

Особенности и значимость оценки подгрупп в подтверждающих клинических исследованиях

О. И. Басова*, И. В. Лыскова, О. Ю. Иванова

Федеральное государственное бюджетное учреждение
«Научный центр экспертизы средств медицинского применения»
Министерства здравоохранения Российской Федерации,
Петровский б-р, д. 8, стр. 2, Москва, 127051, Российская Федерация

Резюме. Анализ подгрупп пациентов играет важную роль в интерпретации результатов подтверждающих клинических исследований и в большинстве случаев обязателен к выполнению. Целью анализа является оценка согласованности (гетерогенности) терапевтического эффекта в подгруппах пациентов, выделенных на основании демографических характеристик, стадии и степени тяжести основного заболевания, наличия определенной генетической мутации и др. Однако существующие методологические проблемы (проблема множественных сравнений, случайные различия между подгруппами и т.д.) затрудняют его проведение и часто приводят к спорным выводам. Цель работы — анализ и систематизация подходов зарубежных регуляторных органов к оценке подгрупп в подтверждающих клинических исследованиях, формирование научно обоснованных требований к ее выполнению и интерпретации результатов спонсорами клинических исследований и экспертами при проведении экспертизы отношения ожидаемой пользы к возможному риску применения лекарственных препаратов с целью их регистрации по национальным правилам и по правилам Евразийского экономического союза. В работе рассмотрены цели проведения анализа подгрупп и статистические подходы к его выполнению, приведены примеры такого анализа из регуляторной практики. Описаны подходы к интерпретации анализа подгрупп в зависимости от наличия или отсутствия доказательства основной гипотезы исследования, характера обнаруженных гетерогенных эффектов исследуемого препарата в подгруппах и выбора самих подгрупп. Освещены проблемные аспекты анализа подгрупп, возможные противоречия интерпретации полученных результатов, регуляторные ожидания. Представленные рекомендации могут быть использованы экспертами при проведении экспертизы отношения ожидаемой пользы к возможному риску применения лекарственного препарата, а также разработчиками лекарственных средств при подготовке протоколов и отчетов клинических исследований.

Ключевые слова: подгруппы; подтверждающий анализ подгрупп; поисковый анализ подгрупп; взаимодействие; гетерогенность; статистические методы; клинические исследования; регуляторные органы; эффективность и безопасность; стратификация

Для цитирования: Басова ОИ, Лыскова ИВ, Иванова ОЮ. Особенности и значимость оценки подгрупп в подтверждающих клинических исследованиях. *Ведомости Научного центра экспертизы средств медицинского применения.* 2021;11(2):81–93. <https://doi.org/10.30895/1991-2919-2021-11-2-81-93>

* **Контактное лицо:** Басова Ольга Игоревна; basovaoi@expmed.ru

Specific Aspects and Significance of Subgroup Assessment in Confirmatory Clinical Trials

O. I. Basova*, I. V. Lysikova, O. Yu. Ivanova

Scientific Centre for Expert Evaluation of Medicinal Products,
8/2 Petrovsky Blvd, Moscow 127051, Russian Federation

Abstract. Patient subgroup analysis plays an important role in interpretation of confirmatory clinical trial results and is mandatory in most cases. The purpose of subgroup analysis is to assess the consistency (heterogeneity) of the treatment effect in subgroups of patients identified based on such characteristics as demographics, stage and severity of the underlying disease, presence of a certain genetic mutation, etc. However, existing methodological issues (the problem of multiple comparisons, detection of differences between subgroups by chance alone, etc.) make it difficult to carry out the analysis and often lead to controversial conclusions. The aim of the study was to analyse and summarise foreign regulatory approaches to subgroup analysis in confirmatory clinical trials, and to elaborate science-based requirements for subgroup analysis and interpretation of the results by clinical trial sponsors and experts when assessing the risk-benefit ratio of medicinal products for the purpose of their authorisation in Russia and the Eurasian Economic Union. This paper discusses the objectives of subgroup analysis and statistical approaches to its performance, provides relevant examples of such analysis from regulatory practice. It describes approaches to interpretation of subgroup analysis depending on the presence/absence of evidence supporting the primary hypothesis of the study, the nature of the experimental medicinal product's heterogeneous effects in the subgroups, and selection of the subgroups. The paper highlights areas of concern in subgroup analysis, potential controversies in interpretation of the obtained results, and regulatory expectations. The recommendations presented in the paper can be used by experts in the assessment of the risk-benefit ratio, as well as by medicine developers in the preparation of clinical trial protocols and reports.

Key words: subgroups; confirmatory subgroup analysis; exploratory subgroup analysis; interaction; heterogeneity; statistical methods; clinical trials; regulatory authorities; efficacy; safety; stratification

For citation: Basova OI, Lysikova IV, Ivanova OYu. Specific aspects and significance of subgroup assessment in confirmatory clinical trials. *Vedomosti Nauchnogo tsentra ekspertizy sredstv meditsinskogo primeneniya = The Bulletin of the Scientific Centre for Expert Evaluation of Medicinal Products.* 2021;11(2):81–93. <https://doi.org/10.30895/1991-2919-2021-11-2-81-93>

* **Corresponding author:** Olga I. Basova; basovaoi@expmed.ru

Вопросами научной методологии анализа подгрупп в клинических исследованиях (КИ) и его применением в регуляторной практике занимаются научные коллективы по всему миру. В 2015 г. рабочая группа Управления по контролю за качеством продуктов питания и лекарственных средств США (Food and Drug Administration, FDA) подготовила технический документ по анализу подгрупп в подтверждающих исследованиях [1], в 2019 г. Европейское агентство по лекарственным средствам (European Medicines Agency, EMA) выпустило руководство по методологии и интерпретации результатов анализа подгрупп¹, содержание которого практически полностью соответствует вышеупомянутому документу FDA. Еще один концептуальный документ, основные выводы которого в целом согласуются с позицией регуляторных органов, был подготовлен совместной рабочей группой Европейской федерации статистиков в фармацевтической промышленности (European Federation for Statisticians in the Pharmaceutical Industry, EFSPi) и профессиональной организацией статистиков в фармацевтической промышленности (Statisticians in the Pharmaceutical Industry, PSI) PSI/EFSPi [2], представляющей интересы фармацевтических компаний — разработчиков лекарственных средств.

Фактически между регуляторными органами и спонсорами КИ был достигнут консенсус по общим вопросам анализа подгрупп (например, необходимость предварительной спецификации ключевых подгрупп, понятие биологического правдоподобия и т.д.), хотя имеющиеся разночтения в деталях еще не урегулированы и на практике могут приводить к серьезным спорным ситуациям. В частности, одна из центральных идей документа EMA², касающаяся поискового анализа подгрупп при изучении отношения ожидаемой пользы к возможному риску применения лекарственных препаратов (соотношение «польза — риск») в подтверждающих КИ, заключается в методологическом подходе «генерирования сигналов». Под «сигналом» подразумевается вероятность обнаружения подгруппы, в которой эффект лечения отличается по величине/знаку от значений, наблюдаемых в остальной популяции исследования, при этом номинальный уровень значимости теста взаимодействия p должен составлять $>0,05$. Предложенный методологический подход обусловлен необходимостью проверки согласованности терапевтического эффекта между разными подгруппами исследования для определения границ охвата целевой популяции и подтверждения заявленных разработчиком показаний к применению лекарственного препарата (ЛП). Согласно техническому документу FDA для обнаружения потенциально

важных статистических взаимодействий уровень значимости теста взаимодействия p может составлять 0,10 или даже 0,20, однако в этом документе подчеркивается некорректность утверждения о применимости данной рекомендации для каждого конкретного случая вне зависимости от контекста [1]. В подготовленном PSI/EFSPi документе высказаны опасения касательно принятия ошибочных регуляторных решений вследствие ложноположительно-го обнаружения взаимодействия при номинальном уровне значимости $p = 0,10$ и предложены альтернативные статистические подходы к выполнению анализа подгрупп [2]. На практике такие разночтения означают противоположные стратегии проведения анализа подгрупп и, следовательно, интерпретации результатов терапевтической эффективности ЛП в подгруппах пациентов.

Цель работы — анализ и систематизация подходов зарубежных регуляторных агентств к оценке подгрупп в подтверждающих клинических исследованиях; формирование научно обоснованных требований к ее выполнению и интерпретации спонсорами клинических исследований и экспертами при проведении экспертизы отношения ожидаемой пользы к возможному риску применения лекарственных препаратов с целью их регистрации по национальным правилам и по правилам Евразийского экономического союза.

КЛЮЧЕВЫЕ ПОНЯТИЯ АНАЛИЗА ПОДГРУПП

Основные термины, используемые в области анализа подгрупп в КИ³, перечислены ниже.

Гетерогенность (heterogeneity) отражает степень различий пациентов в целевой популяции или в популяции КИ в отношении прогностических факторов клинического исхода/величины терапевтического эффекта. Изучение эффектов лечения в четко определенных подгруппах пациентов тем важнее, чем более неоднородной является исследуемая популяция.

Согласованность (consistency) эффектов лечения в представляющих интерес подгруппах отражает степень применимости оценки общего терапевтического эффекта ко всей исследуемой популяции. По каждому конкретному предполагаемому показанию к применению требуется оценить степень различий эффекта лечения между клинически значимыми подгруппами (то есть выявление потенциальной несогласованности). Выявленная несогласованность в оценках эффектов лечения между подгруппами является поводом для проведения дальнейшего анализа подгрупп.

Достоверность (credibility) отражает степень доверия к полученным выводам по итогам проведенного анализа подгрупп и, следовательно,

¹ Guideline on the investigation of subgroups in confirmatory clinical trials. EMA/CHMP/539146/2013. EMA; 2019.

² Там же.

³ Там же.

возможность их использования в принятии соответствующих регуляторных решений. Достоверность зависит от степени научной обоснованности, в том числе клинической, и биологических предпосылок к получению таких результатов, а также их воспроизводимости.

Биологическое правдоподобие (biological plausibility) отражает возможность прогнозирования терапевтического эффекта (то есть различных эффектов лечения между подгруппами) на основании клинических и/или фармакологических знаний. Под указанной концепцией подразумевается наличие определенных клинических и фармакологических суждений, не поддающихся какой-либо количественной оценке и являющихся аналитическими по своей природе. Любые имеющиеся предположения, подкрепленные фактами (как внешними, так и внутренними, основанными на результатах собственных доклинических/клинических исследований), должны быть рассмотрены еще на стадии планирования исследования.

Репликация (replication) — воспроизводимость полученных результатов проведенного анализа подгрупп в нескольких источниках данных (как внутренних, так и внешних). Под этим термином подразумевается одинаковое влияние определенной ковариаты на величину терапевтического эффекта или наличие различных эффектов в определенных подгруппах, наблюдаемое не только в одном подтверждающем КИ, но и в независимых КИ (другое КИ фазы III, поисковое КИ фазы II или КИ вне программы клинической разработки препарата, но со схожими экспериментальными условиями).

ЦЕЛИ АНАЛИЗА ПОДГРУПП В ПОДТВЕРЖДАЮЩИХ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЯХ

Обобщенная оценка терапевтического эффекта по всей гетерогенной популяции исследования / совокупности исследований позволяет судить о терапевтическом охвате исследуемого препарата, то есть подразумевает проведение КИ в формате «pragmatic clinical trials» — максимально приближенных к реальной клинической практике и не ограниченных узко заданными критериями включения/невключения в исследование. Распространена практика выбора в качестве основной популяции для первичного анализа эффективности популяции по назначенному лечению (ITT популяции), подразумевающей наличие пациентов с разной комплаентностью и различными клиническими сценариями в исследовании.

Однако обобщенные по всей популяции исследования данные по первичному/вторичным критериям эффективности препарата не позволяют выявить потенциальные клинические фенотипы с разными ответами на лечение, в то время как сведения о терапевтических эффектах исследуемого

препарата у пациентов с различными клиническими характеристиками могут быть использованы для дальнейших исследований оценки эффективности и безопасности лечения и выработки клинических стратегий.

Таким образом, основными целями проведения анализа подгрупп являются:

1) оценка согласованности (при выявлении несогласованности — дальнейшая оценка степени гетерогенности) терапевтического эффекта между разными подгруппами исследования для определения показаний к применению (границ охвата целевой популяции);

2) поиск подгруппы, в которой применение препарата наиболее эффективно и/или безопасно (лучшее соотношение «польза — риск»);

3) поиск подгруппы с неблагоприятным соотношением «польза — риск».

ФАКТОРЫ ИДЕНТИФИКАЦИИ ЦЕЛЕВЫХ ПОДГРУПП

Факторы идентификации подгрупп можно разделить на прерандомизационные (например, стратификация при рандомизации и/или исходные значения переменных на скрининге) и пострандомизационные (например, подгруппа пациентов с терапевтическими неудачами). Прерандомизационные факторы выбирают исходя из существующих предпосылок потенциальной прогностической значимости фактора в отношении результатов лечения; пострандомизационные факторы идентификации подгрупп, как правило, носят исключительно поисковый характер. Ниже перечислены факторы, лежащие в основе идентификации различных подгрупп *a priori*⁴:

1) существуют настолько веские основания ожидать неравномерного ответа на лечение при разных уровнях некоторого фактора, что общая оценка терапевтического эффекта исследуемого препарата будет невозможна в рамках одного КИ. В этом случае необходимо запланировать проведение независимых КИ;

2) фактор считается прогностическим в отношении клинического исхода, или, по крайней мере, существуют некоторые биологические предпосылки или сторонние доказательства существования неравномерного терапевтического эффекта. В дополнение к стратификационным факторам при рандомизации в данную категорию можно включить демографические факторы, в том числе регион проживания пациента, и факторы, связанные с механизмом действия и/или фармакологическими свойствами препарата. Следует рассмотреть другие факторы, которые могут учитываться при прогнозировании ответа на лечение: стадия, тяжесть или фенотип заболевания, наличие генетических мутаций, использование сопутствующей терапии. Для того чтобы сделать заключение об эффективности препарата

⁴ Там же.

во всей популяции исследования, не требуется формальное доказательство его эффективности для всех ключевых подгрупп, идентифицированных на основании вышеперечисленных факторов (за исключением факторов, указанных в п. 1);

3) для заключения о наличии эффектов препарата по всей популяции исследования не требуется формальное доказательство эффективности препарата в подгруппе, характеризующейся идентификационным фактором, для которого приведены убедительные аргументы (фармакологическое обоснование и клинические данные) отсутствия влияния на оцениваемые эффекты исследуемого препарата (или же если имеющиеся данные не позволяют оценить его возможное влияние на правдоподобность согласованности эффектов препарата).

СТАТИСТИЧЕСКИЕ АСПЕКТЫ АНАЛИЗА ПОДГРУПП В КЛИНИЧЕСКИХ ИССЛЕДОВАНИЯХ

Традиционно статистические и методологические проблемы анализа подгрупп разделяют на объективные (инфляция ошибки I рода вследствие множественных сравнений, низкая мощность) и субъективные (отсутствие предварительной спецификации подгрупп и допущенные ошибки в статистическом анализе). Необходимость проведения анализа подгрупп не ставится под сомнение, речь идет о том, как максимально корректно его выполнять, не допуская методологических ошибок.

Статистические подходы к анализу подгрупп следует разделять в зависимости от цели его выполнения: анализ подгрупп как часть стратегии подтверждающего тестирования статистических гипотез в КИ фазы III и поисковый анализ подгрупп как часть оценки соотношения «польза — риск» в контексте согласованности эффекта лечения в популяции КИ. При проведении анализа подгрупп как части стратегии подтверждающего тестирования ключевые подгруппы должны быть определены заранее на этапе планирования КИ с использованием обязательной научной аргументации и стратификации по идентификационным факторам подгрупп при рандомизации. Наличие подгрупп должно учитываться на этапе оценки размера выборки КИ фазы III (например, указание о включении в КИ не менее 30% пациентов от общего запланированного числа участников для формирования ключевой подгруппы) с учетом необходимости введения поправки на множественное сравнение.

Количество ключевых подгрупп, подлежащих оценке, должно быть небольшим. Анализ по первичной конечной точке в такой подгруппе может быть указан в протоколе КИ как тестирование гипотезы, составляющей семейство первичных конечных точек с контролем ошибки I рода или как часть

иерархического тестирования и анализ ко-первичной конечной точки (т.е. без введения поправки на множественное сравнение). Подтверждающий анализ подгрупп целесообразен в случае равноценной клинической значимости субпопуляции пациентов и общей популяции, что в случае подтверждения эффективности исследуемого препарата в подгруппе позволит зарегистрировать его для применения в данной субпопуляции (в случае иерархического тестирования и ко-первичной конечной точки — при обязательном доказательстве эффективности в общей популяции; в случае конечной точки, составляющей семейство первичных конечных точек, достаточно доказательства эффективности только в ключевой подгруппе). Вопросы дизайна КИ и статистического анализа в случае наличия множественных сравнений описаны в проектах руководств FDA и ЕМА⁵.

Спланированный еще на этапе разработки проекта протокола подтверждающий анализ подгрупп особенно актуален в случае сложных адаптивных дизайнов с промежуточными анализами эффективности, таких как двухэтапные и бесшовные дизайны КИ фазы II/III. Для КИ указанных дизайнов, особенно распространенных в онкологии, подтверждение эффективности в подгруппе пациентов, определенной по экспрессии биомаркера / гистологическому типу опухоли / линиям предшествующей терапии, позволяет оптимизировать процесс клинической разработки препарата и ускорить процесс его регистрации для применения пациентами, которые могут получить наибольшую пользу от лечения. Обсуждение концепции и статистических аспектов мультипопуляционных КИ представлено в обзоре В.А. Millen и соавт. [3].

Таким образом, если ожидается значительное взаимодействие идентификационных факторов ключевых подгрупп и исследуемой терапии на этапе планирования КИ, то статистическая мощность исследования должна быть достаточной для выполнения подгруппа-специфических статистических тестов (то есть выявления терапевтического эффекта в ключевых подгруппах). При выявлении значительных различий терапевтического эффекта в подгруппах рекомендуется провести анализ методом контрастов по разности и знаку величины эффекта. Если эффекты в подгруппах различаются по знаку, следует провести сравнение средних значений эффектов в подгруппах методом наименьших квадратов.

В случае поискового анализа подгрупп как части оценки соотношения «польза — риск» в контексте согласованности эффекта лечения в отчете о проведенном КИ первоначально приводится описательная статистика результатов оценки эффективности по подгруппам, заранее указанным в протоколе КИ (но при этом не обязательно ограничивается только

⁵ Multiple endpoints in clinical trials. Guidance for Industry. Draft guidance. FDA; 2017.

Guideline on multiplicity issues in clinical trials. Draft. EMA/CHMP/44762/2017. EMA; 2016.

ими). Основным способом представления результатов является график типа «форест-плот», основным статистическим методом анализа — тесты взаимодействия.

Статистические взаимодействия зависят от используемой шкалы и статистической модели. Взаимодействие в моделях линейной регрессии представляет собой отклонение от аддитивности (различия в эффекте лечения по абсолютной шкале), в то время как взаимодействие в модели логистической регрессии и/или регрессии Кокса является отклонением от мультипликативной модели (различия в эффекте лечения по относительной шкале). Рекомендуется оценивать взаимодействие «подгруппа*терапия» (interaction tests) и сами эффекты лечения в подгруппе по изначально выбранной шкале оценки первичной конечной точки. Дополнительно может быть проведен анализ для ковариат или подгрупп с выявленной несогласованностью терапевтического эффекта по другой шкале оценки конечной точки (в том числе для изучения зависимости несогласованности от используемой шкалы).

Уровень статистической значимости теста взаимодействия p может быть $\geq 0,05$ при реализации подхода «генерирования сигналов» в контексте оценки согласованности эффекта лечения; при проведении анализа поправка на множественность не рекомендуется. Так или иначе, требуется дальнейший анализ данных при обнаружении «сигналов» или статистически значимого взаимодействия с позиции не только статистики, но и клинической интерпретации (принимая во внимание результаты по вторичным конечным точкам эффективности и профиль безопасности в исследуемых подгруппах).

Используемые для целей поискового анализа подгрупп статистические подходы имеют свои преимущества и недостатки. Так, например, тесты взаимодействия «подгруппа*терапия» приводят к ложноположительным и ложноотрицательным результатам вследствие низкой мощности теста, в том числе по причине различающихся по численности сравниваемых подгрупп. Вероятность того, что терапевтический эффект исследуемого препарата будет ниже, чем в группе сравнения, при его положительных статистически значимых результатах на общей популяции стремительно возрастает при увеличении числа подгрупп для анализа и уровней факторов идентификации подгрупп (в особенности при неравномерном распределении пациентов по уровням фактора и небольшом размере эффекта). Поэтому для реализации идеи «генерирования сигналов» предлагается проводить тестирование статистической гипотезы об отсутствии гетерогенности эффекта лечения в подгруппах при номинальном уровне значимости $>5\%$ (приоритет мощности над величиной ошибки I рода).

Для отображения эффекта лечения в подгруппах широко используются графики типа «форест-плот», по которым путем визуального сравнения можно выявить подгруппы с менее или, наоборот, более выраженным эффектом. Однако подобные графики строятся без учета корреляции между идентификационными факторами перекрывающихся подгрупп и без учета влияния конфаундинг-факторов.

Методы, основанные на байесовской интерпретации вероятности, могут применяться в поисковом анализе заранее определенных подгрупп как при изучении согласованности эффекта лечения в популяции КИ, так и для определения размера выборки планируемого КИ, целью которого является подтверждение наличия выраженного терапевтического эффекта, ранее обнаруженного в подгруппе пациентов при проведении поискового анализа в другом КИ. Как правило, байесовская оценка эффекта лечения в подгруппе более точна, чем выборочная оценка для той же подгруппы. Среди байесовских методов можно выделить методы оценки усадки (shrinkage estimation), применяемые в байесовском иерархическом моделировании [4, 5]. В качестве основного предположения в нем выступает вероятностная перестановочность априорного распределения подгруппа-специфических эффектов лечения, что, однако, не всегда является корректным допущением. При выборе указанных методов в целях поискового анализа подгрупп в подтверждающих КИ следует критически оценивать исходные допущения, проводить оценку адекватности созданной статистической модели. Методы оценки усадки не являются основанием для игнорирования результатов в подгруппах, в которых эффект лечения меньше его среднего значения в популяции КИ в целом. При обнаружении таких результатов следует проверять их на достоверность⁶. Кроме того, методы оценки усадки не подходят для анализа данных по безопасности в подгруппах пациентов.

Таким образом, ключевыми задачами при интерпретации анализа подгрупп являются:

- обнаружение статистической значимости различий терапевтического эффекта между подгруппами — значимые тесты на взаимодействие;
- оценка распределения ковариат в исследуемых подгруппах;
- оценка возможного влияния сопутствующих факторов, способных повлиять на конечные результаты КИ (конфаундинг-факторы);
- итоговая оценка величины эффекта (правдоподобный эффект / возможный эффект / отсутствие терапевтического эффекта).

В таблице 1 в качестве примера представлены некоторые статистические методы, используемые при анализе подгрупп в подтверждающих КИ (подтверждающий и поисковый анализы подгрупп).

⁶ Guideline on the investigation of subgroups in confirmatory clinical trials. EMA/CHMP/539146/2013. EMA; 2019.

Таблица 1. Статистические методы, используемые при анализе подгрупп

Table 1. Statistical methods used in subgroup analysis

Типы методов Types of methods	Реализация Implementation
Проверка распределения ковариат/корректировка Covariate distribution/ covariate adjustment	<p>Сводная исходная статистика по основным ковариатам. Сравнение средних значений или медиан непрерывных ковариат и распределение категориальных ковариат между пациентами (например, стандартизированное различие). Поправка на прогностические ковариаты может привести к увеличению мощности, поправка на остальные ковариаты может привести к увеличению стандартной ошибки среднего (SE) и, следовательно, снижению мощности. Корректировка может привести к инфляции ошибки I рода также в случае небольшой выборки и бинарной первичной переменной или времени-до-события [6]. Некоторые подходы к выбору ковариат в регрессионных моделях: прямой отбор, обратное исключение, пошаговая регрессия, критерии отбора моделей на основании информационных критериев (информационный критерий Акаике (AIC), байесовский информационный критерий (BIC), скорректированный коэффициент детерминации, Ср-критерий Мэллоу) и др. Для обобщенных линейных моделей или нелинейных моделей скорректированные на ковариату оценки эффекта лечения следует сравнивать с нескорректированными оценками. Для сравнения распределений непрерывных исходных ковариат между подгруппами можно использовать графические методы: диаграммы размаха, квантиль-квантиль графики рассеяния, кумулятивные функции распределения, непараметрические эмпирические функции распределения Baseline summaries with respect to main covariates. Comparison of the means or medians of continuous covariates and the distribution of their categorical counterparts between subjects (for example, standardised difference). Adjustment for prognostic covariates can lead to increased power, adjustment for nonprognostic covariates can lead to increased standard error, and thus a decrease in power. Covariate adjustment can lead to inflated type I error rates when there is a small sample size and a binary or time-to-event outcome [6]. Some examples of covariate selection methods in regression models include but are not limited to: forward selection, backward elimination, stepwise regression, criterion-based procedures (Akaike Information Criterion, Bayesian Information Criterion, adjusted R², Mallow's Cp Statistic), and others methods. For generalised linear models or non-linear models, adjusted estimates of the treatment effect should be compared to unadjusted estimates. Graphical methods: side-by-side boxplots, Q-Q plots, cumulative distribution functions, empirical nonparametric density plots can be used to compare the distribution of continuous baseline covariates between subgroups</p>
Тесты взаимодействия Statistical tests for interaction	<p>Включение взаимодействия между факторами в типичные методы анализа (ANOVA, ANCOVA, регрессионный анализ для непрерывных данных и модели продольного анализа для динамических данных). В большинстве случаев взаимодействие «терапия*подгруппа» можно представить с одной степенью свободы с указанием PE и CI. Для такого взаимодействия однофакторная ANOVA с 4 группами (например, два вида терапии по двум регионам или двум уровням фактора стратификации) позволит оценить PE и оценку общей дисперсии (и затем методом контрастов оценить PE и CI для взаимодействия) [7]. Графические методы: график взаимодействия (отображает аппроксимированные значения зависимой переменной по оси Y, по оси X отображаются значения независимой переменной) Inclusion of interaction terms in the typical methods of analysis (ANOVA, ANCOVA, regression for continuous data and longitudinal analytical models for data collected over time). In most cases, the treatment-by-subgroup interaction can be reduced to a 1-degree of freedom setting with PE and CI. For such interaction, 1-way ANOVA with 4 groups (e.g., 2 treatments by 2 regions, or 2 levels of the stratification factor) provides both PEs and a common variance estimate. Then, using the contrast statement, the PE and CIs for this interaction can be constructed [7]. Graphical methods: interaction plot (displays the fitted values of the dependent variable on the y-axis, while the x-axis shows the values of the first independent variable)</p>
Поправки на множественное сравнение Multiplicity adjustment methods	<p>Спектр возможных вариантов колеблется от базовых непараметрических множественных проверок гипотез (например, поправки на основе процедуры Бонферрони) до более сложных модельных процедур, например параметрических (включая параметрические цепные процедуры и процедуры обратной связи) [8] Available options range from basic nonparametric multiple testing procedures (for example, Bonferroni-based procedures), to more sophisticated model-based multiple testing procedures, for example, parametric procedures (including parametric chain procedures and feedback procedures) [8]</p>
Отображение эффекта лечения в подгруппах Visualisation of treatment effect in subgroups	<p>Описательная статистика (PE, SE). Графические методы: график типа «форест-плот», график Гэлбрейта Descriptive statistics (PE, SE). Graphical methods: forest plot, Galbraith plot</p>

Примечание. SE — стандартная ошибка, PE — точечная оценка, CI — доверительные интервалы, ANOVA — дисперсионный анализ, ANCOVA — ковариационный анализ.

Note. SE—standard error, PE—point estimate, CI—confidence interval, ANOVA—analysis of variance, ANCOVA—analysis of covariance.

В руководстве ЕМА⁷ подчеркивается, что заключение о наблюдаемой неоднородности эффекта лечения и принятие дальнейших регуляторных решений не могут базироваться исключительно на статистических выводах (в частности, репортирование отдельного *p*-значения из теста взаимодействия не может использоваться в качестве единственного аргумента при принятии таких решений) и требует обязательной верификации с использованием релевантной научно-медицинской информации. В большинстве случаев результаты анализа подгрупп *post hoc* могут быть использованы только для построения гипотез будущих исследований (в особенности если речь идет о КИ, в которых основная статистическая гипотеза не была доказана).

ОЦЕНКА РЕЗУЛЬТАТОВ ПРОВЕДЕННОГО АНАЛИЗА ПОДГРУПП С ТОЧКИ ЗРЕНИЯ РЕГУЛЯТОРНОГО ОРГАНА

Оценка результатов проведенного анализа подгрупп подтверждающего КИ при проведении экспертизы клинических данных регистрационного досье может быть различной в зависимости от статистической и клинической значимости полученных результатов.

Вариант 1: достигнута статистическая значимость результатов и продемонстрирована терапевтическая эффективность в первичной популяции для анализа⁸. Изучение согласованности эффектов лечения в различных субпопуляциях должно включать анализ подгрупп в соответствии с методами, описанными выше. Если анализ ключевых подгрупп был должным образом запланирован, в ходе КИ не были допущены отклонения, ставящие под сомнение научную правдоподобность анализа, и отсутствуют доказательства противоречивости результатов, то анализ подгрупп можно считать завершенным.

Несо согласованные или экстремальные значения по целевым переменным в поисковых подгруппах, где эксперт может подтвердить отсутствие правдоподобной связи ответа и терапии, обычно могут быть проигнорированы при условии, что такие результаты не повторяются более чем в одном исследовании. Если же экстремальные результаты воспроизводятся более чем в одном исследовании, для таких ситуаций оценка правдоподобия должна быть пересмотрена.

В случае если обсуждение полученных результатов и предварительная спецификация ключевых подгрупп были неполными, эксперт будет вынужден больше полагаться на приведенные данные, опираясь на *ad hoc* подход и собственное суждение о правдоподобности.

Несо согласованность терапевтических эффектов в подгруппе и в общей популяции исследования

может считаться вероятной и, следовательно, подлежать дальнейшей оценке спонсором исследований и регуляторными органами в случае наличия:

а) биологических предпосылок неравномерности эффекта в ожидаемом направлении (такая вероятность особенно велика, если доказательства встречаются в нескольких различных источниках, например в научной литературе или данных других проведенных КИ). В случае наличия в материалах регистрационного досье результатов только одного КИ, в котором должным образом оценена конкретная подгруппа и обнаружен несогласованный терапевтический эффект, последний не может быть проигнорирован только по причине отсутствия репликации в рамках КИ фазы III;

б) фактов обнаружения неравномерного эффекта в нескольких различных источниках данных (такая вероятность особенно велика, если есть и биологическая правдоподобность наблюдаемого неравномерного эффекта).

Правдоподобность неравномерности эффекта между подгруппами дополнительно подтверждается, если тесты взаимодействия являются статистически значимыми или погранично значимыми, а также если существуют некоторые свидетельства взаимодействия типа «терапия*ковариата» для различных конечных точек, охватывающих различные аспекты эффективности лечения и не связанных напрямую.

Вариант 2: продемонстрирована статистическая значимость результатов в первичной популяции анализа, но терапевтическая эффективность или соотношение «польза — риск» неубедительны⁹. Формальное доказательство эффективности (статистическая значимость) имеет первостепенное значение для разработки новых лекарственных средств. Тем не менее, даже если результаты анализа статистически достоверны, клиническая значимость эффекта лечения во всей популяции исследования или в отдельной субпопуляции может быть неубедительной. Для доказательства терапевтической эффективности также необходимо подтвердить благоприятное соотношение «польза — риск». Даже в случаях, когда фактические данные об эффективности убедительны со статистической точки зрения, целесообразно идентифицировать подгруппу, которая не была предварительно указана в качестве части стратегии подтверждающего тестирования, но для которой эффективность препарата и соотношение «польза — риск» были бы убедительны. Основные сценарии, при которых это может произойти:

1) польза фармакотерапии по всей популяции исследования является статистически значимой, но неубедительной с клинической точки зрения;

⁷ Там же.

⁸ Там же.

⁹ Там же.

2) польза фармакотерапии в популяции всех рандомизированных пациентов является статистически и клинически значимой, но существуют риски и неопределенности, которые не позволяют сделать вывод, что ожидаемая польза превышает возможный риск во всей популяции исследования;

3) польза фармакотерапии в популяции всех рандомизированных пациентов является статистически и клинически значимой, но существуют риски и неопределенности для отдельной субпопуляции исследования, которые не позволяют сделать вывод, что ожидаемая польза для нее превышает возможный риск.

При таких сценариях возникает не только проблема множественных сравнений, но и проблема предвзятости выбора подгруппы для оценки, поскольку идентификация интересующей подгруппы происходит уже после того, как станут известны результаты КИ. Таким образом, эксперт и заявитель будут обращать внимание на наиболее экстремальные результаты, в связи с чем оцениваемый эффект лечения и связанные с ним доказательства не обязательно окажутся достоверными. Для того чтобы исследуемая подгруппа считалась правомерной для оценки, обычно необходимо соблюдение всех ниже приведенных критериев. Указанные критерии применяются независимо от того, кто инициировал дополнительные исследования, — спонсор или регуляторный орган:

1) существуют внешние источники доказательств того, что исследуемая подгруппа является четко определенной и клинически значимой субпопуляцией пациентов. Обычно предполагается, что соответствующая подгруппа была определена на этапе планирования КИ (например, в результате стратифицированной рандомизации или выделения ее среди ключевых подгрупп) и приведена аргументация ее клинической релевантности;

2) существует фармакологическое обоснование причин различной эффективности (или соотношения «польза — риск») исследуемого препарата в субпопуляции и общей популяции пациентов (во внимание принимают также масштаб выявленных отличий и используемую шкалу оценки);

3) эффект лечения, наблюдаемый в подгруппе, обычно больше, чем во всей рандомизированной популяции исследования. Совокупность статистических доказательств, основанных на отдельных КИ и объединенных анализах, должна соответствовать тем же стандартам доказательств, которые обычно предъявляются для всей рандомизированной популяции, что указывает на выраженность эффекта в подгруппе по сравнению с имеющейся фоновой вариабельностью явления;

4) результаты анализа подгруппы воспроизводятся в других КИ (как проведенных спонсором, так и внешних КИ). Особые затруднения возникают

при рассмотрении заявки на регистрацию препарата, доказательства эффективности и безопасности применения которого основаны на результатах одного ключевого КИ, поскольку основным признаком достоверности выводов является идентичность результатов, полученных в различных КИ. В случае проведения только одного КИ биологические предпосылки и полученные клинические данные для исследуемой подгруппы должны быть исключительно убедительными;

5) в случае, когда рекомендации по лечению основаны на анализе подгруппы, необходимо провести тщательный анализ соотношения «польза — риск» в данной подгруппе, а также удостовериться в применимости данных по безопасности, полученных для популяции всех рандомизированных пациентов. Редкие серьезные нежелательные явления могут случайным образом не встречаться в интересующей подгруппе в рамках проведенного КИ, но должны быть учтены при оценке соотношения «польза — риск», если не обосновано иное.

Необходима тщательная проверка исходных клинических характеристик пациентов из исследуемой подгруппы для сравнения терапевтического эффекта в подгруппе между группами лечения. Может потребоваться корректировка различий по исходным клиническим характеристикам, чтобы гарантировать, что наблюдаемый в подгруппе пациентов терапевтический эффект исследуемого препарата не вызван или не усиливается из-за дисбаланса между группами лечения.

При соблюдении всех вышеперечисленных требований регистрация исследуемого препарата может быть ограничена показаниями к применению пациентами только из подгруппы КИ с выраженным терапевтическим эффектом; если же существенные различия между величиной эффекта лечения или соотношением «польза — риск» во всем КИ по-прежнему не будут объяснены, то регистрация препарата может оказаться невозможной.

Пример (гетерогенность эффекта лечения между подгруппами). В международном КИ фазы III PLATO, проведенном с участием более чем 18 000 пациентов с острым коронарным синдромом, было продемонстрировано преимущество тикагрелора перед клопидогрелом в общей популяции в отношении первичной комбинированной конечной точки «смерть от сосудистых причин / инфаркт миокарда / инсульт» (отношение риска (HR) 0,84) [9]. Однако по результатам предусмотренного протоколом КИ анализа подгрупп была выявлена гетерогенность результатов по эффективности для когорты пациентов из региона Северная Америка, составляющей ~10% от общей популяции КИ (HR 1,25), при этом для пациентов из других географических регионов были получены согласованные результаты (HR 0,80–0,86). Несмотря на то что анализ подгрупп в этом КИ носил

исключительно описательный характер (для идентификации подгрупп кроме географического было заранее выбрано более тридцати иных факторов и статистических поправок в связи с этим введено не было) и статистическая значимость теста взаимодействия «терапия*географический регион» составила немногим менее 5% ($p = 0,045$) — FDA первоначально не одобрило применение тикагрелора. Позднее независимыми статистическими группами был проведен дополнительный анализ данных для количественной оценки того, какая часть взаимодействия «терапия*регион» может быть объяснена клиническими характеристиками пациентов и сопутствующим лечением, включая поддерживающую терапию аспирином. В США больше пациентов (53,6%), чем в остальных странах (1,7%), принимали аспирин в средней дозе ≥ 300 мг/сут. Методом скорректированного по средней поддерживающей дозе аспирина регрессионного анализа Кокса и регрессионного анализа в восьми ориентирных временных точках (landmark analyses) было показано, что у пациентов, принимающих аспирин в низких дозах (≤ 100 мг/сут), тикагрелор ассоциировался с лучшими результатами по сравнению с клопидогрелом (HR 0,77). Однако по результатам статистического моделирования на основании частоты клинических исходов в разных странах и распределения пациентов было показано, что наблюдаемая структура результатов могла возникнуть случайным образом. Учитывая также отсутствие четкого биологического обоснования того, почему тикагрелор может быть менее эффективным, чем клопидогрел в сочетании с высокой поддерживающей дозой аспирина, был сделан вывод, что наблюдаемая неоднородность эффекта лечения могла быть случайной из-за большого количества нескорректированных анализов подгрупп, и препарат в итоге получил одобрение регуляторных органов США — с указанием, что суточная поддерживающая доза аспирина должна составлять 75–100 мг.

Вариант 3: статистическая значимость результатов в первичной популяции анализа не была достигнута, но существует заинтересованность в определении подгруппы, в которой можно оценить соответствующий эффект лечения и найти убедительные доказательства благоприятного соотношения «польза — риск»¹⁰. Подобная ситуация возникает в случае, если основная цель КИ не была достигнута, исследование потерпело неудачу (невозможность отклонить нулевую гипотезу, что обычно классифицируется как двусторонний $p > 0,05$). С формальной статистической точки зрения дальнейшее подтверждающее тестирование по дополнительным задачам КИ недопустимо, так как первичная статистическая гипотеза не была доказана.

Как правило, в таких ситуациях требуется проведение дополнительных КИ. В редких случаях

возможно одобрение препарата регуляторными органами без проведения дополнительных исследований. Основанием такого одобрения может быть невозможность повторного проведения КИ в больших популяциях (например, КИ препаратов для лечения сердечно-сосудистых заболеваний). В этом случае даже субпопуляции рандомизированных пациентов могут представлять значительное количество надежных доказательств для оценки эффективности лечения и соотношения «польза — риск».

Необходима тщательная оценка всех имеющихся доказательств, и спонсорам следует иметь в виду, что вышеописанный сценарий в большинстве случаев будет рассматриваться как не отвечающий предъявляемым требованиям для регистрации препарата. Если, тем не менее, в исключительных случаях рассматривается возможность вынесения положительного решения о регистрации, то должно быть представлено четкое обоснование того, почему должным образом спланированное КИ потерпело неудачу, несмотря на то что препарат считается эффективным, и почему проведение дополнительных проспективных КИ для установления формального доказательства эффективности неосуществимо или неоправданно.

Пример (препарат, не отвечающий предъявляемым требованиям). Исследования применения препарата адуканумаб по показанию «болезнь Альцгеймера» потерпели неудачу по результатам промежуточного анализа бесполезности в двух идентичных по дизайну КИ III фазы ENGAGE и EMERGE — соответствие критериям бесполезности $< 20\%$ условной мощности. Однако по окончании этих КИ компания Biogen подала заявку в FDA на получение маркетингового одобрения препарата на основании *post hoc* анализа — были повторно проанализированы данные КИ с дополнительным включением пациентов (12 и 18% от запланированного количества соответственно), продолживших участие в исследовании после даты проведения анализа бесполезности. В результате в КИ ENGAGE статистическая значимость по первичной переменной эффективности CDR-SB в отношении превосходства над плацебо не была достигнута в группе пациентов, получавших препарат в высокой дозе (6% ухудшение по сравнению с плацебо, $p = 0,627$), но в КИ EMERGE продемонстрировано улучшение на 23% по сравнению с плацебо ($p = 0,031$). Также было проведено сравнение с группами плацебо (численностью пациентов $n = 545$ и 548 соответственно) подгруппы пациентов ($n = 116$ и 147 соответственно), получавших адуканумаб в высокой дозе 10 мг/кг в течение 78 нед. исследования и не имеющих значительных отклонений от протокола [10]. Данный случай вызвал резкую критику как среди представителей фармацевтической индустрии, так и академического сообщества [11]. Причины, по которым проведенный

¹⁰ Там же.

компанией анализ неправомерен: 1) анализ подгруппы, меньшей, чем планировалось численности (40% пациентов не завершили КИ EMERGE); 2) необоснованность сравнения выборки пациентов с высокой комплаентностью с более гетерогенной группой плацебо; 3) указанная подгруппа составляет только 20% от численности пациентов, получавших препарат в высокой дозе, и предположительно включает большинство пациентов с отсутствием аллели APOE ε4 (прогностический биомаркер более тяжелого течения заболевания); 4) отсутствие репликации (внутренней — КИ ENGAGE, внешней — КИ препаратов подобного класса неоднократно терпели неудачу); 5) слабая биологическая правдоподобность (этиология заболевания неизвестна, его патогенетические механизмы в динамике в рамках амилоидной гипотезы изучены недостаточно).

Пример (проведение дополнительного КИ). В КИ SYNTAX не удалось доказать основную гипотезу наименьшей эффективности чрескожного коронарного вмешательства (ЧКВ) по сравнению с аортокоронарным шунтированием (АКШ) у пациентов с ишемической болезнью сердца с трехсосудистым поражением коронарных артерий или стенозом ствола левой коронарной артерии — частота серьезных нежелательных сердечно-сосудистых или цереброваскулярных явлений в группе АКШ 12,4% vs. в группе ЧКВ 17,8%, $p = 0,002$, абсолютная разность 5,4%, верхняя граница одностороннего 95% ДИ 8,3% при границе не меньшей эффективности δ 6,6% (наличие подгрупп пациентов по локализации поражения было учтено при оценке мощности и размера выборки на этапе протокола КИ) [12]. Но проведенный *post-hoc* анализ подгрупп выявил, что для пациентов с поражением левой главной коронарной артерии (кроме пациентов с высокой анатомической сложностью по шкале SYNTAX Score) ЧКВ может являться приемлемой альтернативой АКШ [13]. В результате было проведено отдельное КИ EXCEL на указанной популяции пациентов с низкой и средней анатомической сложностью, в котором была продемонстрирована не меньшая эффективность ЧКВ АКШ в отношении комбинированной конечной точки «смерть / инфаркт миокарда / инсульт» в течение трех лет с момента проведения вмешательства (ЧКВ 15,4% vs. АКШ 14,7%, абсолютная разность 0,7%, верхняя граница одностороннего 97,5% ДИ 4,0%, $p = 0,02$ при δ 4,2%) [14]. Однако в 2019 г. возникли серьезные разногласия в кардиологическом сообществе относительно интерпретации результатов данного исследования [15, 16].

РЕКОМЕНДАЦИИ ПО АНАЛИЗУ ПОДГРУПП НА СТАДИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЯ

Полностью описать запланированный анализ подгрупп на этапе проекта протокола возможно

только в редких случаях. Несмотря на ограниченное количество информации об эффективности исследуемого препарата на этапе планирования подтверждающего КИ фазы III следует использовать всю доступную информацию для выбора оптимальной стратегии анализа данных. Поскольку наибольшее доверие вызывают результаты заранее запланированного в проекте протокола анализа подгрупп, следует уделить значительное внимание выбору подгрупп пациентов, представляющих интерес с клинической точки зрения, проводя анализ научной литературы по исследуемой нозологии. Кроме этого, еще на этапе КИ фазы II можно проводить поисковый анализ клинически значимых подгрупп¹¹.

Учет гетерогенности в целевой популяции.

Во время разработки дизайна и планирования статистического анализа КИ обсуждение известных прогностических (разграничивающих группы по разным клиническим исходам) и предиктивных (разграничивающих группы с различным ответом на лечение) факторов является одним из наиболее важных этапов подготовки исследования. На данном этапе необходимо выбрать целевую группу пациентов для КИ. В частности, следует определить, должны ли критерии включения/невключения ограничивать популяцию пациентов, например, одним уровнем определенного фактора, или же предполагается, что исследуемый препарат предназначен для всей популяции при условии, что пациенты во всех субпопуляциях, определяемых различными уровнями фактора, получают в значительной степени пользу от лечения.

КИ с включением широкого круга пациентов дает возможность регистрации препарата с расширенными показаниями к применению, но это также потребует от спонсора изучения согласованности ответа на лечение. Для дальнейшего принятия регуляторных решений по возможности рекомендуется представлять результаты одного или нескольких крупных КИ, охватывающих всю целевую популяцию пациентов для возможной оценки влияния различных факторов, модифицирующих терапевтический эффект, с целью последующего анализа отношения ожидаемой пользы к возможному риску и общей оценки прогнозируемой терапевтической эффективности исследуемого препарата. В проекте протокола КИ должна обсуждаться ожидаемая степень гетерогенности популяции пациентов с точки зрения как прогностических, так и предиктивных факторов. Ограничение популяции исследования субпопуляцией целевой популяции пациентов (то есть всеми пациентами в соответствии с предполагаемыми показаниями к применению) должно быть обосновано с учетом того, установлено ли такое ограничение из-за проблем безопасности, ожидаемой

¹¹ Там же.

недостаточной эффективности или каких-либо других соображений. Целесообразно собрать доказательства того, что исследуемый препарат не приносит пользы всем пациентам из популяции КИ (например, в случае использования предиктивных биомаркеров)¹².

Очевидно, что по мере проведения подтверждающих КИ объем данных будет увеличиваться и, следовательно, не все потенциальные источники гетерогенности могут быть предсказаны заранее. Сведения, полученные в ходе исследования, должны быть приведены на этапе анализа независимо от того, были ли они отражены в протоколе КИ или нет. Заслепленный промежуточный анализ данных (в случае его наличия в КИ) дает возможность изучить имеющиеся клинические данные и затем пересмотреть запланированный анализ подгрупп, избегая при этом предвзятости в оценке.

Стратегия выбора и определения подгрупп для оценки¹³. После тщательного выбора исследуемой популяции в КИ и исключения подгрупп пациентов, о которых имеются веские основания полагать, что терапевтический эффект лечения будет несогласованным, стратегия последующего анализа данных по окончании КИ, тем не менее, должна предусматривать доказательство правомерности предположения о согласованности эффекта. Без обсуждения и проведения дальнейших анализов предположение о согласованности эффекта лечения по всей популяции КИ, основывающееся на высокой вероятности ответа пациентов на лечение, является необоснованным.

Включенная в КИ популяция должна отражать эпидемиологию заболевания в целевой популяции пациентов (внешняя валидация). Следует рассмотреть необходимость стратификации при рандомизации (в сочетании с блоковой рандомизацией) для того, чтобы, во-первых, снизить риск несбалансированного распределения пациентов с различными уровнями факторов в группы терапии; во-вторых, выделить некоторые факторы подгрупп для анализа того, будут ли пациенты с различным профилем риска иметь одинаковую пользу от применения исследуемого препарата. В частности, необходимо изучить, являются ли известные прогностические стратификационные факторы также прогностическими для оценки эффективности и/или соотношения «польза — риск». Стратифицированная рандомизация допускает включение только ограниченного количества прогностических факторов в модель (корректировка по исходным значениям ковариат), поэтому на этапе планирования следует обсудить с врачами-исследователями потенциально важные прогностические и предиктивные факторы для исследования.

В проекте протокола должны быть отдельно выделены ключевые подгруппы (соответствующие определению фактора идентификации подгрупп 2 (нумерацию подгрупп см. выше)), а также перечислены поисковые подгруппы (соответствующие определению фактора идентификации подгрупп 3).

Четкое понимание приоритизации анализа различных подгрупп на этапе планирования КИ помогает минимизировать апостериорные суждения и способствует созданию достаточной доказательной базы для оценки клинически значимых подгрупп и снижению риска ошибочных заключений об эффективности или ее отсутствии в субпопуляциях пациентов. Важно достичь консенсуса между спонсором и регулятором по вопросам проведения анализа подгрупп еще на этапе планирования КИ. В противном случае, если в проекте протокола (и затем в подготовленном отчете о проведенном КИ) указанные вопросы были освещены в недостаточной мере, регуляторная оценка анализа подгрупп будет проводиться преимущественно в формате *post-hoc*.

В целом план первоначального анализа должен учитывать тип фактора (например, бинарный/категориальный/непрерывный) в отношении того, насколько он является прогностическим или предиктивным. Функциональная форма связи фактора с эффектом лечения (например, линейная зависимость) непрерывной переменной-ковариаты легко интерпретируется и делает возможной надлежащую классификацию подгрупп.

Если эффект лечения варьируется в зависимости от уровня конкретного фактора, последующие анализы могут включать категоризацию (для непрерывного фактора) или сжатые категории (для порядковой категориальной переменной с большим числом уровней), поскольку эти категории относятся к критериям, которые могут в конечном итоге использоваться при составлении инструкции или принятии клинических решений. Различные подходы к выполнению анализа подгрупп должны быть тщательно описаны в проекте протокола.

ПРЕДСТАВЛЕНИЕ СВЕДЕНИЙ ПО АНАЛИЗУ ПОДГРУПП В РЕГИСТРАЦИОННОМ ДОСЬЕ

При регистрации исследуемого препарата по правилам Евразийского экономического союза в экспертном отчете по оценке безопасности, эффективности и качества (далее — экспертный отчет по оценке) должны представляться результаты проведенного заявителем анализа подгрупп, включая представленное заявителем обсуждение результатов в контексте наличия гетерогенности в целевой популяции и потенциальных источников несогласованности терапевтического эффекта между субпопуляциями пациентов. Результаты анализа,

¹² Там же.

¹³ Там же.

демонстрирующие согласованность терапевтического эффекта, также должны представляться в экспертном отчете по оценке безопасности, эффективности и качества (экспертный отчет по оценке)¹⁴.

Помимо этого, в экспертном отчете по оценке приводятся выводы регуляторного органа относительно методологического качества проведенного анализа подгрупп, его научной обоснованности и клинической интерпретируемости полученных результатов¹⁵. В случае признания достоверности выявленной несогласованности терапевтического эффекта в экспертном отчете по оценке приводится соответствующее обсуждение с точки зрения клинической значимости, соотношения «польза — риск» и представления указанной информации в разделе «Общая характеристика лекарственного препарата» (ОХЛП)¹⁶. Если результаты, указывающие на несогласованность терапевтического эффекта, признаются недостоверными, то это также должно быть указано в экспертном отчете по оценке.

Представление сведений относительно согласованности терапевтического эффекта в разделе 5.1 «Фармакодинамические свойства» ОХЛП, как правило, не требуется, но может быть целесообразным при наличии важной информации для врача¹⁷. Если в ключевой подгруппе пациентов были получены неопределенные результаты по первичным переменным эффективности/безопасности, несмотря на положительные результаты в исследовании в целом, то такие сведения могут быть указаны в качестве предупреждения в разделе 4.4 «Особые указания и меры предосторожности при применении» ОХЛП вместе с данными, представленными в разделе 5.1. Результаты анализа подгрупп, указывающие на отсутствие терапевтической эффективности или неблагоприятное соотношение «польза — риск» в определенных подгруппах (в случае признания их достоверности), должны быть отражены в разделе 4.1 «Показания к применению» и/или 4.3 «Противопоказания» ОХЛП¹⁸.

Описание в экспертном отчете по оценке результатов КИ с выполненным анализом подгрупп,

соответствующим вариантам 2 или 3 сценариев оценки экспертом (о сценариях см. выше), должно содержать обстоятельное обсуждение с обоснованием принятого положительного решения о регистрации препарата по результатам такого анализа. Если целевая популяция ограничена подгруппой пациентов из популяции проведенного КИ, то результаты исследования эффективности и безопасности препарата сначала должны быть представлены в общем по всей популяции КИ, а затем отдельно по данной подгруппе в разделе 5.1. «Фармакодинамические свойства» ОХЛП. Объяснение того, почему результаты по эффективности и безопасности в данной подгруппе можно считать достоверными, могут быть представлены там же в краткой форме. Если аналогичные результаты по эффективности и безопасности для данной подгруппы повторяются сразу в двух проведенных КИ, то целесообразно представить результаты обоих КИ для описания степени согласованности результатов анализа.

ЗАКЛЮЧЕНИЕ

Описанные в работе требования отражают общие принципы планирования и проведения анализа подгрупп в подтверждающих КИ, обозначенные в международных руководствах и специализированной статистической литературе. Разработка указанных требований продиктована востребованностью анализа подгрупп и его зачастую ошибочным выполнением и некорректной интерпретацией результатов.

Представленные рекомендации целесообразно использовать как экспертам, осуществляющим экспертизу отношения ожидаемой пользы к возможному риску применения ЛП, так и фармацевтическим компаниям при подготовке протоколов и отчетов клинических исследований ЛП.

Вклад авторов. *О. И. Басова* — разработка концепции, сбор и анализ данных, написание текста; *И. В. Лыскова* — обоснование актуальности работы, литературный обзор и написание текста; *О. Ю. Иванова* — научное редактирование, написание текста.

¹⁴ Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 78 «О Правилах регистрации и экспертизы лекарственных средств для медицинского применения». Приложение № 16. Форма Экспертного отчета об оценке безопасности, эффективности и качества (IV. Оценка соотношения пользы и риска).

¹⁵ Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 78 «О Правилах регистрации и экспертизы лекарственных средств для медицинского применения». Приложение № 15. Указания по составлению экспертного отчета об оценке клинических исследований (3. Клиническая эффективность, 3.4. Основные исследования, 3.4.1.8. Статистические методы и 3.4.2.7. Дополнительные анализы).

¹⁶ Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 78 «О Правилах регистрации и экспертизы лекарственных средств для медицинского применения». Приложение № 15. Указания по составлению экспертного отчета об оценке клинических исследований (3. Клиническая эффективность, 3.8. Общее заключение эксперта по оценке клинической эффективности, 3.8.3. Данные об эффективности и дополнительные анализы).

¹⁷ Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 88 «Об утверждении требований к инструкции по медицинскому применению лекарственных препаратов и общей характеристике лекарственных препаратов для медицинского применения» («Требования к инструкции по медицинскому применению лекарственного препарата и общей характеристике лекарственного препарата для медицинского применения», III. Разделы ОХЛП, 5. Фармакологические свойства).

¹⁸ Решение Совета Евразийской экономической комиссии от 3 ноября 2016 г. № 88 «Об утверждении требований к инструкции по медицинскому применению лекарственных препаратов и общей характеристике лекарственных препаратов для медицинского применения» («Требования к инструкции по медицинскому применению лекарственного препарата и общей характеристике лекарственного препарата для медицинского применения», III. Разделы ОХЛП, 4. Клинические данные).

Authors' contributions. *Olga I. Basova*—elaboration of the study concept, collection and analysis of data, writing of the text; *Irina V. Lysikova*—literature review, writing of the paper and justification of its relevance; *Olga Yu. Ivanova*—writing of the text and scientific editing of the paper.

Благодарности. Работа выполнена в рамках государственного задания ФГБУ «НЦЭСМП» Минздрава России № 056-00005-21-00 на проведение прикладных научных исследований (номер государственного учета НИИ 121021800098-4).

Acknowledgements. The study reported in this publication was carried out as part of a publicly funded research project No. 056-00005-21-00 and was supported by the Scientific Centre for Expert Evaluation of Medicinal Products (R&D public accounting No. 121021800098-4).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

Conflict of interest. The authors declare no conflict of interest requiring disclosure in this article.

ЛИТЕРАТУРА / REFERENCES

1. Alosch M, Fritsch K, Huque M, Mahjoob K, Pennello G, Rothmann M, et al. Statistical considerations on subgroup analysis in clinical trials. *Stat Biopharm Res.* 2015;7(4):286–303. <https://doi.org/10.1080/19466315.2015.1077726>
2. Dane A, Spencer A, Rosenkranz G, Lipkovich I, Parke T, PSI/EFSP Working Group on Subgroup Analysis. Subgroup analysis and interpretation for phase 3 confirmatory trials: White paper of the EFSP/PSI working group on subgroup analysis. *Pharm Stat.* 2019;18(2):126–39. <https://doi.org/10.1002/pst.1919>
3. Millen BA, Dmitrienko A, Ruberg S, Shen L. A statistical framework for decision making in confirmatory multipopulation tailoring clinical trials. *Drug Inf J.* 2012;46(6):647–56. <https://doi.org/10.1177/0092861512454116>
4. Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. *Clin Trials.* 2011;8(2):129–43. <https://doi.org/10.1177/1740774510396933>
5. Simon R. Bayesian subset analysis: application to studying treatment-by-gender interactions. *Stat Med.* 2002;21(19):2909–16. <https://doi.org/10.1002/sim.1295>
6. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials.* 2014;15:139. <https://doi.org/10.1186/1745-6215-15-139>
7. Ting N. Statistical interactions in a clinical trial. *Ther Innov Regul Sci.* 2018;52(1):14–21. <https://doi.org/10.1177/2168479017716491>
8. Dmitrienko A, Millen B, Lipkovich I. Multiplicity considerations in subgroup analysis. *Stat Med.* 2017;36(28):4446–54. <https://doi.org/10.1002/sim.7416>
9. Mahaffey KW, Wojdyla DM, Carroll K, Becker RC, Storey RF, Angiolillo DJ, et al. Ticagrelor compared with clopidogrel by geographic region in the Platelet Inhibition and Patient Outcomes (PLATO) trial. *Circulation.* 2011;124(5):544–54. <https://doi.org/10.1161/circulationaha.111.047498>
10. Schneider L. A resurrection of aducanumab for Alzheimer's disease. *Lancet Neurol.* 2020;19(2):111–2. [https://doi.org/10.1016/s1474-4422\(19\)30480-6](https://doi.org/10.1016/s1474-4422(19)30480-6)
11. Carroll J. In a stunning turnaround, Biogen says that aducanumab does work for Alzheimer's — but data mining incites controversy and questions. *Endpointsnews.* 2019. October 22. <https://endpts.com/in-a-stunning-turnaround-biogen-says-that-aducanumab-does-work-for-alzheimers-and-theyre-prepping-a-pitch-to-the-fda/>
12. Serruys PW, Morice MC, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al. Percutaneous coronary intervention versus coronary artery bypass grafting for severe coronary artery disease. *N Engl J Med.* 2009;360(10):961–72. <https://doi.org/10.1056/nejmoa0804626>
13. Pocock SJ, Stone GW. The primary outcome fails — what next? *N Engl J Med.* 2016;375(9):861–70. <https://doi.org/10.1056/nejmra1510064>
14. Stone GW, Sabik JF, Serruys PW, Simonton CA, Généreux P, Puskas J, et al. Everolimus-eluting stents or bypass surgery for left main coronary artery disease. *N Engl J Med.* 2016;375(23):2223–35. <https://doi.org/10.1056/nejmoa1610227>
15. Fornell D. EXCEL trial authors say european surgery society claims of bad science are groundless. *DAIC.* 2020. January 2. <https://www.dicardiology.com/content/excel-trial-authors-say-european-surgery-society-claims-bad-science-are-groundless>
16. O'Riordan M. Former EXCEL investigator alleges trial manipulation, prompting vehement denials. *TCTMD.* 2019. October 7. <https://www.tctmd.com/news/former-excel-investigator-alleges-trial-manipulation-prompting-vehement-denials>

ОБ АВТОРАХ / AUTHORS

Басова Ольга Игоревна. *Olga I. Basova.* ORCID: <https://orcid.org/0000-0002-6300-2092>

Лыскова Ирина Викторовна, канд. мед. наук. *Irina V. Lysikova, Cand. Sci. (Med.).* ORCID: <https://orcid.org/0000-0002-7032-5957>

Иванова Ольга Юрьевна. *Olga Yu. Ivanova.* ORCID: <https://orcid.org/0000-0003-1698-2670>

Статья поступила 11.09.2020

После доработки 10.02.2021

Принята к печати 31.05.2021

Article was received 11 September 2020

Revised 10 February 2021

Accepted for publication 31 May 2021