

Адаптивный дизайн в клинических исследованиях: преимущества и риски

Бондарева И.Б.

Федеральное государственное бюджетное учреждение «Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства»

Резюме. В последние десятилетия постепенно становится всё более популярным использование методов адаптивного дизайна как средства повышения эффективности ранних и поздних фаз клинических исследований, в частности, путём снижения стоимости, ресурсов и длительности, а также повышения вероятности успеха исследования и получения пациентами более эффективной и безопасной терапии. Проспективно запланированным модификациям могут подвергаться: критерии отбора, размер выборки и/или длительность исследования, параметры рандомизационной процедуры, режимы дозирования, включая число терапевтических групп, сопутствующая терапия, схема оценки состояния пациента, конечные точки, статистические методы. Возможность изменения дизайна в ходе исследования по результатам промежуточного анализа имеет ряд преимуществ перед традиционными исследованиями с фиксированным дизайном. Тем не менее, адаптивные исследования более сложные в исполнении, требуют более тщательного планирования, могут сопровождаться в различной степени статистическими, операционными, логистическими и регуляторными проблемами. Целью этой статьи является обзор основных определений, анализ “за” и “против” применения адаптивного дизайна на разных фазах клинических исследований со статистической и регуляторной точек зрения для лучшего понимания методологии.

Ключевые слова: адаптивные конструкции, гибкие конструкции, групповые последовательные методы, повторная оценка размера выборки, прекращение бесперспективности, промежуточный анализ, операционный уклон, статистический уклон

Adaptive design in clinical trials: benefits and risks

Bondareva I.B.

Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency

Abstract. Over the past several decades, adaptive design methods in clinical trials has become more popular as a way to increase the efficiency of the early and late phases of clinical development by means of reducing costs, resources and duration as well as increasing the likelihood of trial success and the chance for patients of being allocated to the treatment that's more effective and safe. The following prospectively planned modifications can be implemented in an adaptive study: eligibility criteria, sample size and/or study duration, randomization procedure, treatment regimens, including the number of treatment groups, concomitant medication, patient evaluation schedule, endpoints, statistical methods. Opportunity to change some elements of the design based on interim results during the course of the study provides advantages over the traditional fixed design. Nevertheless, the implementation of adaptive designs is more complex, such trials require more thorough planning, and are often accompanied with various degrees of statistical, procedural, logistic and regulatory issues. The aim of this article is to provide the main definitions, to highlight the pros, cons of adaptive design at different phases of drug development from statistical and regulatory point of view for better understanding of the methodology.

Keywords: adaptive designs, flexible designs, group sequential methods, sample size re-estimation, futility stopping, interim analyses, operational bias, statistical bias

Автор, ответственный за переписку:

Бондарева Ирина Борисовна — в.н.с, д.б.н.; Федеральное государственное бюджетное учреждение «Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства»; 119435, г. Москва ул. Малая Пироговская дом 1а; тел.: +7 (985) 924-12-06; e-mail: i_bondareva@yahoo.com

Введение

В последние десятилетия использование методов адаптивного дизайна в клинических исследованиях постепенно становится всё более популярным. Концепция адаптивного дизайна прослеживается с 70 годов XX века, когда в клинических исследованиях

начали использовать адаптивную рандомизацию и групповые последовательные методы. Методы адаптивного дизайна в клинических исследованиях привлекательны по разным причинам. Они отражают реальную клиническую практику, уменьшают вероятность получения пациентами неэффективной и/или высокотоксичной терапии, являются не

только гибкими, но и эффективными на всех фазах клинических исследований. Тем не менее, применять их нужно с осторожностью, поскольку методологические ошибки могут привести к некорректным *p*-значениям (доверительным интервалам) в отношении эффекта терапии, а значит, к некорректным выводам об эффективности изучаемой терапии и к невозможности интерпретации полученных результатов [1—3]. Адаптивный дизайн развивался как средство повышения эффективности клинических исследований, потенциально увеличивающее пользу для участников исследования и будущих пациентов путём снижения стоимостей, и повышения вероятности нахождения эффекта и действительного преимущества изучаемого лечения, если оно существует. В литературе появлялось всё больше работ, посвящённых теоретическим и регуляторным аспектам адаптивного дизайна в клинических исследованиях, а также рекомендациям по практическому использованию этих методов [4—10]. Тем не менее, достаточно долго среди специалистов не было консенсуса ни относительно единообразного использования терминов, ни относительно методологии, ни относительно отношения регуляторов к практическому использованию различных адаптивных методов и процедур. До появления регуляторных документов наиболее часто используемыми методами адаптивного дизайна были: адаптивная рандомизация, групповые последовательные методы с гибкими критериями для ранней остановки исследования по причинам безопасности, бесполезности (futility), эффективности, а также пересчёта размера выборки на промежуточном анализе для достижения желаемой статистической мощности [11—23].

Первый регуляторный документ, касающийся методологии применения адаптивного дизайна в клинических исследованиях, был выпущен в 2007 г. Комитетом по лекарственным средствам для медицинского применения Европейского медицинского агентства (*Committee for Human Medicinal Products (CHMP) European Medicines Agency, EMA*) [24]. В 2010 вышел документ Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США (*Food and Drug Administration, FDA*) [25]. Статьи, впервые представившие концепцию адаптивного дизайна, были опубликованы раньше этих регуляторных документов, но после появления рекомендаций тема адаптивного дизайна и его практического применения стала ещё более популярной в фармацевтической индустрии, академических исследованиях и среди регуляторов. Эти документы дали определения адаптивного дизайна, а главное, отразили позицию регуляторных органов — использование возможно, но с осторожностью. Между этими документами нет принципиальных разногласий. Документ FDA, более детальный, охватывает разные аспекты, связанные с практическим примени-

ем адаптивного дизайна на разных фазах клинических исследованиях. Документ EMA касается применения адаптивного дизайна в подтверждающих (confirmatory) исследованиях. В сравнении с документом FDA документ EMA менее обнадеживающий в плане применимости адаптивного дизайна в подтверждающих клинических исследованиях, но допускает его использование при корректном планировании, проведении и анализе. Авторы документа EMA предупреждают о недопустимости большого числа адаптаций или большого числа промежуточных анализов, которые могут превратить подтверждающее исследование в поисковое.

Поскольку интерес к адаптивному дизайну возрастает и в нашей стране, целью этой статьи является обзор основных определений, анализ дополнительных возможностей и рисков при его применении на разных фазах клинических исследований для лучшего понимания методологии такого подхода. В статье будут также рассмотрены несколько примеров практического применения адаптивного дизайна в исследованиях III фазы. Рамки статьи не позволяют представить многочисленные варианты дизайнов и разные методы адаптации в деталях, подробности можно найти в литературе, в частности, из предложенного списка. Мы также не будем отдельно рассматривать в статье особенности так называемого «бесшовного» дизайна (seamless design).

Что такое адаптивный дизайн?

Прежде всего, нужно определить, что такое адаптивный дизайн и какова мотивация для его использования в клинических исследованиях. Для обычного традиционного (неадаптивного) клинического исследования с фиксированным размером выборки много разных предположений об основных элементах дизайна должно быть сделано на стадии планирования: целевая популяция пациентов, ожидаемый размер эффекта (клинически значимые различия, вариабельность, доля ответчиков), зависимость «доза—эффект», доля досрочно выбывших пациентов и т.п. Это — именно предположения, поскольку заранее эти параметры исследования не известны. В ходе исследования собираются данные, на основе которых после завершения исследования проводится статистический анализ. Успех такого исследования во многом зависит от точности сделанных предположений. Известно, например, что многие плацебо-контролируемые исследования не смогли продемонстрировать преимущество изучаемой терапии и оказались негативными из-за недооценки плацебо—эффекта. Адаптивный дизайн позволяет уточнять наши предположения и, при необходимости, модифицировать характеристики исследования по мере накопления информации в ходе проведения исследования. То есть интуитивно

понятно, что адаптивный дизайн, являясь более гибким и эффективным, чем обычный фиксированный дизайн, может помочь при наличии неопределённости на стадии планирования исследования, повышая шансы исследования на успех.

Большая доля негативных клинических исследований вносит вклад в возрастающую стоимость процесса разработки лекарственных препаратов [26]. По данным обзоров, порядка 36% экспериментальных веществ (agents), вошедших в фазу I клинических исследований, не доходит до фазы II, 68—70% вошедших в фазу II не попадает на фазу III [27], неуспех исследований фазы III, по разным оценкам, составляет 40—50%. Как показали результаты обзора [28], причинами неуспеха и отказа в регистрации, помимо проблем с безопасностью, могут быть: 1) неопределённость в отношении оптимальной дозы, максимизирующей эффективность и минимизирующей риск токсичности; 2) популяция не отражала целевую популяцию пациентов, 3) неправильный выбор конечных точек, то есть выбор самого показателя и/или временных моментов его оценки (например, раньше достижения полного эффекта не позволяли отразить клинически значимое преимущество); 4) несоответствие результатов по нескольким выбранным конечным точкам; 5) несоответствие результатов эффективности для разных «частей» изучаемой популяции (эффективность была показана только в нескольких исследовательских центрах или только в подгруппах, которые заранее не планировалось изучать отдельно); 6) преимущество над плацебо было продемонстрировано, но размер эффекта был признан неадекватным по сравнению со стандартной терапией. Последняя причина относилась к серьёзным показаниям (например, аритмия, шизофрения), для которых уже были зарегистрированы более эффективные препараты, а исследования новых препаратов не показали какие-то другие преимущества, например, в отношении безопасности.

Так, в онкологии только порядка 34% исследований фазы III достигают статистической значимости в отношении первичной конечной точки [29]. Неуспех больших подтверждающих исследований фазы III имеет значительное влияние на научные инвестиции. В этих условиях спонсоры и регуляторы ищут новые стратегии и методологии, которые помогли бы сделать клинические исследования более продуктивными и информативными. Применение адаптивного дизайна является в этом смысле привлекательной перспективой. Кроме того, в эпоху растущего интереса к персонализированной медицине больше внимания уделяют адаптивным дизайнам с выбором подгрупп [30].

В последние десятилетия в клинических исследованиях применялись различные модификации дизайна и/или статистических процедур в ходе исследования на основе данных, наблюдаемых обычно на промежуточном(ых) анализе(ах). Целью такой модификации является не только эффективная иденти-

фикация клинической пользы изучаемого лечения, но и увеличение вероятности успеха самого исследования, то есть вероятности получения статистически значимого заключения о наличии эффекта, если он существует, или остановка исследования как можно быстрее, если эффекта нет. Среди элементов дизайна, которые модифицировались, можно назвать: критерии включения / не включения, изучаемая(ые) доза(ы), длительность терапии, конечные точки, схема наблюдения, число и моменты проведения промежуточных анализов, диагностические процедуры, определения ответа. Среди изменений, относящихся к статистике — рандомизация, цели и гипотезы, размер выборки, первичные и вторичные переменные, методы статистического анализа данных.

В вышедшем в 2010 г. руководстве по использованию адаптивного дизайна в клинических исследованиях FDA (“Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics”) [25] адаптивный дизайн был определён как исследование, которое включает проспективно запланированную возможность модификации одного или более элементов дизайна и гипотез на основе анализа полученных данных (обычно данных промежуточного анализа). «Прспективно» означает, что все изменения планируются заранее, детали определяются как можно раньше и до того, как разослепленные данные могут стать доступны специалистам, участвующим в планировании адаптации [25].

В документе подчеркивалось, что целостность и валидность исследования не должны были скомпрометированы в ходе применяемых адаптаций. В документе ЕМА тоже говорится о проспективном планировании адаптаций.

Таким образом, важное отличие регуляторного определения адаптивного дизайна от того, что было до этого принято называть «гибким» дизайном (flexible design), состоит в том, что все правила для изменения элементов дизайна на основе анализа данных промежуточных анализов полностью определяются заранее [31]. В отличие от запланированных изменений дизайна, все незапланированные изменения на основе наблюдаемых разослепленных данных являются потенциальной причиной систематической ошибки и могут явиться источником критики, в том числе, критики со стороны медицинского сообщества и регуляторов. Кроме того, и все параметры возможных изменений (например, порог остановки по бесполезности (futility)) и временные моменты любых адаптаций должны быть также оговорены заранее. Возможность остановки по бесполезности важна с этической и финансовой точек зрения, но выбор этого порога — достаточно сложное решение. Слишком низкий порог может «пропускать» в дальнейшее изучение неэффективные терапии, слишком высокий — несправедливо «браковать» эффективные.

Хотя определение адаптивного дизайна и накладывает определённые ограничения, но даже оно позволяет проводить достаточно широкий спектр адаптаций [25]: пересматривать критерии отбора (выбор популяции, для которой терапия имеет наибольшее преимущество), размер выборки и/или длительность исследования, параметры рандомизационной процедуры, режимы дозирования (дозовые уровни, схема, длительность и т.п.), число терапевтических групп, сопутствующую терапию, схему оценки состояния пациента, первичную конечную точку (например, единичная (unitary) против композитной (composite)), компоненты композитной, временные моменты оценки), вторичные конечные точки (выбор, порядок тестирования), методы анализа данных (ковариаты для финального анализа, статистическая методология, контроль ошибки первого рода).

Один из наиболее популярных элементов адаптации — изменение размера выборки и/или длительности исследования. При пересчёте размера выборки на основе данных промежуточного анализа решение об изменении может быть принято только в сторону увеличения [25]. Максимальный размер для потенциального увеличения определяется заранее. Снижение размера выборки не рекомендуется, поскольку высокая вариабельность оценки размера эффекта, частоты событий и дисперсии характерна для ранних этапов исследования. Возможность снижения запланированного размера выборки заложена в групповых последовательных методах, контролируя вероятность ошибки первого рода, позволяют осуществлять остановку по ранней демонстрации эффективности [25].

Некоторые из потенциальных адаптаций, например, касающиеся первичного анализа, включены в список, скорее, для его полноты, поскольку эти адаптации вряд ли будут приемлемы с регуляторной точки зрения, если даже будут запланированы. По поводу изменения или модификации первичной переменной в подтверждающих исследованиях в документе ЕМА [24] сказано, что очень трудно обосновать такие изменения, поскольку попытка комбинировать результаты разных стадий исследования с разными первичными переменными неизбежно вызовет статистические трудности (например, при попытке отвергнуть глобальную нулевую гипотезу).

Определение адаптивного дизайна в регуляторном документе [25] согласуется с определением, предложенным рабочей группой Американской ассоциации фармацевтических исследователей и производителей (Pharmaceutical Research and Manufacturers of America, PhRMA) в 2006 г. Адаптивный дизайн был определён как исследование с несколькими этапами (a multi-stage study), которое использует накапливающиеся в ходе исследования данные для решения, как модифицировать некоторые аспекты исследования без нарушения его валидности (validity) и

целостности (integrity) [4]. Термины «валидность», «целостность» и «систематическая ошибка» часто можно встретить в литературе, касающейся темы адаптивного дизайна в клинических исследованиях. Под валидностью понимают получение корректных статистических выводов, а целостность означает согласованность различных этапов клинического исследования. То есть адаптивный дизайн не должен вносить статистическую и операционную систематическую ошибку. Систематическую ошибку (bias) в этом контексте можно определить как систематическое влияние любых факторов, связанных с дизайном, проведением исследования, анализом данных, приводящее к отклонению оценки эффекта терапии от своего истинного значения.

Хотя формулировки регуляторных определений адаптивного дизайна несколько различаются, есть два основных аспекта, в которых они единодушны: изменения в адаптивном клиническом исследовании 1) планируются проспективно, 2) основаны на промежуточном анализе данных, собранных в самом исследовании (не из внешних источников, как например, результаты других исследований). Протоколы неадаптивных клинических исследований тоже могут подвергаться изменениям в ходе исследования, такие изменения могут вноситься в виде поправок к протоколу. Разница состоит в том, что в адаптивных исследованиях эти изменения запланированы заранее и проспективно включены в протокол и/или план статистического анализа.

Нужно предупредить, что нарушение валидности и целостности не является обязательным свойством адаптивного дизайна. Поддержание валидности означает, что, несмотря на применяемые в ходе исследования адаптации, корректность статистических выводов не нарушается. Для того чтобы это было возможно, операционная систематическая ошибка должна быть минимизирована и должна быть применена необходимая статистическая процедура для контроля вероятности ошибки первого рода, например, корректировка уровня значимости или доверительных интервалов. Целостность исследования поддерживается в основном за счёт сохранения «слепоты» (насколько возможно), а также проспективной природой самих адаптаций.

Типы исследований

Регуляторный документ FDA [25] выделяет два типа исследований, для которых применение адаптивного дизайна может рассматриваться: 1) так называемые адекватные и хорошо контролируемые исследования (adequate and well-controlled, A&WC) эффективности, подтверждающие (confirmatory), проводимые для регистрации препарата, и 2) поисковые (exploratory) исследования, которые могут рассматриваться как некоторое дополнение к A&WC

исследованиям. С точки зрения адаптивного дизайна, различия между этими типами исследований состоят в том, что для исследований 1) необходим строгий контроль уровня ошибки первого рода, в то время как для исследований типа 2) эта проблема гораздо менее критична. Основное внимание в документе уделяется исследованиям типа 1), но исследования типа 2) также обсуждаются.

В плане методологии использования адаптивного дизайна для исследований типа 1) основное беспокойство, выраженное в регуляторном документе, относится к возможной инфляции уровня ошибки первого рода и систематической ошибке в оценке эффектов изучаемой терапии. Это особенно важно для модификаций на основе разослепленных данных, в то время, как «слепой» анализ в этом плане не вызывает подобных опасений. Для планируемого разослепленного анализа в документе предлагается предоставить доказательства, желательно аналитические, что используемые статистические методы действительно контролируют вероятность ошибки первого рода.

Большое внимание в документе уделяется также необходимости избежать потенциальной операционной систематической ошибки (*operational bias*), причиной которой может стать «утечка» разослепленных результатов до завершения исследования. Если такое произойдет, научная валидность исследования может быть поставлена под сомнение, и результаты могут оказаться плохо интерпретируемыми. Изменение популяции пациентов после разослепленной адаптации можно привести в качестве примера операционной систематической ошибки. Конечно, изменение популяции пациентов в ходе клинического исследования может происходить и без адаптации. Это, например, возможно, если какие-то центры или регионы начинают набор в исследование позднее остальных. Однако основная регуляторная рекомендация в отношении адаптивного дизайна — необходимо обеспечить и документировать действия, предпринятые для предотвращения операционной систематической ошибки, связанной с разослепленными в ходе исследования данными.

Опасения относительно возможной связанной с адаптациями неоднородностью в результатах между различными этапами, например, из-за «сдвига» в популяции пациентов в ходе исследования, звучат в обоих регуляторных документах. Но гораздо большее внимание этому аспекту уделяют авторы документа ЕМА [24], они рекомендуют изучить согласованность демографии и оценок эффектов терапии между стадиями исследования (например, до и после адаптации) и представить эти результаты в финальном Отчёте. Неоднозначное отношение к проблеме тестирования однородности результатов в адаптивном исследовании обсуждается в статье [9].

Если адаптации требуют разослепления данных в процессе проведения исследования, нужно быть уверенными, чтобы промежуточные результаты не могли стать известны и как-то повлиять на поведение и объективизм исследователей в проведении лечения пациентов и оценок их состояния. Поскольку причина операционной систематической ошибки не является статистической, статистические методы не могут скорректировать эту часть систематической ошибки. Чтобы предотвратить операционную систематическую ошибку необходимо обеспечить «слепоту» исследователей в отношении выбранных адаптивных изменений, поскольку знания о разослепленных данных, на основании которых принимались адаптационные решения, или даже знания только о специфическом адаптивном выборе потенциально могут повлечь операционную систематическую ошибку в оценке эффекта терапии. Например, если знания исследователей о принятых адаптационных решениях влияют на то, как они лечат или оценивают состояние пациентов, или если исследователи под воздействием этих знаний начинают идентифицировать больше (или, наоборот, меньше) изучаемых событий во всех терапевтических группах. Последнее может привести к ложноположительным заключениям в исследованиях «не хуже» (*non-inferiority*) или ложноотрицательным в исследованиях преимущества. Кроме того, если есть какое-то разослепление на уровне пациентов, например, из-за наблюдаемых нежелательных явлений, систематическая ошибка также может повлиять на оценку межгрупповых различий. То есть, если терапии не могут быть полностью заслеплены, или оценка результата лечения предполагает некоторый субъективизм, все это может стать причиной систематической ошибки.

Нужно обратить внимание на определения «слепого» и разослепленного анализа в документе [25]. «Слепой» анализ означает, что принадлежность участников терапевтическим группам неизвестна и никаким образом не используется в анализе данных. Разослепленный анализ — принадлежность участников терапевтическим группам известна и используется в анализе данных, обычно (но не всегда) в формальном межгрупповом сравнении. Представление результатов по терапевтическим группам лицам, принимающим решения, открыто идентифицируя группы или как-то маскируя идентификаторы групп, рассматривается в обоих случаях как разослепленный анализ.

Что касается поисковых исследований типа 2), можно сказать, что регуляторный документ призывает использовать адаптивный дизайн в таких исследованиях. В документе говорится, что адаптивный дизайн позволяет лучше изучить соотношение доза — ответ, эффект в подгруппах и т.п., может существенно улучшить знания об эффективности изучаемой терапии, что, в свою очередь, повышает шансы

на успех исследований более поздней фазы. Кроме того, такое использование ведёт к накоплению знаний о практическом применении методов разослепленной адаптации. Одной потенциальной проблемой реализации этой рекомендации можно считать желание спонсоров проводить исследования фазы II как «мини — фаза III» в надежде, что полученные хорошие результаты можно будет использовать в процессе регистрации.

Типы адаптивных дизайнов:

«Хорошо изученные» (Well-Understood) против «Менее изученных» (Less Well-Understood)

Ещё одним важным аспектом документа FDA [25] является классификация типов адаптивного дизайна: «хорошо изученные» (*well-understood*) и «менее изученные» (*less well-understood*) внутри класса A&WC исследований.

К так называемым «хорошо изученным» типам адаптивного дизайна был отнесён широкий спектр групповых последовательных методов (*group sequential methods*, например, методы O'Brien-Fleming, Pocock, Peto, Lan-DeMets) с возможностью ранней остановки как по демонстрации эффективности, так и по бесполезности (*futility*), а также адаптации, которые не включают анализ разослепленных данных, собранных после начала терапии (*post-baseline unblinded data*). В качестве примеров можно назвать адаптацию критериев отбора на основе исходных данных, «слепой» пересчёт размера выборки или длительности исследования. В общем можно сказать, что адаптация на основе заслепленных и/или исходных данных, проводимая персоналом, не имеющим доступа к разослепленным данным, не влечёт регуляторные проблемы. Хотя групповые последовательные методы на основе разослепленных данных тоже были включены в категорию «хорошо изученных».

Остальные методы адаптации на основе разослепленных данных, собранных после начала терапии, были отнесены к категории «недостаточно изученных». Примерами могут служить пересчёт размера выборки / длительности исследования на основе разослепленных данных, адаптивная рандомизация на основе ответа (*response adaptive randomization*), адаптивный поиск подгрупп и/или конечных точек на основе наблюдаемого эффекта терапии, адаптивный выбор дозы. Внутри категории таких «менее изученных» адаптивных дизайнов в документе выделяются дизайны с множественными адаптациями и с адаптациями в исследованиях «не хуже» (*non-inferiority*). По поводу первой группы дизайнов документ предупреждает о возможной сложности с интерпретацией получаемых результатов. Модификации в исследованиях «не хуже» возможна, но не должна затрагивать такие элементы дизайна, как

установленная граница (*non-inferiority margin*).

Первоначально появление в регуляторном документе подобной классификации было неправильно интерпретировано, как то, что применение только так называемых «хорошо изученных» типов адаптивного дизайна допускается FDA. Однако в действительности категоризация относится, скорее, к состоянию регуляторных знаний, касающихся различных типов дизайна, на момент выхода документа. Из-за ограниченного опыта применения «недостаточно изученных» адаптивных дизайнов уровень риска систематической ошибки и возможный размер потенциальной систематической ошибки, а также пути её предотвращения нуждаются в дальнейшем изучении.

Регуляторный документ не запрещает применение «недостаточно изученных» дизайнов, а рекомендует приложить все усилия для адекватного контроля вероятности ошибки первого рода и минимизации влияния связанной с адаптацией статистической и операционной систематической ошибки на оценку эффектов терапии и интерпретируемость получаемых в исследовании результатов. Помимо перечисленных выше проблем, для этого типа адаптивного дизайна авторы обращают внимание на потенциально возможное увеличение уровня ошибки второго рода [25]. Причиной, например, могут быть слишком либеральные правила остановки из-за бесполезности (*futility*) или неоптимальный выбор дозы. Все эти вопросы должны быть рассмотрены при планировании таких адаптаций.

Возможно, по мере накопления знаний о практическом применении разослепленных методов адаптации, классификация методов адаптации будет меняться, и методы из категории «недостаточно изученных» будут постепенно переходить в категорию «хорошо изученных».

Роль моделирования и симуляционных исследований

В области разработки и практического внедрения адаптивных дизайнов моделирование всегда играло центральную роль в оценке и понимании их свойств и операционных характеристик. Даже планирование использования достаточно простых адаптаций, например, «слепого» пересчёта размера выборки требует симуляций различных сценариев (величины возможных различий, вариабельности) для принятия решения о целесообразности такой адаптации и об оптимальном времени проведения промежуточного анализа. Моделирование может помочь оценить влияние характеристик набора и выбывания, профилей «доза—эффект», корреляции между конечными точками. Моделирование и симуляции могут помочь в оценке влияния адаптаций на уровень ошибки первого рода и мощность исследования.

Регуляторный документ FDA [25] признаёт и подчеркивает важность моделирования и симуляционных расчётов для оценки операционных характеристик адаптивного дизайна, а также для сравнения альтернативных дизайнов и обоснования выбора элементов адаптации. Среди характеристик дизайна, которые могут быть изучены с помощью моделирования и симуляций, авторы документа называют влияние на вероятность ошибки первого и второго рода, систематическую ошибку в оценке эффектов терапии.

В соответствии с документом FDA детали проведения моделирования, включающие предположения для создания модели, сами модели, программы и временные схемы возможных адаптаций, которые использовались в процессе моделирования, должны быть тщательно документированы для представления в регуляторные органы [25].

Важную роль в исследованиях с адаптивным дизайном отводят плану статистического анализа. Для поддержания целостности исследования план статистического анализа должен быть финализирован до разослепления данных, а в некоторых случаях до начала сбора данных. Все результаты, включаемые в финальный отчёт, должны быть выполнены в строгом соответствии с планом. Большое значение для обеспечения целостности исследования имеют независимые группы, ответственные за проведение разослепленных промежуточных анализов и принятие адаптационных решений.

Основные преимущества, барьеры и риски использования адаптивного дизайна

Фармацевтическая индустрия, столкнувшись со значительными организационными и финансовыми проблемами, ищет пути повышения эффективности клинических исследований, таких как снижение стоимости и сокращение временных рамок. Мотивацией для применения адаптивного дизайна является более эффективное получение информации по сравнению с неадаптивным дизайном. Было бы неправильно говорить, что адаптивные исследования всегда дешевле и короче, чем неадаптивные. Речь идёт о получении большего объёма поддающейся оценке информации за те же инвестиции и время из-за возможности корректировать критические элементы дизайна в ходе исследования. В качестве дополнительных преимуществ также упоминается 1) возрастание вероятности успеха, например, за счет промежуточного анализа в середине исследования, правильного выбора целевой популяции, корректной дозы, адекватного размера выборки, 2) возможность ранней остановки из-за бесполезности (futility), 3) возможность лучшего понимания эффекта терапии (например, лучшая оценки профиля доза — ответ или эффекта в подгруппах пациентов). Если,

условно говоря, терапия «не работает», адаптивный дизайн поможет остановить её использование как можно быстрее, что позволит сэкономить ресурсы и избавить пациентов от получения ненужного лечения, в то время, как эффективная и безопасная терапия может стать доступной для пациентов гораздо быстрее.

Учитывая, что организация клинического исследования часто наталкивается на бюджетные ограничения, финансовым аспектом преимуществ адаптивного дизайна может быть возможность начать исследование с относительно небольшим бюджетом, а запросить его увеличение после того, как подающие надежды результаты будут получены на промежуточном анализе.

Несмотря на имеющийся потенциал для улучшения эффективности процесса оценки новых препаратов и методов лечения по сравнению с традиционным дизайном и возможность смягчить некоторые ограничения последнего, а также несмотря на популярность адаптивных методов в научной литературе [32—40] и появление регуляторных документов, способствующих практическому внедрению [41—45], адаптивный дизайн не очень широко используется в рутинной практике. Работы, изучавшие барьеры на пути использования адаптивного дизайна, отмечали в качестве причин: отсутствие практических знаний и опыта, примеры прерывания или задержки при выполнении адаптивных исследований, потенциальные проблемы с распределением и поставками препарата в исследовательские центры, потенциальные проблемы с мониторингом данных без возникновения операционной систематической ошибки, ограниченная доступность независимых комитетов по мониторингу данных, недостаточно глубокое понимание типов адаптивного дизайна, опасение поставить под угрозу регуляторное одобрение, консерватизм [3, 34, 46—48].

По данным на 2013 г. [48], приблизительно 20% клинических исследований использовали простые методы адаптивного дизайна. Однако такие оценки варьируют в разных источниках, в том числе в зависимости от временного периода оценки и от выборки исследований для оценки этой доли.

Пытаясь оценить динамику в использовании адаптивного дизайна, авторы работы [49] проанализировали данные о клинических исследованиях с 2001 по 2014 гг., представленные на ClinicalTrials.gov (159 645 исследований) и в регистре Национального института исследований в области здравоохранения (National Institute of Health Research, NIHR register) (порядка 2 300 исследований). Авторы отмечали, хотя и не бурный, но рост практического использования адаптивных исследований. Так, общее число адаптивных исследований на 10 000 зарегистрированных исследований было оценено как 11 с 95% доверительным интервалом (4—17) для перио-

да 2001—2005 гг., 22 (11—29) — для 2006—2008 гг., 29 (21—37) — для 2009 — 2011 гг. и 38 (27 — 50) — для 2012 — 2014 гг. Большинство адаптивных исследований относились к фазе II, что соответствует литературным и регуляторным данным о более широком внедрении различных методов адаптивного дизайна именно в поисковых исследованиях ранних фаз [50—52]. Результаты отличались в зависимости от региона и медицинской области.

Среди наиболее популярных методов адаптации были названы групповые последовательные методы, в том числе с пересчётом размера выборки как на фазе II, так и на фазе III, а также адаптации, связанные с выбором дозы на фазе II клинических исследований. Популярность групповых последовательных методов, возможно, связана с их глубокой теоретической проработкой, а также упоминанием в регуляторном документе FDA как «хорошо изученных».

Адаптивные методы неравномерно применяются в разных областях медицины [49]. Областью, в которой адаптивные методы находят наиболее широкое применение, в работе была названа онкология. И для этого есть много причин. В частности, можно сказать, что исследователи и регуляторы в этой области более восприимчивы к адаптации, и им проще принять адаптивный дизайн. Если для определённого типа рака существуют ограниченные возможности текущего стандарта лечения, медицинское сообщество хотело бы как можно быстрее понять, будет ли новая терапия разрешена для использования, чтобы пациенты в исследовании могли получить к ней доступ, и их можно было бы перевести со стандартного лечения. Учитывая необходимость ждать много лет, чтобы получить ответ на исследуемый вопрос на основании основного клинического показателя общей выживаемости (обычно используемая первичная переменная для онкологических исследований фазы III), было бы полезно иметь возможность принимать решение о терапии на основе промежуточных результатов по такому показателю, как, например, безрецидивная выживаемость, пока не будут получены финальные результаты по основному показателю.

Несмотря на все перечисленные дополнительные возможности, адаптивный дизайн не всегда лучше фиксированного. Моделирование может помочь с разных сторон оценить преимущества применения определённых адаптаций для данного исследования по сравнению с неадаптивным дизайном. Если вопрос о применении адаптивного дизайна решается в отношении конкретного исследования, следующие риски нужно учитывать при принятии такого решения. Одним из рисков является операционный риск, связанный с проспективным планированием логистики и поставок препарата, особенно если адаптация позволяет менять число дозовых уровней, число терапевтических групп, переводить пациентов в группу оптимальной терапии, менять размер выбор-

ки. Ещё один важный риск связан с регуляторным одобрением адаптивного исследования. До настоящего времени в мире однозначно «принимаемые» регуляторами статистические методы существуют не для всех типов адаптаций. Кроме того, статистические методы должны использоваться адекватно, чтобы не вносить статистическую систематическую ошибку. В частности, такая ошибка может быть связана с недостаточным контролем заданного уровня альфа может служить тестирование нескольких статистических гипотез, несколько тестирований одной гипотезы из-за промежуточных анализов, наличие нескольких конечных точек (первичных, вторичных, композитных), сравнение нескольких терапевтических групп или анализ в подгруппах. Кроме того, в исследованиях с адаптивным дизайном общий эффект терапии получается, в некотором смысле, как комбинация эффектов, наблюдаемых на каждом из этапов (до и после модификации), и оценки этого общего эффекта используется при тестировании гипотезы. То, каким образом будет проводиться объединение данных каждого этапа, может влиять на валидность оценки общего эффекта терапии. Хотя адаптивный дизайн часто используется для уменьшения неопределённости на этапе планирования, это не означает, что процесс планирования при этом становится проще, чем для фиксированного дизайна. Наоборот, адаптивное исследование нуждается в более глубоком продумывании, что требует участия специалистов высокой квалификации в различных областях клинических исследований.

Разработки в области проведения исследований, включая статистические и информационные технологии, а также представление методологии адаптивного дизайна в научной литературе играют важную роль в поддержке адаптивных исследований [53—58]. Адаптивный дизайн может быть и контрпродуктивным. В частности, маловероятно, что неожиданные результаты будут обнаружены во время ограниченного и быстро проводимого промежуточного анализа данных, кроме того, в результате «закрытия» некоторых терапевтических групп соотношение «доза—эффект» не будет изучено полностью. Это может также привести к неадекватному выявлению проблем с безопасностью, что является одной из основных целей исследования. Адаптивный дизайн может также несколько «сдвинуть» целевую популяцию пациентов, что приведёт к ограниченной генерализации полученных результатов. Потенциальным недостатком адаптивного дизайна является возможность переоценивать эффект терапии. Например, если адаптивный дизайн выбирает наилучший из наблюдаемых на промежуточном анализе эффектов терапии, который может быть проявлением случайности, то такая оценка эффекта будет превышать его реальный размер.

К каким последствиям ошибочный прогноз на основе оценки размера эффекта на промежуточном анализе приводит в реальных клинических исследованиях, иллюстрируют приведённые ниже примеры.

Практические примеры применения адаптивных дизайнов в клинических исследованиях фазы III

Можно привести достаточно большой список успешных подтверждающих клинических исследований с применением адаптивного дизайна. Однако хотелось бы остановиться на двух негативных примерах, которые иллюстрируют сформулированное в [25] положение о том, что характеристики и свойства многих элементов адаптивного дизайна ещё недостаточно изучены. В этих примерах применялись разные методы адаптации, но их объединяет один «подводный камень» адаптивного дизайна — наблюдаемые промежуточные результаты не всегда позволяют правильно предсказать результаты в конце исследования, даже при корректном применении статистических методов.

Пример 1. Рандомизированное двойное слепое исследование фазы III LUME-Lung 2 (nintedanib в комбинации с remetrexed против плацебо+remetrexed у пациентов с распространённой формой или рецидивирующим немелкоклеточным раком лёгких с рецидивом или неэффективностью после первой линии химиотерапии) с элементом адаптивного дизайна — возможность ранней остановки по бесполезности (futility). Результаты исследования представлены в [59] и подробно проанализированы авторами статьи [60]. В исследование планировалось рандомизировать 1 300 пациентов на основе предположения, что 713 PFS-событий необходимо для поддержания 90% мощности продемонстрировать ожидаемое увеличение на 27,5% в медианах PFS (HR=0,78).

В клинических исследованиях, в том числе в области онкологии, часто используется в качестве элемента адаптивного дизайна запланированная заранее остановка по бесполезности (futility) по результатам промежуточного анализа. Так исследование LUME-Lung 2 было остановлено в ходе такого анализа, выполненного независимым Комитетом по мониторингу данных на основе оценок исследователями выживаемости без прогрессирования (PFS, оценка с помощью modified Response Evaluation Criteria In Solid Tumors [RECIST] v1.0), после того как произошли 48% событий (из 713 запланированных) для анализа PFS. Порог для остановки по бесполезности был установлен как <20% условной мощности (вероятность, что новая терапия окажется лучше контрольной в конце исследования, оцененная на основе результатов промежуточного анализа и предположений об эффективности в исследовании). Во время проведения этого запланированного промежуточного анализа условная вероятность успеха

исследования на основе наблюдаемых данных (условная мощность) была оценена как 10,3%. Независимый комитет, проводивший промежуточный анализ, принял решение остановить набор в исследование. На момент остановки набора было рандомизировано 713 пациентов. Проблем с безопасностью не было, пациенты продолжили лечение и наблюдение в соответствии с протоколом. Проведённый позднее финальный анализ PFS (498 событий) по оценкам независимых оценщиков показал статистически значимое преимущество изучаемой терапии: HR=0,83; 95% доверительный интервал 0,70—0,99; $p=0,0435$; медиана PFS — 4,4 против 3,6 мес. Анализ PFS на основе оценок независимых оценщиков был запланирован как первичный в исследовании, а на основе оценок исследователей — вторичный анализ, но основной на промежуточном анализе данных.

Ретроспективный анализ выявил значительную вариабельность во времени условной мощности для PFS как по оценкам исследователей, так и оценкам независимых оценщиков. При этом мощность падала ниже порога остановки только на момент проведения промежуточного анализа и только для анализа PFS на основе оценок исследователей. Поэтому ретроспективно был сделан вывод, что, если бы промежуточный анализ был проведён в другой момент времени или основывался бы на оценках независимых оценщиков, условная мощность не оказалась бы ниже порогового значения, исследование было бы продолжено и успешно завершено.

Нужно заметить, что размер эффекта в этом исследовании был средним, финальная оценка отношения рисков HR была 0,83. Для исследований с маленьким или средним размером эффекта такие «подводные камни» — не редкость. Большие размеры эффекта обычно «проявляются» в исследовании раньше, и низкая условная мощность на промежуточном анализе маловероятна.

Причины таких колебаний условной мощности не были установлены, но, независимо от конкретных причин и в отсутствии ошибок в проведении промежуточного анализа, величина этих колебаний привела к ошибочному заключению, что вероятность улучшения, наблюдаемого на промежуточном анализе размера эффекта, слишком мала, и исследование нужно досрочно остановить.

Основные уроки, которые авторы [60] рекомендуют извлечь из этого примера, следующие. Необходимо тщательно выбирать порог остановки по бесполезности, нет одного значения такого порога, подходящего для всех исследований. В частности, для дополнительной, или добавочной (add-on) терапии в исследовании LUME-Lung 2 пороговое значение 20% для анализа 50% событий PFS можно считать «агрессивным» (завышенным), 15 или 10% могут рассматриваться в качестве порогов остановки для таких терапий [60]. Если оценка услов-

ной мощности оказалась все-таки ниже порога для остановки, возможно, имеет смысл глубже проанализировать ситуацию, например, оценить степень и направление «дрейфа» (drift, отклонение от ожидаемых результатов), связанного с тестируемой гипотезой, и даже более информативно, с альтернативной гипотезой. Это может дать полезную информацию для принятия важного решения об остановке исследования.

В свете наблюдаемых различий результатов по PFS на основе оценок исследователей и независимых оценщиков авторы статьи поддерживают рекомендации принимать решение об остановке исследования по результатам анализа более чем одной конечной точки [61, 62]. Кроме того, если суррогатная конечная точка используется на промежуточном анализе, это должно приниматься во внимание при принятии решения об остановке исследования.

Пример 2. Исследование VALOR [63] — фаза III, плацебо-контролируемое рандомизированное двойное слепое исследование (Vosaroxin плюс cytarabine против плацебо плюс cytarabine у пациентов после первого рецидива или с рефрактерной формой острого миелоидного лейкоза). 101 клинический центр в разных странах участвовал в исследовании, было рандомизировано 711 пациентов, стратифицированных по возрасту, географическому региону, и статусу заболевания. Первичная переменная в исследовании — общая выживаемость (Overall Survival, OS).

В исследовании использовался один из самых распространённых элементов адаптации — перерасчёт размера выборки (Sample Size Re-estimation, SSR).

На основе данных промежуточного анализа приблизительно в середине исследования планировалось:

- 1) остановиться ранее по демонстрации эффективности;
- 2) остановиться ранее по бесполезности (futility) при низкой условной мощности
- 3) продолжить исследование без изменения размера выборки;
- 4) увеличить планируемый размер выборки при необходимости (увеличение на 50% числа событий общей выживаемости).

Первоначально планировалось наблюдать в исследовании 375 OS-событий у 450 пациентов, что поддерживало бы 90% мощность (для двустороннего уровня значимости $\alpha=5\%$) в предположении медиан выживаемости 5 и 7 мес в группе плацебо/cytarabine и группе vosaroxin/cytarabine, соответственно (40% увеличение, HR=0,71). На основе данных промежуточного анализа (173 OS-события, приблизительно 50% от запланированных) независимый комитет рекомендовал вариант 4 — увеличение до 562 наблюдаемых событий с соответствующим увеличением размера выборки до 675 пациентов. Эта рекомендация была основана на более консервативных предположениях и альтернативном сценарии. Увеличение

размера выборки до 675 пациентов позволяло поддерживать 90% статистическую мощность для выявления 30% различий в общей выживаемости (медиана = 5 мес против 6,5 мес, HR=0,77).

Однако реальные наблюдаемые различия оказались ещё меньше, чем консервативные ожидания, хотя и клинически значимыми. Несмотря на все дополнительные усилия (планирование адаптивного дизайна, привлечение независимого комитета по рассмотрению данных, проведение промежуточного анализа, пересчёт размера выборки) в исследовании не удалось продемонстрировать статистически значимое улучшение по первичной переменной эффективности на финальном анализе: медиана общей выживаемости = 7,5 мес для группы vosaroxin и cytarabine против 6,1 мес для группы плацебо и cytarabine (HR=0,87; 95% доверительный интервал 0,73—1,02; логранговый тест без учёта стратификационных факторов, $p=0,0610$).

Исследование VALOR планировалось с высокой статистической мощностью, адаптация была проведена корректно с точки зрения статистической методологии. Основным вывод, который можно сделать из этого примера, — адаптивный дизайн позволяет в какой-то степени снизить неопределённость и повысить мощность в ходе проведения исследования. Но он не может полностью устранить неопределённость и гарантировать успех исследования.

Заключение

Адаптивный дизайн позволяет проводить запланированные модификации на основе наблюдаемых и накапливаемых в исследовании данных. При планировании исследования свойства различных адаптивных методов должны быть аккуратно рассмотрены в контексте конкретного исследования как с научной стороны, так и с точки зрения лежащих в основе статистических предположений. Если методы адаптивного дизайна используются корректно, их эффективность может выражаться в снижении размера выборки, более эффективном процессе прохождения фаз клинических исследований, увеличении вероятности ответа на клинический вопрос(ы), стоящий(е) перед клиническим исследованием. Тем не менее, методологические ошибки адаптации могут привести к систематической ошибке и невозможности интерпретировать результаты.

Можно сказать, что общий «сигнал» регуляторных документов — применение адаптивного дизайна в клинических исследованиях допустимо, но с осторожностью и особым вниманием к грамотному планированию, внедрению и документированию. Адаптивные исследования должны быть достаточно простыми, необходимо избегать слишком большого числа и/или слишком сложных адаптаций в одном исследовании.

Однако, несмотря на позитивный тон регуляторных документов, всё возрастающий интерес и постепенно накапливаемый опыт, в области применения адаптивного дизайна остаётся ещё много «белых пятен» и потребность в разработке новых и совершенствование используемых статистических методов.

Несмотря на все преимущества, адаптивный дизайн в клинических исследованиях используется по-прежнему не очень широко. Возможной причиной является так называемая «плата» за адаптивность. Больших усилий требует этап планирования исследования, а также документирования процесса адаптации и действий, направленных на обеспечение целостности и валидности. Большие операционные

усилия затрачиваются на получение качественных данных на промежуточных анализах, поддержание валидности исследования, сохранение «слепоты» полученных в ходе исследования результатов. Для решения последних задач привлекаются независимые эксперты и комитеты по рассмотрению, анализа данных, а также принятию запланированных адаптационных решений. Кроме того, как было показано на примерах, даже грамотное применение методологии адаптивного дизайна не может однозначно гарантировать успех исследования даже в случае, если преимущество терапии, особенно с относительно небольшим, хотя и клинически значимым, размером эффекта действительно существует.

Литература

1. Chow S.C., Chang M. Adaptive Design Methods in Clinical Trials. Chapman and Hall/CRC Press, Taylor and Francis, New York, NY; 2006.
2. Chow S.C., Chang M. Adaptive design methods in clinical trials — a review. *Orphanet Journal of Rare Diseases* 2008; 3: 11—24.
3. Kairalla J.A., Coffey C.S., Thomann M.A. and Muller K.E. Adaptive trial designs: a review of barriers and opportunities. *Trials* 2012; 13: 145—153
4. Gallo P, Chuang-Stein C., Dragalin V., Gaydos B., Krams M., Pinheiro J. Adaptive design in clinical drug development — an executive summary of the PhRMA Working Group (with discussions). *Journal of Biopharmaceutical Statistics* 2006; 16 (3): 275—283.
5. Chang M. Adaptive Design Theory and Implementation Using SAS and R. Chapman and Hall/CRC Press, Taylor and Francis, New York, NY; 2007.
6. Chow S.C., Chang M., Pong A. Statistical consideration of adaptive methods in clinical development. *Journal of Biopharmaceutical Statistics* 2005; 15: 575—591.
7. Ellenberg S.S., Fleming T.R., DeMets D.L. Data Monitoring Committees in Clinical Trials: A Practical Perspective. John Wiley and Sons, New York; 2002.
8. Dragalin V. Adaptive designs: terminology and classification. *Drug Inf J* 2006; 40: 425—435.
9. Gaydos B., Anderson K.M., Berry D., Burnham N., Chuang-Stein C., Dudinak J., et al. Good practices for adaptive clinical trials in pharmaceutical product development. *Drug Information Journal* 2009; 43: 539—556.
10. Quinlan J., Krams M. Implementing adaptive designs: logistical and operational considerations. *Drug Inf J* 2006; 40 (4): 437—44.
11. Lachin J.M. Statistical properties of randomization in clinical trials. *Controlled Clinical Trials* 1988; 9: 289—311.
12. Rosenberger W.F., Stallard N., Ivanova A., Harper C.N., Ricks M.L. Optimal adaptive designs for binary response trials. *Biometrics* 2001; 57: 909—913.
13. Hardwick J.P., Stout Q.F. Optimal few-stage designs. *Journal of Statistical Planning and Inference* 2002; 104: 121—145.
14. Lan K.K.G., DeMets D.L. Group sequential procedures: calendar versus information time. *Statistics in Medicine* 1987; 8: 1191—1198.
15. Posch M., Bauer P. Adaptive two-stage designs and the conditional error function. *Biometrical Journal* 1999; 41: 689—696.
16. Lehman W., Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics* 1999; 55: 1286—1290.
17. Liu Q., Proschan M.A., Pledger G.W. A unified theory of two-stage adaptive designs. *Journal of American Statistical Association* 2002; 97: 1034—1041.
18. Jennison C., Turnbull B.W. Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall, New York, NY; 2000.
19. Jennison C., Turnbull B.W. Meta-analysis and adaptive group sequential design in the clinical development process. *J Biopharm Stat* 2005; 15 (4): 537—558.
20. Cui L., Hung H.M.J., Wang S.J. Modification of sample size in group sequential trials. *Biometrics* 1999; 55: 853—857.
21. Shih W.J. Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison. *Statistics in Medicine* 2006; 25: 933—941.
22. Chung-Stein C., Anderson K., Gallo P., Collins S. Sample size reestimation: a review and recommendations. *Drug Information Journal* 2006; 40: 475—484.
23. Vandemeulebroeke M. Group sequential and adaptive designs—a review of basic concepts and points of discussion. *Biometrical J* 2008; 50: 541—557.
24. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (2007). European Medicines Agency. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf
25. Guidance for Industry (2010) Adaptive Design Clinical Trials for Drugs and Biologicals. Food and Drug Administration. www.fda.gov
26. Morgan S., Grootendorst P., Lexchin J., Cunningham L., Greyson D. The cost of drug development: a systematic review. *Health Policy* 2011; 100 (1): 4—17.
27. Hay M., Thomas D.W., Craighead J.L., Economides C., Rosenthal J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* 2014; 32 (1): 40—51.
28. Sacks L.V., Shamsuddin H.H., Yasinskaya Y.I., Bouri K., Lanthier M.L., Sherman R.E. Scientific and Regulatory Reasons for Delay and Denial of FDA Approval of Initial Applications for New Drugs, 2000—2012. *JAMA* 2014; 311 (4): 378—384.
29. Hay M., Rosenthal J., Thomas D., Craighead J. Bio/BioMedTracker clinical trial success rates study. Presented at: BIO CEO & Investor Conference, 15 February 2011.
30. Bauer P., Bretz F., Dragalin V., Konig F., Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Stat Med* 2016; 35: 325-47.
31. Brannath W., Koenig F., Bauer P. Multiplicity and flexibility in clinical trials. *Pharm Stat* 2007; 6: 205—216.
32. Bauer P., Einfalt J. Application of adaptive designs — a review. *Biom J* 2006; 48 (4): 493—506.
33. Coffey C.S., Kairalla J.A. Adaptive clinical trials. *Drugs R&D*. 2008; 9 (4): 229—42.
34. Coffey C.S., Levin B., Clark C., Timmerman C., Wittes J., Gilbert P., et al. Overview, hurdles, and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop. *Clin Trials*. 2012; 9 (6): 671—80.
35. Retzios A.D. Innovation in Drug Development: Adaptive Designs for Clinical Trial. 2010. Bay Clinical R&D Services.
36. van der Graaf R., Roes K.C., van Delden J.J. Adaptive trials in clinical research: scientific and ethical issues to consider. *JAMA* 2012; 307: 2379—2380.
37. Bornkamp B., Bretz F., Dmitrienko A., Enas G., Gaydos B., Hsu C., et al. Innovative approaches for designing and analyzing adaptive dose-ranging trials. *JBiopharm Stat* 2007; 17: 965—995.
38. Freidlin B., Simon R. Evaluation of randomized discontinuation design. *J Clin Oncol* 2005; 23: 5094—5098.

39. Wang S.J., Hung H.M.J., O'Neill R.T. Adaptive patient enrichment designs in therapeutic trials. *Biometrical J* 2009; 51: 358—374.
40. Bretz F, Koenig F, Brannath W., Glimm E., Posch M. Adaptive designs for confirmatory clinical trials. *Stat Med* 2009; 28: 1181—1217.
41. Chuang-Stein C., Beltangady M. FDA draft guidance on adaptive design clinical trials: Pfizer's perspective. *J Biopharm Stat* 2010; 20 (6): 1143—9.
42. Cook T., DeMets D.L. Review of draft FDA adaptive design guidance. *J Biopharm Stat* 2010; 20 (6): 1132—42.
43. Gallo P, Anderson K., Chuang-Stein C., Dragalin V., Gaydos B., Krams M., et al. Viewpoints on the FDA draft adaptive designs guidance from the PhRMA working group. *J Biopharm Stat* 2010; 20 (6): 1115—24.
44. Brannath W., Burger H.U., Glimm E., Stallard N., Vandemeulebroecke M., et al. Comments on the draft guidance on “adaptive design clinical trials for drugs and biologics” of the U.S. Food and Drug Administration. *J Biopharm Stat* 2010; 20: 1125—1131.
45. Rong Y. Regulations on Adaptive Design Clinical Trials. *Pharmaceut Reg Affairs* 2014; 3: 116—123.
46. Jaki T. Uptake of novel statistical methods for early-phase clinical studies in the UK public sector. *Clin Trials* 2013; 10 (2): 344—6.
47. Dimairo M., Boote J., Julious S.A., Nicholl J.P., Todd S. Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. *Trials*. 2015; 16 (1): 430—446.
48. Development TCtSoD. The Adoption and Impact of Adaptive Trial Designs. Boston, MA, USA: TUFTS University. 2013.
49. Hatfield I., Allison A., Flight L., Julious S.A., Dimairo M. Adaptive designs undertaken in clinical research: a review of registered clinical trials. *Trials* 2016; 17: 150—163.
50. Quinlan J., Gaydos B., Maca J., Krams M. Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clin Trials* 2010; 7 (2): 167—73.
51. Elsässer A., Regnstrom J., Vetter T., Koenig F., Hemmings R.J., Greco M., et al. Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. *Trials* 2014; 15 (1): 383—393.
52. Morgan C.C., Huyck S., Jenkins M., Chen L., Bedding A., Coffey C.S., et al. Adaptive design: results of 2012 survey on perception and use. *Ther Innov Regul Sci* 2014; 48 (4): 473—81.
53. Chow S.C., Corey R. Benefits, challenges and obstacles of adaptive clinical trial designs. *Orph J Rare Dis* 2011; 6: 79—89.
54. Cheung K., Kaufmann P. Efficiency perspectives on adaptive designs in stroke clinical trials. *Stroke* 2011; 42: 2990—2994.
55. Coffey C.S. You may worked on more adaptive designs than you think. *Stroke* 2015; 46 (2): 26—28.
56. Berry D.A. Adaptive clinical trials: the promise and the caution. *J Clin Oncol* 2011; 29 (6): 606—9.
57. Kaplan R., Maughan T., Crook A., Fisher D., Wilson R., Brown L., et al. Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol*. 2013; 31 (36): 4562—8.
58. Elman S.A., Ware J.H., Gottlieb A.B., Merola J.F. Adaptive Clinical Trial Design: An Overview and Potential Applications in Dermatology. *Journal of Investigative Dermatology* 2016; 136: 1325—1329.
59. Hanna N.H., Kaiser R., Sullivan R.N., Aren O.R., Ahn M.J., Tiangco B., et al. Nintedanib plus pemetrexed versus placebo plus pemetrexed in patients with relapsed or refractory, advanced non-small cell lung cancer (LUME-Lung 2): a randomized, double-blind, phase III trial. *Lung Cancer* 2016; 102: 65—73.
60. Lesaffre E., Edelman M.J., Hanna N.H., Park K., Thatcher N., Willemsen S., et al. Statistical controversies in clinical research: Futility analyses in oncology — lessons on potential pitfalls from a randomized controlled trial. *Annals of Oncology* 2017; 28 (7): 1419—1426.
61. Jitlal M., Khan I., Lee S.M., Hackshaw A. Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies. *Br J Cancer* 2012; 107: 910—917.
62. Hughes S., Cuffe R.L., Liefucht A., Garrett Nichols W. Informing the selection of futility stopping thresholds: case study from a late-phase clinical trial. *Pharm Stat* 2009; 8: 25—37.
63. Ravandi F., Ritchie E.K., Sayar H., Lancet J.E., Craig M.D., Vey N., et al. Vosaroxin plus cytarabine versus placebo plus cytarabine in patients with first relapsed or refractory acute myeloid leukaemia (VALOR): a randomised, controlled, double-blind, multinational, phase 3 study. *Lancet Oncol* 2015; 16 (9): 1025—1036.