назначения доз или к задачам определения формы кривой доза–эффект (например, ее монотонность). В протоколе необходимо представить подробное описание планируемых статистических процедур.

5.3.3.4. Групповой последовательный дизайн (Group Sequential Designs)

Групповые последовательные дизайны используют для облегчения проведения промежуточного анализа (см. раздел 5.4.5 и Глоссарий). Несмотря на то что групповые последовательные дизайны не единственный приемлемый вид дизайна, позволяющего проведение промежуточного анализа, их используют наиболее часто ввиду большей практичности оценки групповых исходов через определенные интервалы в течение исследования, чем непрерывные данные от каждого становящегося доступным субъекта. Необходимо полностью описать планируемые статистические процедуры до получения данных об исходах лечения и распределении субъектов по сравниваемым группам (то есть до снятия ослепления, см. раздел 5.4.6). Указанный дизайн наиболее часто и с высокой результативностью используется в крупных длительных исследованиях смертности или основных неконечных точек летальности, однако его значимость в других областях также повышается. В частности, общепризнано, что во всех исследованиях следует оценивать безопасность, поэтому необходимо иметь формальные процедуры, позволяющие досрочно прекратить исследование по соображениям безопасности.

5.3.4. Размер выборки

Количество субъектов, участвующих в клиническом исследовании, всегда должно быть достаточно большим, чтобы правильно ответить на поставленные вопросы. Это количество, как правило, определяют на основании основной цели исследования. Если расчет размера выборки осуществляется с помощью других методов, на это необходимо открыто указать и обосновать его. Например, исследование, размер выборки которого рассчитывается с целью изучения безопасности или важных вторичных целей, может потребовать большего количества субъектов, чем исследование, размер выборки которого определяется первичной переменной эффективности (см., в частности, ICH E1).

При использовании стандартного метода расчета размера выборки необходимо учитывать следующие параметры: первичная переменная, статистический тест, нулевая гипотеза, альтернативная («рабочая») гипотеза для выбранных доз (учитывающая подлежащее выявлению или отклонению терапевтическое различие при определенной дозе у определенной группы пациентов), вероятность ошибочного отклонения нулевой гипотезы (ошибка I рода) и вероятность ошибочно принять нулевую гипотезу (ошибка II рода), а также способы обращения с исключенными данными и нарушениями протокола. В некоторых случаях частота явления служит основой для расчета мощности, для этого задают определенное количество явлений и на основании этого вычисляют размер выборки.

В протоколе необходимо представить способ расчета размера выборки, а также все используемые в расчетах переменные (как то дисперсии, средние, частота ответов, частота явлений, искомые различия) с обоснованием их значений. Необходимо определить чувствительность размера выборки к различным отклонениям от выдвинутых допущений, что можно осуществить, представив набор различных размеров выборок, соответствующих обоснованному ряду отклонений от указанных допущений. В подтверждающих исследованиях такие допущения должны основываться на опубликованных данных или результатах ранее проведенных исследований. Искомое терапевтическое различие может основываться на предположении о минимальном эффекте, имеющем клиническую значимость при ведении пациентов или суждении об ожидаемом эффекте нового лекарственного препарата, если он больше первого. Традиционно вероятность ошибки I рода принимается за 5% и менее или в соответствии поправками, необходимыми для учета множественности измерений. На точный выбор ее величины влияют предварительная вероятность правильности рассматриваемой гипотезы и желаемого влияния результатов. Вероятность ошибки II рода обычно принимается за 10 или 20%, в интересах спонсора — придерживаться наименьшей возможной величины, особенно в случае трудности или невозможности повторного проведения исследования. В некоторых случаях приемлемы и даже предпочтительны альтернативные значения общепринятым уровням ошибок I и II рода.

Расчет размера выборки осуществляют в целях основного анализа. Если это «полная совокупность, подлежащая анализу», то значения величины эффекта могут потребовать уменьшения по сравнению с совокупностью по протоколу (см. Глоссарий). Это осуществляется для уменьшения влияния терапевтического эффекта, возникающего вследствие включения данных от исключенных или слабо приверженных лечению пациентов. Допущения о вариации также могут потребовать пересмотра.

Размер выборки исследования эквивалентности или не меньшей эффективности (см. раздел 5.3.3.2), как правило, основывается на необходимости получения достоверного интервала для разницы терапевтического эффекта, подтверждающего, что лекарственные препараты отличаются не более чем на клинически приемлемую величину. Если мощность исследования эквивалентности рассчитывается для различий, равных нулю (отсутствие различий), то размер выборки, необходимый для достижения указанной мощности, будет занижен, если истинные различия отличны

от нуля. Если мощность исследования не меньшей эффективности рассчитывается для нулевых различий, то размер выборки, необходимый для достижения заданной мощности будет занижен, если эффект исследуемого лекарственного препарата будет меньше, чем у активного контроля. Выбор «клинически приемлемого» различия требует обоснования с позиций интересов потенциальных пациентов, оно может быть меньше описанного выше «клинически значимого» различия, изучаемого в исследованиях превосходства, направленных на выявление такого различия.

Заранее рассчитать точный размер выборки в групповом последовательном исследовании невозможно ввиду влияния случая, а также принятых критериев остановки исследования и истинного различия между сравниваемыми лекарственными препаратами. При составлении правил остановки необходимо учитывать последующее распределение количества пациентов (выборки), заключающееся, как правило, в ожидаемом и максимальном размере выборки.

Если частота явлений ниже ожидаемой или ее вариабельность выше предусмотренной протоколом, существуют методы перерасчета размера выборки без снятия ослепления или сравнения лекарственных препаратов (см. раздел 5.4.4).

5.3.5. Сбор и обработка данных

Сбор данных и передача их от исследователя спонсору может осуществляться различными способами, включая бумажные индивидуальные регистрационные карты, удаленные системы мониторинга, компьютерные медицинские системы и электронную передачу данных. Какой бы метод сбора данных не использовался, форма и содержание полученных сведений должны полностью соответствовать протоколу, требования к нему необходимо оговорить до начала проведения исследования. Следует сосредоточиться на данных, необходимых для проведения предусмотренного анализа, включая фоновые обстоятельства (как то сроки осуществления оценки в зависимости от приема лекарственного препарата), требуемые для подтверждения соблюдения протокола и выявления важных отклонений от него. Необходимо отличать «недостающие значения» от «нулевых» и «отсутствия, предусмотренного протоколом» (characteristic absent).

Процесс сбора данных до закрытия базы данных необходимо проводить в соответствии с правилами Надлежащей клинической практики (см. раздел 5 Национального стандарта Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика»). В частности необходимо предусмотреть своевременные и надежные процессы регистрации данных, исправления ошибок и упущений, чтобы обеспечить высокое качество заполнения базы данных и достичь целей исследования за счет проведения предусмотренного анализа.

5.4. ВОПРОСЫ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЯ

5.4.1. Мониторинг исследования и промежуточный анализ

Качественное проведение исследования согласно протоколу оказывает ключевое влияние на достоверность результатов (см. Национальный стандарт Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика»). Тщательный мониторинг обеспечивает раннее выявление ошибок и снижает вероятность их повторного возникновения.

Выделяют два разных вида мониторинга, характерных для подтверждающих клинических исследований, спонсируемых фармацевтической отраслью. Первый вид заключается в надзоре за качеством исследования, тогда как второй включает снятие ослепления с целью сравнения лекарственных препаратов (то есть промежуточный анализ). Указанные виды мониторинга исследования в дополнение к различным обязанностям персонала имеют различный уровень доступа к документам и данным,

поэтому к ним применяют различные принципы контроля потенциальных статистических и операционных систематических ошибок.

С целью надзора за качеством исследования проверки, осуществляемые в ходе мониторинга, могут включать проверку соблюдения протокола, приемлемости собираемых данных, успешности целевых показателей набора пациентов, правильности элементов дизайна, способности удерживать пациентов в исследовании и т. д. (см. разделы 5.4.2–5.4.4). Этот вид мониторинга не требует доступа к данным сравнительных терапевтических эффектов и снятия ослепления, таким образом, не влияя на ошибку I рода. Указанный мониторинг является обязанностью спонсора (см. Национальный стандарт Российской Федерации ГОСТ Р 52379-2005 «Надлежащая клиническая практика») и может проводиться им самостоятельно или с привлечением независимых лиц. Такой мониторинг начинается непосредственно после выбора исследовательских центров и заканчивается сбором и выверкой данных последнего субъекта.

Мониторинг, осуществляемый в целях промежуточного анализа, включает рассмотрение результатов сравнения лекарственных препаратов. Промежуточный анализ требует открытого (то есть неослепленного) доступа к распределению субъектов по сравниваемым группам и сравнительное резюме по сравниваемым группам. Для этого в протоколе (или в соответствующей поправке, введенной до первого анализа) должен содержаться статистический план промежуточного анализа, направленный на предотвращение возникновения некоторых видов систематических ошибок, что обсуждается в разделах 5.4.5 и 5.4.6.

5.4.2. Изменения критериев включения и невключения

На протяжении всего периода набора субъектов критерии включения и невключения должны оставаться неизменными и соответствовать протоколу. Изредка допускаются исключения, например, для длительных исследований, когда накопление медицинских данных, получаемых вне исследования или по результатам промежуточного анализа, может потребовать изменения указанных критериев. Изменения могут быть внесены, если мониторами выявляются регулярные нарушения критериев отбора или ввиду их строгости отмечается низкая скорость набора субъектов. Изменения необходимо осуществлять без снятия ослепления, их необходимо описать в поправке к протоколу, в которой следует представить все статистические последствия, как то изменение размера выборки, обусловленное различной частотой возникновения явлений, или коррекция запланированного анализа, например, разделение анализа в соответствии с модифицированными критериями включения и невключения.

5.4.3. Скорость набора субъектов

В исследованиях с длительным набором субъектов необходимо осуществлять мониторинг скорости такого набора, и если она опускается значительно ниже запланированной, необходимо выявить причины происходящего и принять шаги к их устранению, чтобы сохранить мощность исследования и снизить озабоченность в отношении избирательного включения пациентов и других аспектов качества исследования. В многоцентровых исследованиях эти факторы необходимо учесть для отдельных центров.

5.4.4. Изменение (коррекция) размера выборки

В длительных исследованиях зачастую появляется возможность проверить предпосылки, на которых основывались выбор исходного дизайна и расчет размера выборки. Это особенно важно, если дизайн исследования был составлен на предварительных и (или) неопределенных допущениях. Промежуточный анализ, осуществ-

вленный на ослепленных данных, может выявить, что общие переменные ответа, частота явления или выживаемость не соответствуют ожидаемым. Допускается на основании исправленных допущений произвести перерасчет размера выборки, который необходимо обосновать и представить в поправке к протоколу и отчете о клиническом исследовании. Необходимо обосновать меры, принятые для сохранения ослепления, и описать влияние, если таковое обнаружено, на ошибку I рода и ширину доверительных интервалов. По возможности, в протоколе необходимо оговорить потенциальную необходимость перерасчета размера выборки (см. раздел 5.3.4).

5.4.5. Промежуточный анализ и досрочное прекращение

Промежуточный анализ — это любой анализ, направленный на сравнение исследуемых групп в отношении показателей эффективности или безопасности, осуществляемый до момента официального завершения исследования. Ввиду того, что количество, методы и последствия таких сравнений влияют на интерпретацию результатов исследования, все промежуточные анализы необходимо заранее тщательно спланировать и описать в протоколе. Ряд обстоятельств может потребовать проведения не предусмотренного протоколом промежуточного анализа. В этом случае, до снятия ослепления с целью анализа данных, необходимо составить поправку к протоколу, описывающую промежуточный анализ. Промежуточный анализ, проводимый с целью решения вопроса о продолжении или прекращении исследования, реализуется, как правило, с помощью группового последовательного дизайна, в котором используются соответствующие статистические схемы мониторинга (см. раздел 5.3.3.4). Целью такого промежуточного анализа является раннее прекращение исследования, если превосходство лекарственного препарата достоверно подтверждается; если обнаруживается, что значимое терапевтическое различие маловероятно или неприемлемо; если возникают недопустимые нежелательные явления. В целом, для досрочного прекращения исследования критерии подтверждения эффективности требуют большего обоснования (то есть они более консервативны), чем критерии безопасности. Если дизайн исследования и цель мониторинга основываются на множестве конечных точек, то эту множественность также необходимо принимать во внимание.

В протоколе необходимо описать схему промежуточного анализа или, как минимум, принципы его проведения, например, будет ли использоваться подход, основанный на гибкой функции зависимости вероятности ошибки I рода от полученного объема данных. Более подробные сведения допускается представить в поправке к протоколу непосредственно перед первым промежуточным анализом. В протоколе или поправках необходимо четко описать критерии остановки и их особенности. Необходимо рассмотреть возможное влияние ранней остановки на анализ прочих важных переменных. Это должен осуществить или утвердить Комитет по мониторингу данных (см. раздел 5.4.6), если таковой предусмотрен. Отклонения от запланированных процедур способны исказить результаты исследования. Если изменения в исследовании необходимы, то последующую коррекцию статистических процедур необходимо как можно раньше описать в поправке к протоколу, особенно в части обсуждения влияния на анализ или выводы, которые такие изменения могут привести. Выбранные процедуры должны всегда гарантировать, что полная (совокупная) вероятность ошибки I рода находится под контролем.

Осуществление промежуточного анализа должно носить конфиденциальный характер, так как открывается доступ к неослепленным данным и результатам. Ни одно лицо, участвующее в проведении исследования, не должно знать о результатах таких анализов, так как это может повлиять на отношение к исследованию и вызвать изменение характеристик вновь включаемых пациентов или ввести систематические ошибки при сравнении лекарственных препаратов. Этот принцип справедлив в от-

ношении всего исследовательского персонала и представителей спонсора, за исключением лиц, выполняющих промежуточный анализ. Исследователям необходимо сообщать лишь о решении прекратить или продолжить исследование или осуществить изменение его процедур.

Большинство клинических исследований, направленных на подтверждение эффективности и безопасности исследуемого лекарственного препарата, должны продолжаться до включения всего запланированного количества субъектов. Досрочное прекращение исследования допускается только по этическим соображениям или при обнаружении, что мощность исследования неприемлема. Однако общеизвестно, что в ходе разработки лекарственного препарата спонсору по различным причинам может понадобиться доступ к данным исследования, например, с целью планирования последующих. Также признается, что только часть исследований включает изучение тяжелых, угрожающих жизни исходов или смертности, что может потребовать последовательного мониторинга накапливающихся сравнительных данных по этическим соображениям. При любой из этих ситуаций, до снятия ослепления со сравнительных данных, в протоколе или поправках к нему необходимо описать план промежуточного статистического анализа, чтобы предупредить возникновение возможных статистических или операционных систематических ошибок.

Для многих клинических исследований новых лекарственных препаратов, особенно оказывающих значимое влияние на общественное здоровье, обязанность сравнения исходов эффективности и (или) безопасности в ходе мониторинга должна возлагаться на внешнюю независимую группу, часто называемую Независимый комитет по мониторингу данных, Комиссия по мониторингу данных и безопасности или Комитет по мониторингу данных, чьи обязанности должны быть хорошо описаны.

Если спонсор осуществляет мониторинг сравнения эффективности или безопасности и, таким образом, имеет доступ к неослепленным сравнительным данным, необходимо принять особые меры осторожности, чтобы защитить целостность исследования и надлежащим образом использовать и ограничивать полученные сведения. Спонсор обязан обеспечить и документально подтвердить, что внутренний комитет по мониторингу действовал в соответствии с письменными стандартными операционными процедурами и вел протокол заседаний по принятию решений, включая записи о промежуточных результатах.

Любой плохо спланированный промежуточный анализ (с последующим досрочным прекращением исследования или без такового) может исказить результаты исследования и ослабить надежность полученных выводов, поэтому таких анализов необходимо избегать. Если проведен незапланированный промежуточный анализ, в отчете о клиническом исследовании следует обосновать его необходимость, степень снятия ослепления, представить оценку потенциальной величины возникших систематических ошибок и влияние, которое они оказали на интерпретацию результатов.

5.4.6. Роль Независимого комитета по мониторингу данных (НКМД) (см. разделы 1.25 и 5.52 ГОСТ)

Спонсор вправе учредить НКМД с целью регулярной оценки течения клинического исследования, данных по безопасности и ключевых переменных эффективности и составления рекомендаций спонсору о продолжении, модификации или прекращении исследования. Для НКМД необходимо составить письменную операционную процедуру и вести записи всех его встреч, включая промежуточные результаты. По завершении исследования эти записи должны быть доступны для ознакомления. Независимость НКМД направлена на контроль над раскрытием важных сравнительных данных и защиту целостности клинического исследования от нежелательного влияния, обусловленного доступом к информации об исследовании. НКМД не свя-

зан с Этическим советом организации (ЭСО) и Независимым этическим комитетом (НЭК), в его состав должны входить специалисты в отдельных областях клинических исследований, включая статистиков.

Если в НКМД входят представители спонсора, их роль необходимо четко описать в операционной процедуре комитета (например, указание на право голоса при рассмотрении ключевых вопросов). Ввиду того что представители спонсора получают доступ к неослепленным данным, в процедуре также необходимо описать способ контроля над распространением промежуточных результатов исследования среди персонала спонсора.

5.5. ВОПРОСЫ АНАЛИЗА ДАННЫХ

5.5.1. Предварительное описание анализа

При планировании клинического исследования ключевые аспекты последующего статистического анализа необходимо описать в статистическом разделе протокола. В нем необходимо представить все основные элементы предполагаемого подтверждающего анализа первичных переменных и пути преодоления возможных аналитических затруднений. Для поисковых исследований допускается ограничиться более общими принципами и указаниями.

План статистического анализа (см. Глоссарий) может быть представлен в виде отдельного от протокола документа. В него включаются более подробные технические детали основных принципов, отраженных в протоколе (см. раздел 5.7.1). План может включать подробное описание процедур выполнения статистического анализа первичных и вторичных переменных и других данных. План необходимо проверить, возможно, обновить по результатам проверки данных без снятия ослепления (определение представлено в разделе 5.7.1) и окончательно оформить до снятия ослепления. Необходимо сохранить документы, подтверждающие дату завершения плана статистического анализа, а также дату последующего снятия ослепления.

Если по результатам проверки данных без снятия ослепления возникает необходимость в коррекции ключевых аспектов, обозначенных в протоколе, ее осуществляют путем представления поправки к протоколу. В иных случаях достаточно обновить план статистического анализа в соответствии с итогами проверки данных без снятия ослепления. Подтверждающими результатами признаются лишь такие, которые получены на основании предусмотренных протоколом (включая поправки) анализов.

В статистическом разделе отчета о клиническом исследовании необходимо подробно описать статистическую методологию, включая этапы принятия решений, затрагивающих методологию проведения клинического исследования (см. главу 6 настоящего тома «Структура и содержание отчетов о клинических исследованиях»).

5.5.2. Совокупность, подлежащая анализу

В статистическом разделе протокола необходимо обозначить совокупность субъектов, чьи данные будут включены в основной анализ. В дополнение к этому рекомендуется представить документацию по всем субъектам, в отношении которых были использованы исследовательские процедуры (например, подготовительный период). Содержание документации о субъекте зависит от особенностей отдельного исследования, тем не менее, по возможности во всех случаях необходимо обладать как минимум демографическими данными и исходными данными о заболевании.

Если все включенные в клиническое исследование субъекты удовлетворяют всем критериям отбора, успешно прошли все процедуры исследования с полноценным последующим наблюдением, по ним имеется полный набор данных, то совокупность

субъектов, подлежащих включению в анализ, очевидна. Дизайн и его реализация должны способствовать максимальному соответствию указанному идеалу, однако в действительности полное достижение такой цели маловероятно. В связи с этим в статистическом разделе протокола необходимо заранее отразить способы решения ожидаемых затруднений с позиций их влияния на субъектов и данные, подлежащие анализу. В протоколе также необходимо предусмотреть процедуры, направленные на недопущение ожидаемых затруднений в проведении исследования, способных помешать правильному анализу, включая различные виды нарушений протокола, исключения из исследования и недостающие данные. В протоколе необходимо описать и способы снижения частоты таких затруднений, и методы решения возникших во время анализа данных проблем. По результатам проверки данных без снятия ослепления следует рассмотреть необходимость внесения поправок с целью коррекции анализа в связи с нарушениями протокола. Рекомендуется установить все важные нарушения протокола с учетом времени их возникновения, причин и влияния на результаты исследования. В отчете о клиническом исследовании необходимо указать частоту и вид нарушений протокола, недостающие данные и другие проблемы, возникшие в ходе исследования, а также описать их потенциальное влияние на результаты исследования (см. методические рекомендации по структуре и содержанию отчетов о клинических исследованиях).

Решение о выборе совокупности, подлежащей анализу, принимается на основании следующих принципов: (1) минимизация систематических ошибок, (2) недопущение увеличения ошибки I рода.

5.5.2.1. Полная совокупность, подлежащая анализу

Принцип «по намерению лечить» (intention to treat, см. Глоссарий) — это принцип, согласно которому в первичный анализ включают всех рандомизированных пациентов. Приверженность указанному принципу потребует наблюдения за всеми рандомизированными субъектами на предмет исследуемых исходов. В действительности этот идеал по нижеописанным причинам практически недостижим. В настоящем документе понятие «полная совокупность, подлежащая анализу» используется для описания совокупности, подлежащей анализу, которая максимально полно соответствует совокупности всех рандомизированных субъектов. Поддержание начальной рандомизации в анализе необходимо для предотвращения систематических ошибок и обеспечения базы для статистических тестов. Во многих клинических исследованиях использование полной совокупности, подлежащей анализу, обеспечивает консервативный подход. Зачастую это позволяет получить оценки терапевтических эффектов, которые лучше отражают реальную клиническую практику.

Причины исключения рандомизированных субъектов из полной совокупности, подлежащей анализу, ограничены и включают: несоответствие основным критериям отбора (нарушение отбора субъектов), субъектом не принята ни одна доза исследуемых лекарственных препаратов (плацебо), отсутствуют какие-либо данные после рандомизации. Во всех случаях указанные исключения необходимо объяснить. Субъекты, не удовлетворяющие критериям отбора, могут быть исключены из анализа без последующего возникновения систематических ошибок только при следующих обстоятельствах:

- критерий включения был определен до рандомизации;
- выявление соответствующего нарушения отбора полностью объективно;
- проверку соблюдения критериев отбора у всех субъектов выполняют одинаково (это тяжело обеспечивается в открытых исследованиях, а также в двойных слепых, если снятие ослепления произошло до проверки, что указывает на важность проверки данных без снятия ослепления);
- все выявленные нарушения определенного критерия отбора устранены.

В некоторых случаях уместно исключить из совокупности всех тех рандомизированных субъектов, кто не принял ни одну дозу лекарственного препарата. Несмотря на исключение таких субъектов, принцип по намерению лечить в этом случае не нарушится, если, к примеру, на решение о начале лечения не может повлиять осведомленность о распределении субъектов по группам. В других ситуациях может возникнуть необходимость исключения из совокупности всех тех рандомизированных субъектов, у которых отсутствуют какие-либо данные после рандомизации. Анализ считается неполным, если не учтены систематические ошибки, возникающие в связи с указанными обстоятельствами.

Если используется полная совокупность субъектов, подлежащих анализу, нарушения протокола, возникшие после рандомизации, могут оказать влияние на качество данных и выводы, особенно если возникновение связано с распределением субъектов по группам. В большинстве случаев включение данных таких субъектов в анализ допустимо и соответствует принципу по намерению лечить. Особые затруднения возникают в отношении субъектов, исключенных из исследования после получения одной или более доз лекарственного препарата, данные которых после исключения отсутствуют, а также других субъектов, последующее наблюдение которых невозможно по иным причинам, так как невозможность включить указанных субъектов в полную совокупность, подлежащую анализу, может резко подорвать указанный подход. В этой связи важным является измерение первичных переменных, произведенное при невозможности по каким-либо причинам последующего наблюдения за субъектом, или последующий сбор данных в соответствии с намеченной протоколом схемой оценки. Последующий сбор данных особенно важен для исследований, в которых первичной переменной является смертность или тяжелая инвалидность. В протоколе необходимо описать намерение осуществлять сбор данных таким способом. С целью замещения недостающих данных допускается применять вычислительные методы: от использования результатов последнего наблюдения до сложных математических моделей. Другие методы, направленные на обеспечение доступности значений первичных переменных каждого субъекта в полной совокупности, подлежащей анализу, могут потребовать определенных допущений об исходах субъектов или упрощенного выбора между исходами (например, успех/неуспех). В статистическом разделе протокола необходимо описать и обосновать использование любой из указанных стратегий и разъяснить допущения, лежащие в основе любых запланированных математических моделей. Необходимо подтвердить надежность соответствующих результатов анализа, особенно если рассматриваемая стратегия сама способна вводить систематические ошибки в оценку терапевтических эффектов.

Ввиду непредсказуемости ряда проблем в некоторых случаях предпочтительно отложить детальное описание способов обращения с затруднениями до проверки данных без снятия ослепления, проводимой в конце исследований. Этот подход необходимо описать в протоколе.

5.5.2.2. Совокупность «по протоколу»

Совокупность субъектов «по протоколу» (per protocol, «соответствующие протоколу», «надлежащие единицы», выборка «эффективности», выборка «субъектов, которых можно оценить») — подмножество субъектов полной совокупности, подлежащей анализу, которые были значимо близки к требованиям протокола; ей присущи следующие свойства:

соблюдение заранее оговоренного режима дозирования с применением минимальных доз;

доступность значений первичных переменных;

отсутствие каких-либо значительных нарушений протокола, включая нарушение критериев отбора.

Непосредственные причины исключения субъектов из совокупности по протоколу необходимо полностью описать и документировать до снятия ослепления в соответствии с особенностями отдельного клинического исследования.

Использование совокупности по протоколу может увеличить вероятность подтверждения эффективности нового лекарственного препарата в анализе, она также наиболее точно отражает научную модель, заложенную в протокол. Однако в зависимости от исследования соответствующее тестирование гипотезы и оценка терапевтического эффекта могут оказаться неконсервативными. Систематические ошибки, которые могут оказаться достаточно большими, возникают в силу того, что приверженность протоколу исследования может быть обусловлена лекарственным препаратом и исходом.

Необходимо полностью выявить и обобщить все затруднения, приведшие к исключению субъектов с целью формирования совокупности по протоколу и другим нарушениям протокола. К таким нарушениям, в частности, относятся ошибки при распределении пациентов по группам, применение неразрешенных методов лечения, низкая приверженность, невозможность последующего наблюдения и недостающие данные. Надлежащей практикой является выявление закономерности возникновения таких проблем по частоте и времени их возникновения в зависимости от группы лечения.

5.5.2.3. Значение различных совокупностей, подлежащих анализу

В целом, рекомендуется подтвердить отсутствие чувствительности основных результатов исследования к выбору совокупности субъектов, подлежащих анализу. В подтверждающих исследованиях рекомендуется провести анализ как полной совокупности, так и совокупности по протоколу; выявленные различия между совокупностями подлежат открытому обсуждению и интерпретации. В некоторых случаях целесообразно предпринять более глубокое изучение чувствительности выводов к выбору совокупности субъектов, подлежащих анализу. Если анализ полной совокупности и совокупности по протоколу приводит к одинаковым, в сущности, выводам, то надежность результатов исследования повышается, учитывая при этом, что необходимость исключения значительной доли субъектов из анализа по протоколу подрывает общую достоверность результатов исследования.

Полная совокупность, подлежащая анализу, и совокупность по протоколу играют различную роль в исследованиях превосходства (цель которых заключается в подтверждении превосходства исследуемого лекарственного препарата) с одной стороны и в исследованиях эквивалентности и не меньшей эффективности (цель которых заключается в подтверждении сопоставимости исследуемого лекарственного препарата с контролем) с другой. В исследованиях превосходства полная совокупность, подлежащая анализу, используется (за редким исключением) в основном анализе, так как это позволяет избежать чрезмерно оптимистичных оценок эффективности, возникающих при осуществлении анализа по протоколу, так как включение субъектов с низкой приверженностью в полную совокупность, подлежащую анализу, в целом уменьшает оценку терапевтического эффекта. Однако в исследованиях эквивалентности и не меньшей эффективности использование полной совокупности, подлежащей оценке, в целом, не является консервативным подходом, поэтому ее использование необходимо подробно обосновать.

5.5.3. Недостающие данные и выбросы

Недостающие данные в клиническом исследовании являются потенциальным источником систематических ошибок. В связи с этим необходимо принять всевозможные меры, чтобы соблюсти все требования протокола относительно сбора и обращения с данными. Однако в действительности почти всегда приходится сталкивать-

ся с недостающими данными. Тем не менее, результаты исследования признаются достоверными, если применяются рациональные методы обращения с недостающими данными, в особенности, если они заранее описаны в протоколе. Указанные методы могут уточняться при коррекции плана статистического анализа по результатам проверки данных без снятия ослепления. К сожалению, универсальных методов обращения с недостающими данными не существует. Необходимо изучить чувствительность результатов анализа к методу обработки недостающих данных, особенно если объем таких данных значительный.

Аналогичный подход необходимо принять при определении влияния выбросов, статистическое определение которых, в некоторой степени, условно. Однозначное признание некоторого значения в качестве выброса наиболее убедительно, если оно объясняется и с медицинских, и со статистических позиций, тогда, руководствуясь медицинскими обстоятельствами, предпринимаются необходимые действия. Никакая из процедур протокола и плана статистического анализа, направленная на обращение с выбросами, не должна а priori благоприятствовать какому-либо сравниваемому лекарственному препарату. Если в протоколе исследования процедура обращения с выбросами не предусмотрена, необходимо провести анализ со всеми полученными значениями и не менее одного анализа после снижения или исключения влияния выбросов; полученные различия подлежат обсуждению.

5.5.4. Трансформация данных

Решение о трансформации ключевых переменных до анализа необходимо принять на этапе планирования исследования на основании аналогичных данных, полученных по результатам ранее проведенных клинических исследований. В протоколе необходимо описать вид трансформации (например, извлечение квадратного корня, логарифмирование) и объяснить его выбор, особенно в отношении первичных переменных. Общие принципы использования трансформации с целью обеспечения допущений, лежащих в основе статистических методов, представлены в соответствующих статистических руководствах. В ряде отдельных клинических исследований разработаны стандартные подходы к определенным переменным. При решении о необходимости и способе трансформации переменных необходимо учитывать предпочтительную шкалу, способствующую клинической интерпретации результатов.

Те же рассуждения справедливы для других производных переменных, как то изменение по сравнению с исходным, доля изменения по сравнению с исходным, «площадь под кривой» повторных измерений, отношение двух различных переменных. Необходимо подробно объяснить последующую клиническую интерпретацию результатов, в протоколе необходимо обосновать их получение. В разделе 6.2.2.2 описаны вопросы, близкие к обсуждаемой тематике.

5.5.5. Оценка, доверительные интервалы и тестирование (проверка) гипотезы

В статистическом разделе протокола необходимо представить тестируемые гипотезы и (или) терапевтические эффекты, подлежащие оценке для достижения первичных целей исследования. Для первичных (и желательно вторичных) переменных необходимо описать статистические методы, направленные на выполнение указанных задач, и охарактеризовать статистическую модель, лежащую в ее основе. Для значений терапевтических эффектов необходимо, по возможности, указать доверительные интервалы и способы их вычисления. Необходимо оговорить намерение использовать исходные данные для улучшения точности или поправки оценок на потенциальные различия в исходных характеристиках, например, с помощью ковариационного анализа.

Необходимо указать: одно- или двусторонние тесты будут использоваться для установления статистической значимости и заранее обосновать использование односторонних тестов. Если тестирование гипотез считается ненадлежащим, необходимо представить альтернативные методы получения статистических выводов. Проблема использования одно- и двусторонних подходов для выявления взаимосвязи противоречива, в статистической литературе представлены различные мнения на сей счет. С регуляторной точки зрения величину ошибки I рода для односторонних тестов рекомендуется принимать как половину величины ошибки I рода для двусторонних тестов. Это повышает сопоставимость с двусторонними доверительными интервалами, которые, в целом, подходят для оценки возможной величины различий между двумя лекарственными препаратами.

Выбранная статистическая модель должна отражать текущие медицинские и статистические взгляды на анализируемые переменные и статистический план исследования. Необходимо всесторонне описать все эффекты, подлежащие включению в анализ (например, дисперсионный анализ), а также объяснить способ модификации совокупности эффектов (если он имеет место) в ответ на предварительные результаты. Эти требования справедливы и для совокупности ковариат, включаемых в ковариационный анализ (см. также раздел 5.5.7). При выборе статистических моделей необходимо уделить должное внимание статистическому распределению как первичных, так и вторичных переменных. При осуществлении выбора (например, между параметрическими и непараметрическими моделями) следует помнить о необходимости представления величины статистических оценок терапевтических эффектов вместе с доверительными интервалами (в дополнение к тестированию критерия значимости).

Необходимо четко разграничить основной анализ первичной переменной от вспомогательных анализов первичных и вторичных переменных. В статистическом разделе протокола или плане статистического анализа необходимо также изложить способ обобщения и предоставления отчетности по данным, не относящимся к первичным и вторичным переменным. Для соблюдения преемственности между различными исследованиями, например, по данным безопасности, любые принятые подходы для решения указанных задач, необходимо обосновать соответствующими ссылками на источники.

Методы моделирования, включающие сведения об изученных фармакологических параметрах, степени приверженности протоколу отдельных субъектов или данных о биологических эффектах, могут служить ценным источником информации о реальной или потенциальной эффективности, особенно во взаимосвязи с оценкой терапевтических эффектов. Необходимо четко обозначить допущения, лежащие в основе таких моделей, и подробно описать ограничения в отношении выводов.

5.5.6. Поправка на значимость и доверительные уровни

Если имеет место множественность, то частотный подход к анализу данных клинического исследования может потребовать поправки на ошибку I рода. Множественность, к примеру, возникает при нескольких первичных переменных (см. раздел 5.2.2.2), множестве сравнений, повторных оценках во времени и (или) вследствие промежуточного анализа (см. раздел 5.4.5). С целью недопущения или уменьшения множественности в некоторых случаях предпочтительно использовать определенные методы, как то выбор ключевой первичной переменной (если их несколько), выбор одного критического контраста (сравнение проводится только с ним), использование обобщающих методов, например, «площади под кривой» (для повторных измерений). В подтверждающем анализе все вопросы множественности, сохраняющиеся после принятия указанных мер, необходимо изложить в протоколе. Во всех случаях необходимо предусмотреть поправки, а в плане исследования

описать подробности процедуры внесения поправок или объяснения, почему они не были произведены.

5.5.7. Подгруппы, взаимодействия и ковариаты

Первичные переменные зачастую зависят от влияния факторов, не относящихся к лекарственному препарату. Например, возможна взаимосвязь с ковариатами, как то: возраст и пол, или возможны различия между определенными подгруппами субъектов, например, проходящих лечение в различных центрах многоцентрового исследования. В некоторых случаях поправки на влияние ковариат или эффекты подгрупп являются неотъемлемой частью запланированного анализа, что требует описания их в протоколе. С целью повышения точности и достижения баланса между сравниваемыми группами необходимо заранее выявить ковариаты и факторы, которые могут значимо повлиять на первичные переменные, а также предложить способы их учета в анализе. Если для стратификации дизайна используются один или более факторов, необходимо учесть такие факторы в анализе. Если конкретное значение поправочного коэффициента не найдено, анализ без введения поправок рекомендуется представить как основной, а анализ после введения поправок — как вспомогательный. Необходимо уделить особое внимание эффектам центров и влиянию значений исходных измерений на первичную переменную. Не рекомендуется вводить в основные анализы поправки на ковариаты, выявленные после рандомизации, так как они могут быть подвержены влиянию исследуемых лекарственных препаратов.

Терапевтический эффект сам может варьировать в зависимости от подгрупп или ковариат, например, эффект может снижаться с возрастом или быть более выражен у субъектов с определенным диагнозом. В некоторых случаях такие изменения ожидаемы или являются предметом изучения (например, пожилые), в связи с чем анализ подгрупп или статистическая модель, учитывающая взаимодействие, является частью намеченного подтверждающего анализа. Однако в большинстве случаев анализы подгрупп и взаимодействий являются поисковыми и требуют ясного указания на сей счет. С их помощью выявляется единообразие всех терапевтических эффектов в целом. Как правило, такие анализы осуществляются за счет добавления параметра «взаимодействие» в рассматриваемую статистическую модель, а затем она дополняется вспомогательным поисковым анализом в соответствующих подгруппах субъектов или стратах, составленных по ковариатам. Будучи поисковыми, результаты таких анализов требуют тщательной интерпретации; выводы о терапевтической эффективности (или ее отсутствии) или безопасности, основанные исключительно на поисковых анализах подгрупп, неприемлемы.

5.5.8. Целостность данных и надежность программного обеспечения

Достоверность численных результатов анализа зависит от качества и надежности методов и программного обеспечения (созданного самостоятельно или приобретенного), использованных для управления данными (ввод, хранение, верификация, коррекция и извлечение) и их статистической обработки. В связи с этим деятельность по управлению данными должна основываться на продуманных и действенных стандартных операционных процедурах. Для управления данными и статистического анализа, сопровождаемого документацией о надлежащих процедурах его тестирования, необходимо использовать надежное программное обеспечение.

5.6. ОЦЕНКА БЕЗОПАСНОСТИ И ПЕРЕНОСИМОСТИ

5.6.1. Объем оценки

Важным элементом всех клинических исследований является оценка безопасности и переносимости (см. Глоссарий). На ранних фазах такая оценка носит пре-

имущественно поисковый характер и чувствительна только к явным проявлениям токсичности, тогда как на поздних фазах профиль безопасности и переносимости лекарственного препарата описывается более подробно на большой выборке субъектов. Контролируемые исследования более поздних фаз являются важным способом объективного выявления новых потенциальных нежелательных явлений, даже если мощность исследования недостаточна для этих целей.

Определенные исследования планируют с целью подтверждения превосходства или эквивалентности в отношении безопасности и переносимости по сравнению с другим лекарственным препаратом или иной дозой исследуемого лекарственного препарата. Эти гипотезы должны подтверждаться соответствующими данными из подтверждающих исследований подобно тому, как подтверждают эффективность.

5.6.2. Выбор переменных и сбор данных

В каждом клиническом исследовании методы оценки безопасности и переносимости лекарственного препарата зависят от ряда факторов, включая знания о нежелательных явлениях родственных лекарственных препаратов, данные доклинических и ранее проведенных клинических исследований, возможные последствия фармакодинамических и фармакокинетических свойств отдельного лекарственного препарата, путь введения, исследуемая популяция субъектов и длительность исследования. Основной пласт данных о безопасности и переносимости формируют результаты лабораторных исследований в части клинической биохимии и клинических анализов крови, показатели жизнедеятельности и клинические нежелательные явления (заболевания, признаки и симптомы). Особенно важно регистрировать серьезные нежелательные явления и отмену терапии вследствие нежелательных явлений (см. ICH E2A и методические рекомендации по структуре и содержанию отчета о клинических исследованиях).

Более того, с целью объединения данных из различных исследований рекомендуется использовать единую методологию сбора и оценки данных на протяжении всей программы клинической разработки. Структурно такой словарь позволяет обобщать данные о нежелательных явлениях по трем различным уровням: системно-органный класс, предпочтительный термин и включенный термин (см. Глоссарий). Предпочтительный термин — уровень обобщения нежелательных явлений, предпочтительные термины, принадлежащие одному и тому же системно-органному классу допускается объединять вместе с целью описательного представления данных (см. ICH M1).

5.6.3. Совокупность субъектов, подлежащая оценке, и представление данных

Для итоговой оценки безопасности и переносимости совокупность субъектов, подлежащая объединению, как правило, включает всех, кто принял хотя бы одну дозу исследуемого лекарственного препарата. У указанных субъектов необходимо наиболее полно собирать данные о переменных безопасности и переносимости, включая вид нежелательного явления, тяжесть, время начала и продолжительность (см. также руководство ICH E2B). У особых групп пациентов, как то женщины, пожилые (см. методические рекомендации по проведению клинических исследований в особых группах пациентов), тяжелобольные, получающие стандартную сопутствующую терапию, может потребоваться дополнительная оценка безопасности и переносимости. Эти оценки требуют рассмотрения нескольких особых вопросов (см. методические рекомендации по структуре и содержанию отчета о клинических исследованиях).

Во время оценки все переменные безопасности и переносимости требуют внимания, для этого в протоколе исследования необходимо принять широкий подход. Необходимо сообщать обо всех нежелательных явлениях, независимо от того, связаны они или не связаны с лекарственным препаратом. Для оценки необходимо исполь-

зывать все имеющиеся данные об исследуемой популяции. Необходимо тщательно подбирать определения единицам измерения и диапазоны сравнения лабораторных переменных. Если в одном и том же исследовании используются различные единицы или различные диапазоны сравнения (например, при вовлеченности нескольких лабораторий), то их необходимо надлежащим образом стандартизовать с целью унифицированной оценки. Использование шкалы токсичности необходимо заранее оговорить и обосновать.

Частоту возникновения определенных нежелательных явлений обычно выражают в виде отношения количества субъектов, у которых оно возникло, к количеству субъектов, подверженных риску. Однако способ оценки частоты не всегда очевиден. Например, в зависимости от ситуации за знаменатель может приниматься количество субъектов, подвергшихся экспозиции, или степень экспозиции (в человеко-годах). Независимо от того, является ли целью вычисление величины риска или сравнение между исследуемыми группами, в протоколе необходимо описать способ расчета частоты. Это особенно важно для длительного лечения, в котором ожидается большое количество выбываний и смертей. Для таких случаев, во избежание риска недооценки, необходимо предусмотреть анализ выживаемости и вычислить кумулятивную частоту нежелательных явлений.

Если имеется значительное количество признаков и симптомов — так называемый фоновый шум (например, в психиатрических исследованиях), для оценки риска различных нежелательных явлений необходимо предусмотреть методы его коррекции. Одним из таких методов является использование концепции «вызванное лечением» (см. Глоссарий), согласно которой нежелательное явление учитывается только в случае его возникновения или усугубления по сравнению с периодом до начала лечения.

Другими методами снижения фонового шума являются игнорирование нежелательных явлений легкой степени или требование к его сохранению на протяжении ряда визитов, чтобы его можно было включить в числитель. В протоколе необходимо описать и обосновать указанные методы.

5.6.4. Статистическая оценка

Изучение безопасности и переносимости — комплексная проблема. Несмотря на то, что определенные нежелательные явления, как правило, ожидаемы, и в отношении любого лекарственного препарата за ними устанавливается наблюдение. Диапазон возможных нежелательных явлений достаточно большой, поэтому всегда возможны новые, непредсказуемые явления. Более того, нежелательное явление, возникшее вследствие нарушения протокола, как то применение неразрешенного лекарственного препарата, может стать источником систематических ошибок. Все это служит причиной статистических затруднений, связанных с аналитической оценкой безопасности и переносимости лекарственных препаратов, таким образом, однозначные выводы, полученные по результатам подтверждающих клинических исследований, скорее, исключение, чем правило.

В большинстве клинических исследований вопросы безопасности и переносимости наилучшим образом решаются с помощью методов описательной статистики с расчетом, при необходимости, доверительных интервалов. Также рекомендуется использовать графики, в которых закономерность частоты нежелательных явлений отражена как между сравниваемыми группами, так и между субъектами.

В некоторых случаях вычисление p -значений полезно либо для содействия в оценке интересующих различий, либо в качестве маркера, применяемого при большом количестве переменных безопасности и переносимости, с целью обозначения различий, требующих повышенного внимания. Это особенно эффективно в отношении лабораторных данных, которые трудно группировать иным образом. Рекомендуется подвергать лабораторные данные как количественному анализу, например, оце-

нивать терапевтические средние, так и качественному, при котором рассчитываются доли явлений, находящихся выше или ниже определенного порога.

Если тестируется гипотеза, необходимо ввести статистические поправки на множественность с целью надлежащего контроля ошибки I рода, однако ошибка II рода требует большего внимания. Следует осторожно интерпретировать предполагаемые статистически значимые результаты, если поправка на множественность не применялась.

Большинство исследований направлены на доказательство того, что клинически неприемлемые различия по безопасности и переносимости в сравнении с активным контролем или плацебо отсутствуют. Также как и для подтверждения эффективности с помощью исследований эквивалентности и не меньшей эффективности для тестирования гипотезы рекомендуется использовать доверительные интервалы. Однако в таком случае отмечается большая неопределенность, обусловленная низкой частотой возникновения нежелательных явлений.

5.6.5. Обобщенное резюме

Профиль безопасности и переносимости лекарственного препарата на протяжении его разработки и, в особенности, в период государственной регистрации постоянно подвергается пересмотру по результатам клинических исследований. Однако польза такого резюме зависит от отдельных хорошо контролируемых исследований с высоким качеством данных.

Во всех случаях совокупная польза лекарственного препарата представляет собой баланс между риском и пользой, изучение которого можно предусмотреть и в отдельном исследовании, несмотря на то, что оценку отношения ожидаемой пользы к возможному риску осуществляют по совокупным результатам всех проведенных клинических исследований (см. раздел 5.7.2.2).

Более детально отчетность по безопасности и переносимости представлена в главе 6 настоящего издания «Структура и содержание отчетов о клинических исследованиях».

5.7. ОТЧЕТНОСТЬ

5.7.1. Оценка и отчетность

Как указано во введении, структура и содержание отчетов о клиническом исследовании являются предметом главы 6 настоящего издания «Структура и содержание отчетов о клинических исследованиях». В них всесторонне описана статистическая отчетность, которая должным образом взаимосвязана с клиническими и прочими данными. В связи с этим настоящий раздел относительно небольшой.

При планировании исследования в протоколе необходимо отразить описанные в разделе 5.5 настоящей главы ключевые элементы анализа. По завершении исследования, после компоновки данных для предварительного изучения рекомендуется осуществить проверку подлежащих анализу данных без снятия ослепления, описанную в разделе 5.5. Такая предварительная проверка без снятия ослепления должна охватывать значимые решения, например, об исключении субъектов или данных из совокупности, подлежащей анализу; в ее рамках определяется необходимость трансформации, и выявляются выбросы; в модель могут добавиться ковариаты, обнаруженные в ранее проведенных исследованиях; пересматривается решение об использовании параметрических или непараметрических методов. Принятые на этом этапе решения необходимо отразить в отчете, их необходимо отличать от решений, принятых статистиком после раскрытия кодов рандомизации, так как принятые вслепую решения вводят меньше систематических ошибок. Статистиков и другой персонал, принимавший участие в неослепленном промежуточном анализе, нельзя допускать к

проверке данных без снятия ослепления или коррекции плана статистического анализа. Если ослепление может быть под угрозой ввиду явных лекарственных эффектов, прослеживающихся в документах, необходимо с особой осторожностью идти к проверке данных без снятия ослепления.

Многие вопросы представления данных и составление таблиц необходимо решить к моменту проверки данных без снятия ослепления, чтобы к моменту анализа был готов полный план, включающий все интересующие вопросы, в том числе отбор субъектов, выбор данных и их модификацию, обобщенные и табличные данные, оценку и тестирование гипотезы. После завершения валидации данных анализ проводится согласно намеченному плану; чем больше анализ соответствует плану, тем выше достоверность полученных результатов. Необходимо уделить особое внимание различиям между запланированным (в протоколе, поправка к протоколу, обновленному по результатам проверки данных без снятия ослепления плану статистического анализа) и реальным анализом. Необходимо представить подробные основания всем отклонениям от запланированного анализа.

В отчете об исследовании необходимо описать всех субъектов, включенных в исследование, независимо от того, попали они в анализ или нет. Необходимо идентифицировать все основания исключения из анализа, а также основания исключения субъектов, вошедших в полную совокупность, подлежащую анализу, из совокупности по протоколу. Точно так же для всех субъектов, включенных в совокупность, подлежащую анализу, необходимо учесть значения всех важных переменных в значимых временных точках.

Необходимо подробно описать влияние на анализ первичных переменных случаев невозможности последующего наблюдения, недостающих данных, и исключений субъектов из исследования и основных нарушений протокола. Необходимо идентифицировать субъектов, за которыми было невозможно установить наблюдение, исключенных из исследования и грубо нарушавших протокол, представить описательный анализ, включая причины их выбывания и взаимосвязь между приемом лекарственного препарата и исходом.

Описательная статистика является непреложной частью отчетов. В соответствующих таблицах и (или) на графиках необходимо четко представить важные связи первичных и вторичных переменных, а также ключевые прогностические и дефицитические переменные. Предметом особо тщательного описания должны являться результаты основных анализов, относящихся к целям исследования. При описании результатов тестов на значимость необходимо указывать точные p -значения (например, $p=0,034$), а не ограничиваться ссылкой на критические значения.

Несмотря на то что основной задачей анализа клинического исследования является ответ на вопросы, диктуемые основными целями, в ходе неослепленного исследования полученных данных могут возникнуть новые вопросы. Это может потребовать дополнительного, возможно, более сложного статистического анализа. В отчете необходимо четко обозначить все дополнительные действия, не запланированные протоколом.

Существует вероятность возникновения непредвиденного дисбаланса сравниваемыми группами в части исходных показателей, не предусмотренных в качестве ковариат в планируемом анализе, но, тем не менее, обладающих определенной прогностической ценностью. Лучшим способом решения такой проблемы в отчете служит подтверждение, что дополнительный анализ, учитывающий указанный дисбаланс, приводит к тем же заключениям, что и запланированный. Если это сделать не удастся, необходимо вынести на обсуждение влияние дисбаланса на выводы.

В целом, желательно воздерживаться от незапланированных анализов. Такие анализы обычно проводятся, когда предполагается, что терапевтический эффект может изменяться в зависимости от иных факторов. В этом случае следует попы-

установить подгруппы, для которых эффект оказался наиболее полезен. Потенциальные угрозы чрезмерно оптимистичной интерпретации результатов незапланированных анализов подгрупп хорошо известны (см. также раздел 5.5.7), поэтому их следует избегать. Несмотря на то что те же затруднения при интерпретации возникают, если лекарственный препарат оказывается неэффективным или оказывает нежелательное явление в подгруппе субъектов, необходимо надлежащим образом оценить такую возможность и должным образом отразить все в отчете.

Окончательные статистические выводы должны основываться на анализе, интерпретации и представлении результатов клинического исследования. До этого момента статистик исследования должен входить в команду, ответственную за подготовку отчета о клиническом исследовании, он обязан заверить его.

5.7.2. Обобщение клинической базы данных

В регистрационном досье необходимо представить обобщенное резюме и синтез всех доказательств по безопасности и эффективности, полученных по результатам всех представленных клинических исследований (отчет специалиста). В соответствующих случаях в дополнение допускается представлять статистическую комбинацию результатов.

В рамках резюме возникает ряд интересных с точки зрения статистики вопросов: описание демографии и клинических характеристик популяции, получавшей лечение во время программы клинической разработки; рассмотрение ключевых аспектов эффективности с проведением сопоставления результатов значимых (как правило, контролируемых) исследований и обращением внимания на степень их взаимодополнения или противоречия друг другу; обобщение данных по безопасности, доступных из объединенной базы данных всех исследований, результаты которых значимы для регистрационного досье и указывают на потенциальные угрозы безопасности. Во время планирования клинического исследования необходимо уделить особое внимание единообразию определений и единиц измерений, которое будет способствовать последующей интерпретации результатов группы исследований, особенно если они связаны между собой. Необходимо использовать единую терминологию для ведения записей о назначаемом лечении, анамнезе и нежелательных явлениях. Почти всегда целесообразно общее определение первичных и вторичных переменных, оно особенно важно для мета-анализа. Способ измерения ключевых переменных эффективности, сроки оценки в зависимости от рандомизации/включения, обращение с нарушениями и отклонениями от протокола и, возможно, определение прогностических факторов — всё это должно быть единообразно, если только тому нет обоснованных причин.

Необходимо подробно описать статистические процедуры, используемые для объединения данных разных исследований. Необходимо уделить внимание возможному возникновению систематических ошибок, обусловленных выбором исследований, однородностью их результатов и надлежащим моделированием различных источников вариаций. Необходимо изучить чувствительность выводов к допущениям и принятым решениям.

5.7.2.1. Данные по эффективности

Отдельное клиническое исследование всегда должно быть достаточно большим, чтобы достичь поставленных целей. Объединяя ряд клинических исследований, в которых рассматривались идентичные ключевые аспекты эффективности, можно получить дополнительную ценную информацию. Чтобы осуществить сравнения, основные результаты такой совокупности исследований необходимо представить в одной и той же форме, обычно в виде таблиц или графиков, в которых указываются оценки и их доверительные границы. Использование мета-анализа с целью объединения указанных оценок является полезным дополнением, так как это позволяет получить

более точную общую оценку величины терапевтических эффектов и обеспечивая полное и всестороннее обобщение результатов исследований. В исключительных случаях этот подход может служить наиболее подходящим (или даже единственным возможным) способом обеспечения совокупного подтверждения эффективности путем тестирования общей гипотезы. При использовании мета-анализа с этой целью для него необходимо составить отдельный протокол.

5.7.2.2. Данные по безопасности

При обобщении данных по безопасности необходимо подробно изучить базу данных по безопасности с целью обнаружения признаков потенциальной токсичности и последующего наблюдения за любыми признаками путем регистрации любых связанных закономерностей из указанных наблюдений. Объединение данных по безопасности от всех лиц, подвергнувшихся воздействию лекарственного препарата является важным источником данных, так как такой размер выборки повышает вероятность обнаружения более редких нежелательных явлений и, возможно, позволяет установить их частоту. Однако из-за отсутствия группы сравнения на основании такой базы данных сложно оценить частоту, поэтому данные сравнительных исследований имеют особую ценность для преодоления указанных затруднений. Результаты исследований, в которых используется стандартный способ сравнения (плацебо или активный контроль), необходимо объединить и представить отдельно по каждому способу, обеспечивая достаточный объем данных.

Необходимо сообщать обо всех признаках потенциальной токсичности, обнаруживаемых по результатам изучения данных. При оценке реалистичности таких потенциальных нежелательных явлений необходимо учитывать проблему множественности, возникающую в связи с неоднократными сравнениями. При оценке необходимо также надлежащим образом применять методы анализа выживаемости для использования потенциального отношения частоты нежелательных явлений к длительности экспозиции и (или) наблюдения. Необходимо надлежащим образом рассчитать риски, обусловленные выявленными нежелательными явлениями, чтобы обеспечить надлежащую оценку отношения ожидаемой пользы к возможному риску.

ГЛОССАРИЙ

Байесовские подходы (Bayesian Approaches)

Подходы к анализу данных, обеспечивающие апостериорную вероятность распределения некоторого параметра (например, терапевтического эффекта) на основании наблюдений и априорной вероятности распределения этого параметра. Затем апостериорную вероятность используют для построения статистических выводов.

Систематическая ошибка (статистическая и операционная) (Bias (Statistical & Operational))

Систематическая тенденция каких-либо элементов дизайна, проведения, анализа и оценки результатов клинического исследования отклонять оценку терапевтического эффекта от ее истинного значения. Систематические ошибки, возникающие в связи с отклонениями в проведении исследования, именуются «операционными». Систематические ошибки, возникающие из других вышеуказанных источников, — статистические.

Проверка данных без снятия ослепления (Blind Review)

Проверка и оценка данных в промежуток времени между завершением исследования (последнее наблюдение последнего субъекта) и снятием ослепления в целях финализации запланированного анализа.

Содержательная валидность (Content Validity)

Способность переменной (например, рейтинговой шкалы) измерять то, что она должна измерять.

Двойная пустышка (Double-Dummy)

Метод сохранения ослепления в клиническом исследовании, когда невозможно достичь идентичности двух сравниваемых лекарственных препаратов. Для этих целей для лекарственного препарата А готовят не отличимый от него плацебо, ровно как и для лекарственного препарата В готовят не отличимый от него плацебо. Затем субъекты получают два набора лекарственных препаратов: либо А (активный) и В (плацебо), либо А (плацебо) и В (активный).

Выбывание (Dropout)

Субъект клинического исследования, который по каким-либо причинам не способен продолжать исследование до запланированного протоколом последнего визита.

Исследование эквивалентности (Equivalence Trial)

Исследование, основной целью которого является подтверждение, что эффект двух и более лекарственных препаратов не отличается друг от друга на клинически значимую величину. Обычно это достигается подтверждением того, что истинное терапевтическое различие укладывается в интервал, ограниченный верхней и нижней границами эквивалентности, и считается клинически допустимым.

Частотные методы (Frequentist Methods)

Статистические методы, как то: критерии значимости и доверительные интервалы, которые можно оценить с позиций частоты возникновения определенных исходов, возникающих в гипотетических повторениях одного и того же эксперимента.

Полная совокупность, подлежащая анализу (Full Analysis Set)

Совокупность субъектов, которая максимально соответствует идеалу совокупности «по намерению лечить».

Обобщаемость, генерализация (Generalisability, Generalisation)

Степень, с которой результаты клинического исследования можно надежно экстраполировать с субъектов, в нем участвовавших, на большую популяцию пациентов и более широкие клинические условия.

Переменная общей оценки (Global Assessment Variable)

Единая переменная, обычно являющаяся категориальной рейтинговой шкалой, которая включает объективные переменные и общее впечатление исследователя о состоянии или изменении состояния субъекта.

Независимый комитет по мониторингу данных (НКМД) (Комиссия по мониторингу данных и безопасности, Комитет по мониторингу, Комитет по мониторингу данных) (Independent Data Monitoring Committee (IDMC) (Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee))

Независимый комитет по мониторингу данных, который может учредить спонсор с целью оценки через регулярные интервалы течения клинического исследования, данных по безопасности, ключевые конечные точки эффективности и составления рекомендаций для спонсора о необходимости продолжения, модификации или прекращения исследования.

Принцип «по намерению лечить» (Intention-to-Treat Principle, ITT)

Принцип, устанавливающий, что наилучшей совокупностью для оценки терапевтического эффекта является совокупность субъектов, которых намеревались лечить (то есть планируемый режим), а не совокупность, фактически получившая лечение. Вследствие этого, необходимо наблюдать, оценивать и анализировать всех рандомизированных субъектов, независимо от приверженности их предусмотренному режиму.

Взаимодействие (качественное и количественное) (Interaction (Qualitative & Quantitative))

Ситуация, когда различие терапевтических эффектов (например, различие между исследуемым лекарственным препаратом и контролем) зависит от другого фактора (например, центра). Под количественным взаимодействием подразумевается такое, когда степень различий зависит от степени влияния фактора, тогда как при качественном взаимодействии разная степень различий может наблюдаться при изменении фактора.

Межэкспертная надежность (Inter-Rater Reliability)

Способность получать эквивалентные результаты различными экспертами в различных условиях.

Внутриэкспертная надежность (Intra-Rater Reliability)

Способность получать эквивалентные результаты одним и тем же экспертом в различных условиях.

Промежуточный анализ (Interim Analysis)

Любой анализ, направленный на сравнение исследуемых групп по эффективности или безопасности в любой момент до формального завершения исследования.

Мета-анализ (Meta-Analysis)

Формальная оценка количественного подтверждения, полученного из двух и более исследований, проведенных с одной и той же целью. Наиболее часто он заключается в статистическом объединении статистических резюме различных исследований, но в некоторых случаях могут объединяться и исходные (необработанные) данные.

Многоцентровое исследование (Multicentre Trial)

Клиническое исследование, проведенное по единому протоколу более чем в одном центре, в котором таким образом принимали участие более одного исследователя.

Исследование не меньшей эффективности (безопасности) (Non-Inferiority Trial)

Исследование, основной целью которого является подтверждение, что эффект исследуемого лекарственного препарата с клинической точки зрения не меньше, чем лекарственного препарата сравнения (активного контроля или плацебо).

Предпочтительные и включенные термины (Preferred and Included Terms)

В иерархическом медицинском словаре, например, MedDRA, включенный термин является термином низшего уровня, с помощью которого кодируется описание исследователя. Предпочтительный термин — уровень группировки включенных терминов, используемый для описания частоты возникновения явления. Например, указание исследователя следующего содержания: «Боль в левой руке» может быть закодировано включенным термином как «Боль в суставе», которому соответствует предпочтительный термин — «Артралгия».

Совокупность «по протоколу» («соответствующие протоколу», «надлежащие единицы», выборка «эффективности», выборка «субъектов, которых можно оценить») (Per Protocol Set, PP (Valid Cases, Efficacy Sample, Evaluable Subjects Sample))

Совокупность данных, полученная от подмножества субъектов, приверженных протоколу в достаточной степени, чтобы подтвердить, что согласно таким данным, исходя из предложенной научной модели, эффекты лекарственного препарата подтверждаются. Под приверженностью подразумевается применение лекарственного препарата, доступность результатов измерений, отсутствие серьезных нарушений протокола.

Безопасность и переносимость (Safety & Tolerability)

Безопасность лекарственного препарата заключается в медицинском риске для субъектов, оцениваемом, как правило, по результатам лабораторных исследований (включая клиническую биохимию и клинические анализы крови), показателям жизнедеятельности, клиническим нежелательным явлениям (заболеваниям, признакам и симптомам) и другим исследованиям безопасности (например, ЭКГ, данные офтальмологического обследования). Переносимость лекарственного препарата означает способность субъекта переносить явно наблюдающиеся нежелательные явления.

Статистический план анализа (Statistical Analysis Plan)

План статистического анализа — документ, содержащий более подробные технические детали основных элементов анализа, описанных в протоколе; он включает подробное описание процедур проведения статистического анализа первичных и вторичных переменных и других данных.

Исследование превосходства (Superiority Trial)

Исследование, основной целью которого является подтверждение, что эффект исследуемого лекарственного препарата превосходит таковой лекарственного препарата сравнения (активного контроля или плацебо).

Суррогатная переменная (Surrogate Variable)

Переменная, обеспечивающая не прямое измерение эффекта в случаях, когда прямое измерение клинического эффекта невозможно или нецелесообразно.

Терапевтический эффект (Treatment Effect)

Эффект лекарственного препарата, проявляющийся в клиническом исследовании. В большинстве клинических исследований искомый терапевтический эффект заключается в сравнении (или противопоставлении) двух и более лекарственных препаратов.

Вызванный лечением (Treatment Emergent)

Явление, возникающее в период лечения, отсутствующее до его начала, или которое усугубляется относительно состояния, наблюдавшегося перед лечением.

Статистик исследования (Trial Statistician)

Статистик, обладающий знаниями/умениями и опытом, достаточным для реализации принципов, обозначенных в настоящем документе, и ответственный за статистическую сторону исследования.

Литература

1. Statistical Principles for Clinical Trials, E9 // International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use [официальный сайт]. http://www.ich.org/fileadmin/ /Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (дата обращения: 17.09.2012).
2. Клиническая эпидемиология, основы доказательной медицины Р. Флетчер, С. Флетчер, Э. Вагнер / Пер. с англ. М.: Медиа Сфера, 1998. 352 с.
3. Сергиенко В.И., Бондарева И.Б. Математическая статистика в клинических исследованиях. 2-е изд., перераб. и доп. М., ГЭОТАР-Медиа, 2006. 304 с.
4. Гланц С. Медико-биологическая статистика / Пер. с англ. М., Практика, 1998. 459 с.
5. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М.: Медиа Сфера, 2002. 312 с.
6. Ланг Т.А. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов / Т.А. Ланг, М. Сесик / Пер. с англ. / Под ред. В.П. Леонова. М.: Практическая медицина, 2011. 480 с.
7. Thomas D. Cook, David L DeMets. Introduction to Statistical Methods for Clinical Trials (Chapman & Hall / CRC Texts in Statistical Science). Chapman and Hall / CRC 1 edition (November 19, 2007). 464 p.
8. Olga Korosteleva. Clinical Statistics: Introducing Clinical Trials, Survival Analysis and Longitudinal Data Analysis // Jones & Bartlett Publishers; 1 edition (November 11, 2008). 120 p.
9. Alan Agresti. Categorical Data Analysis (Wiley Series in Probability and Statistics Wiley-Interscience; 2 edition (July 22, 2002). 734 p.
10. Joseph Tal. Strategy and Statistics in Clinical Trials: A non-statisticians guide to thinking, designing and executing // Academic Press; 1 edition (July 28, 2011). 278 p.
11. Steven A. Julious. Sample Sizes for Clinical Trials // Chapman and Hall/CRC; 1 edition (August 26, 2009). 317 p.
12. Jozef Nauta. Statistics in Clinical Vaccine Trials. Springer; 1st edition (October 2, 2010). 171 p.