



О.В. Шредер 
Д.В. Горячев 
В.А. Меркулов 

Основные принципы расчета необходимой численности участников клинических исследований. Часть 1. Общие подходы (обзор)

*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*

✉ Шредер Ольга Васильевна; shrederov@expmed.ru

РЕЗЮМЕ

ВВЕДЕНИЕ. Корректное планирование клинического исследования (КИ) является гарантией получения валидных результатов оценки эффективности и безопасности медицинского применения лекарственных средств. В настоящее время отсутствуют четкие критерии выбора базовых элементов, лежащих в основе разработки клинического дизайна, и прежде всего исследовательских гипотез, способов определения ожидаемой величины терапевтического эффекта, уровня статистической значимости и мощности исследования, статистических моделей расчета размера выборки субъектов.

ЦЕЛЬ. Систематизация и гармонизация технических требований к планированию дизайна клинического исследования в части определения размера выборки.

ОБСУЖДЕНИЕ. В работе представлены основные требования и методологические подходы к разработке дизайнов медицинских исследований, направленных на оценку эффективности и подтверждение безопасности лекарственных средств. Приведены базовые принципы расчета необходимого размера выборки для обеспечения необходимой мощности планируемого КИ, а также математические модели, описывающие нулевые и альтернативные гипотезы, используемые при разработке основных статистических дизайнов исследования эффективности и безопасности лекарственных препаратов. Показано, что общим требованием к качеству выборки субъектов исследования является обеспечение ее репрезентативности, то есть соответствие целевой популяции КИ. Выбор математической (вероятностной) модели, на основе которой формулируются исследовательские гипотезы и производится расчет выборки целевой популяции, базируется прежде всего на базовой информации о терапевтическом воздействии и специфических особенностях популяции, полученной из систематических обзоров результатов ранее проведенных исследований, а также в соответствии с классификацией исследуемого препарата. Для расчета размера выборки должны быть определены и обоснованы на этапе разработки дизайна и статистической модели КИ критерии в соответствии с общими требованиями к репрезентативности. Использование программных приложений для расчета мощности и требуемого размера выборки упрощает выполнение рутинных процедур планирования клинических исследований.

ВЫВОДЫ. Основных и базовых статистических моделей определения размера выборки недостаточно для проведения качественного исследования. Большое разнообразие дизайнов КИ, методологических подходов к планированию, реализации схем лечения, сбора и анализа данных КИ требует разработки статистических планов каждого конкретного КИ, включая оценку отдельных случаев, метода анализа выживания, относительного риска, диагностические тесты, адаптивные и другие нечасто используемые планы исследования. Следствием этого является востребованность в разработке дополнительных руководств и других информационных ресурсов, содержащих комментарии и примеры применения вероятностной статистики, и последующей гармонизации созданных национальных стандартов с международными.

Ключевые слова: клиническое исследование; выборка; популяция; размер выборки; дизайн исследования; гипотеза исследования; размер эффекта; статистическая модель

Для цитирования: Шредер О.В., Горячев Д.В., Меркулов В.А. Основные принципы расчета необходимой численности участников клинических исследований. Часть 1. Общие подходы (обзор). *Регуляторные исследования и экспертиза лекарственных средств*. 2024;14(3):338–350. <https://doi.org/10.30895/1991-2919-2024-14-3-338-350>

Финансирование. Работа выполнена в рамках государственного задания ФГБУ «НЦЭСМП» Минздрава России № 056-00026-24-00 на проведение прикладных научных исследований (номер государственного учета НИР № 124022300127-0).

Потенциальный конфликт интересов. Д.В. Горячев — член редакционной коллегии журнала «Регуляторные исследования и экспертиза лекарственных средств» с 2018 г. Остальные авторы заявляют об отсутствии конфликта интересов.

Olga V. Shreder ✉ 
Dmitry V. Goryachev 
Vadim A. Merkulov 

Basic Principles for Calculating the Required Number of Participants in Clinical Trials. Part 1. Common Approaches (Review)

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

✉ **Olga V. Shreder;** shrederov@expmed.ru

ABSTRACT

INTRODUCTION. A well-planned design of a clinical trial (CT) ensures valid results in assessing the efficacy and safety of medicines for human use. However, at present, there are no clear criteria for selecting the basic elements underlying the development of a CT design. This lack of selection criteria primarily concerns planning research hypotheses, calculating the expected therapeutic effect, statistical significance level, and study power, and selecting statistical models for sample size calculation.

AIM. The authors aimed to systematise and harmonise the technical requirements for sample size determination in designing CTs.

DISCUSSION. First, this article describes the basic requirements for and methodological approaches to designing CTs to assess the efficacy of medicines and to confirm their safety. Next, the article presents the basic principles for calculating the sample size to ensure the required CT power. Finally, the article covers the mathematical models describing the null and alternative hypotheses used in the development of basic statistical designs for efficacy and safety studies. A general requirement for the quality of a study sample is to ensure its representativeness, that is, its compliance with the target CT population. The selection of a mathematical (probabilistic) model to formulate research hypotheses and calculate study samples representative of the target population is based on general data from systematic reviews of previous studies on the therapeutic effects of the study product and the specific characteristics of the target population. In addition, model selection relies on the classification of the study product. Sample size calculation requires defining and justifying certain criteria at the stage of CT design and statistical model development, in line with the general requirements for representativeness. Software for calculating the statistical power and required sample size facilitates routine CT planning.

CONCLUSIONS. The sample size determination requires more than the application of basic statistical models. Given the multitude of CT designs and methodological approaches to CT planning, treatment regimens, and data collection and analysis, it is necessary to consider the statistical design of each CT on a case-by-case basis. This consideration should include assessments of individual cases, survival analysis methods, relative risks, diagnostic tests, and adaptive and other infrequent CT designs. The above highlights the need to develop additional guidelines and information resources that would explain and demonstrate the use of probabilistic statistics. The resulting national standards should be harmonised with international standards.

Keywords: clinical trial; study sample; population; sample size; study design; study hypothesis; effect size; statistical model

For citation: Shreder O.V., Goryachev D.V., Merkulov V.A. Basic principles for calculating the required number of participants in clinical trials. Part 1. Common approaches (review). *Regulatory Research and Medicine Evaluation*. 2024;14(3):338–350. <https://doi.org/10.30895/1991-2919-2024-14-3-338-350>

Funding. The study reported in this publication was carried out as part of publicly funded research project No. 056-00026-24-00 and was supported by the Scientific Centre for Expert Evaluation of Medicinal Products (R&D public accounting No. 124022300127-0).

Disclosure. Dmitry V. Goryachev has been a member of the Editorial Board of *Regulatory Research and Medicine Evaluation* since 2018. The other authors declare no conflict of interest.

ВВЕДЕНИЕ

Расчет размера выборки является неотъемлемой частью разработки дизайна большинства клинических исследований (КИ), обеспечивающей валидность, точность и надежность результатов КИ [1–4]. Исключением являются некоторые пилотные исследования, предназначенные для подтверждения целесообразности и возможности проведения более масштабного КИ, определения первичной информации, дающей необходимые представления об основных характеристиках популяции, а также методах сбора данных и расчета размера выборки участников КИ [1–3].

Необходимость научного обоснования размера выборки может быть продемонстрирована на следующих примерах [1, 2, 4]:

- если размер выборки в исследовании с отрицательными результатами достаточен для выявления клинически значимого эффекта, то отрицательные результаты КИ поддаются интерпретации;
- если исследование не продемонстрировало клинически значимый эффект лечения, то эти результаты не могут использоваться для расчета выборки при планировании нового исследования;
- если размер выборки в исследовании с отрицательными результатами недостаточен, то клинически важный (но статистически незначимый) эффект может быть проигнорирован и исследуемое лечение может интерпретироваться как неэффективное (бесполезное).

Проведение КИ должно быть экономически целесообразно, это достигается, в частности, корректным планированием размера выборки [3, 4]:

- исследование, проведенное на выборке небольшого размера, может не привести к обнаружению эффекта и повлечь неоправданные экономические потери, поскольку без достаточного размера выборки собранных данных может быть упущен (не выявлен) клинически значимый эффект, различие между группами или некая взаимосвязь, например, между

дозовыми режимами препарата и показателями излечения пациентов;

- исследование, проведенное на выборке чрезмерно большого размера, может привести к значительным экономическим затратам и при этом дать статистически значимые результаты, которые могут не иметь большого клинического или практического значения. Следует учитывать, что, если исследование основано на очень большой выборке, оно почти всегда приведет к статистически значимым результатам.

При проведении КИ важно придерживаться соблюдения этических принципов [3, 4]:

- исследование, планируемое на выборке небольшого размера, может подвергнуть участников бесполезному, иногда потенциально вредному терапевтическому воздействию без возможности получить клинически важные результаты;
- исследование, планируемое на выборке чрезмерно большого размера, может неоправданно подвергнуть большое число испытуемых потенциально вредному или бесполезному терапевтическому воздействию.

Таким образом, очевидна необходимость сочетания научной, экономической и этической составляющих в программе разработки эффективных и безопасных лекарственных средств (ЛС).

Цель работы — систематизация и гармонизация технических требований к планированию дизайна клинического исследования в части определения размера выборки.

ОСНОВНАЯ ЧАСТЬ

Общие требования, этапы и критерии для расчета выборки

Концепция определения размера выборки и оценки статистической мощности является основой планирования КИ и гарантией получения надежных выводов о клиническом эффекте ЛС. Ключевым инструментом в процессе

планирования и принятия решения о пользе и риске медицинского применения ЛС являются методы доказательной медицины, в частности вероятностной статистики, применяемые в программах клинической разработки в соответствии с требованиями отечественных и зарубежных нормативных актов¹.

Статистическая сущность этого принципа заключается в проверке исследовательских гипотез путем статистического анализа клинических данных [1–12]. Нулевая гипотеза (H_0) в большинстве случаев формулируется в предположении об отсутствии различий между сравниваемыми группами терапии в эффективности медицинского вмешательства и утверждает случайный характер их обнаружения. Подтверждением того, что наблюдаемое различие между группами неслучайно и может рассматриваться как весомый аргумент для отклонения нулевой гипотезы в пользу альтернативной гипотезы (H_1), является установление величины разницы, обоснованной при планировании исследования как «величина клинически значимого эффекта». Следует учитывать, что результат определения только статистической значимости различий между группами сравнения без подтверждения наличия искомой величины разницы, обоснованной на этапе планирования как «клинически значимый эффект», не является корректным представлением доказательства альтернативной гипотезы клинического исследования.

Одним из критериев отклонения H_0 является значение уровня статистической значимости, которое в зависимости от выбранного дизайна КИ устанавливается на уровне 0,05 при двустороннем и 0,025 при одностороннем тестировании [2, 4, 5]. Процесс отклонения нулевой гипотезы сопряжен с потенциальными ошибками I и II рода [1–34]:

- риск ошибки I рода (α) связан с вероятностью неправильного отклонения H_0 , то есть ассоциируется с ложноположительным результатом принятия H_1 ;
- риск ошибки II рода (β) связан с вероятностью неправильного отклонения H_1 , то есть ассоциируется с ложноотрицательным результатом принятия H_0 .

Для исключения таких ошибок в зависимости от условий тестирования статистической

модели клинического исследования определяют значения α - и β -ошибок. Риск ошибки I рода (α) обычно устанавливается равным $p=0,05$ или $p=0,025$, а риск ошибки II рода (β) – 10–20%. Поскольку заданное значение разницы или эффекта может быть и выше, и ниже нуля (гипотеза H_0), риск ошибки β всегда односторонний. Чем меньше риск ошибки β , тем больше статистическая мощность ($P=1-\beta$) – вероятность обнаружения разницы между группами терапии, если она действительно существует, и верного принятия гипотезы H_1 . Статистическая мощность обычно устанавливается на уровне не ниже 80% при допущении 20%-ной вероятности ошибочного принятия нулевой гипотезы ($\beta=0,20$) [2–6].

Общим требованием к качеству выборки субъектов исследования является обеспечение ее репрезентативности, то есть соответствие целевой популяции КИ [2–17].

Выбор математической (вероятностной) модели, на основе которой формулируются исследовательские гипотезы и производится расчет выборки целевой популяции, базируется прежде всего на базовой информации о терапевтическом воздействии, специфических особенностях популяции (демографические характеристики, состояние здоровья, особенности течения заболевания, симптоматические характеристики и т.д.), полученной из систематических обзоров, результатов ранее проведенных исследований, а также классификации исследуемого ЛП (референтный, оригинальный, воспроизведенный, биоаналоговый). Другими факторами, влияющими на определение модели, являются фаза, дизайн, цели, задачи КИ, конечные точки эффективности и безопасности, методы статистической оценки. Выбор статистической модели в большинстве случаев осуществляется из перечня базовых формул (табл. 1 «Классификация и дизайны клинических исследований», опубликована на сайте журнала², табл. 2).

Требование надежности имеет отношение к ожидаемым результатам планируемого КИ в целом и формированию выборки на основе предварительно полученных данных из пилотных исследований или литературных сведений в частности и включает:

- возможность повторения эксперимента с получением сходных результатов;

¹ Рекомендации Коллегии ЕЭК от 03.11.2020 № 19 «О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов».

Федеральный закон Российской Федерации от 12.04.2010 № 61-ФЗ «Об обращении лекарственных средств».

ICH Harmonised tripartite guideline. Statistical principles for clinical trials E9, 1998.

ICH Harmonised tripartite guideline. Structure and content of clinical study reports E3, 1995.

² <https://doi.org/10.30895/1991-2919-2024-14-3-338-350-tabl1>

Таблица 2. Правила определения размера выборки для сравнения средних значений и пропорций (по L. Thabane [2] с изменениями)**Table 2.** Guidelines for sample size determination for comparisons between means and proportions (adapted from L. Thabane [2])

Дизайн Design	Гипотеза Hypothesis	Формулы для сравнения средних Formulae for comparisons between means			Формулы для сравнения пропорций Formulae for comparisons between proportions		
		H ₀	H ₁	Формулы Formulae	H ₀	H ₁	Формулы Formulae
Одновыборочный One-sample	Неравенство средних (пропорций) Equality of means (proportions)	$\mu - \mu_0 = 0$	$\mu - \mu_0 \neq 0$	$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu - \mu_0)^2}$	$\pi - \pi_0 = 0$	$\pi - \pi_0 \neq 0$	$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1 - \pi)}{(\pi - \pi_0)^2}$
	Превосходство Superiority	$\mu - \mu_0 \leq \delta$	$\mu - \mu_0 > \delta$	$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$	$\pi - \pi_0 \leq \delta$	$\pi - \pi_0 > \delta$	$n = \frac{(z_{\alpha} + z_{\beta})^2 \pi(1 - \pi)}{(\pi - \pi_0 - \delta)^2}$
	Эквивалентность Equivalence	$ \mu - \mu_0 \geq \delta$	$ \mu - \mu_0 < \delta$	$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$	$ \pi - \pi_0 \geq \delta$	$ \pi - \pi_0 < \delta$	$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \pi(1 - \pi)}{(\pi - \pi_0 - \delta)^2}$
Параллельный с двумя выборками Two-sample parallel	Неравенство средних (пропорций) Equality of means (proportions)	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2)^2}$	$\pi_1 - \pi_2 = 0$	$\pi_1 - \pi_2 \neq 0$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 (\pi_1(1 - \pi_2) + \pi_2(1 - \pi_1))}{(\pi_1 - \pi_2)^2}$
	Не меньшей эффективности Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{2(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$	$\pi_1 - \pi_2 \geq \delta$	$\pi_1 - \pi_2 < \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 (\pi_1(1 - \pi_2) + \pi_2(1 - \pi_1))}{(\pi_1 - \pi_2 - \delta)^2}$
	Превосходство Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{2(z_{\alpha} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$	$\pi_1 - \pi_2 \leq \delta$	$\pi_1 - \pi_2 > \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 (\pi_1(1 - \pi_2) + \pi_2(1 - \pi_1))}{(\pi_1 - \pi_2 - \delta)^2}$
	Эквивалентность Equivalence	$ \mu_1 - \mu_2 \geq \delta$	$ \mu_1 - \mu_2 < \delta$	$n_i = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$	$ \pi_1 - \pi_2 \geq \delta$	$ \pi_1 - \pi_2 < \delta$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 (\pi_1(1 - \pi_2) + \pi_2(1 - \pi_1))}{(\pi_1 - \pi_2 - \delta)^2}$
Перекрестный с двумя выборками Two-sample crossover	Неравенство средних (пропорций) Equality of means (proportions)	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma_d^2}{2(\mu_1 - \mu_2)^2}$	$\pi_1 - \pi_2 = 0$	$\pi_1 - \pi_2 \neq 0$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma_d^2}{2(\pi_1 - \pi_2)^2}$
	Не меньшей эффективности Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 \sigma_d^2}{2(\mu_1 - \mu_2 - \delta)^2}$	$\pi_1 - \pi_2 \geq \delta$	$\pi_1 - \pi_2 < \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$
	Превосходство Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 \sigma_d^2}{2(\mu_1 - \mu_2 - \delta)^2}$	$\pi_1 - \pi_2 \leq \delta$	$\pi_1 - \pi_2 > \delta$	$n_i = \frac{(z_{\alpha} + z_{\beta})^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$
	Эквивалентность Equivalence	$ \mu_1 - \mu_2 \geq \delta$	$ \mu_1 - \mu_2 < \delta$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma_d^2}{2(\mu_1 - \mu_2 - \delta)^2}$	$ \pi_1 - \pi_2 \geq \delta$	$ \pi_1 - \pi_2 < \delta$	$n_i = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$

Таблица адаптирована авторами из [2] / The table is adapted by the authors from [2]

Примечание. H₀ – нулевая гипотеза; H₁ – альтернативная гипотеза; n – размер выборки; σ – стандартное отклонение; δ – величина разницы (μ – μ₀) или (μ₁ – μ₂) или (π₁ – π₂), минимальная клинически значимая разность; μ – среднее значение распределения переменной отклика (ответа на лечение); μ₀ – исходное (базовое) среднее значение; μ₁ – среднее значение нового лечения; μ₂ – среднее значение стандартного лечения; (μ₁ – μ₂) – величина разницы, имеющей клиническое значение; π₁ – первая доля (значение нового лечения); π₂ – вторая доля (значение стандартного лечения); (π₁ – π₂) – величина разницы, имеющей клиническое значение; σ_d – стандартное отклонение распределения d (где d – изменение по сравнению с исходным уровнем); z_α – величина для подстановки в формулу расчета выборки – зависит от выбора желаемого уровня значимости (см. табл. 3).

Note. H₀, null hypothesis; H₁, alternative hypothesis; n, sample size; σ, standard deviation; δ, size of difference of clinical importance between (μ – μ₀), (μ₁ – μ₂), or (π₁ – π₂) (minimum clinically meaningful difference); μ, mean value of the distribution of the response variable (response to treatment); μ₀, corresponding baseline mean value; μ₁, mean value of the new treatment; μ₂, mean value of the standard treatment; (μ₁ – μ₂), size of difference of clinical importance; π₁, first proportion (value of the new treatment); π₂, second proportion (value of the standard treatment); (π₁ – π₂), size of difference of clinical importance; σ_d, standard deviation of the distribution of d (where d is the change from baseline); z_α, value to be substituted into the sample size calculation formula, depending on the desired significance level (see Table 3).

- уровень доверия к ожидаемым или полученным результатам;
- данные о точности («уровне статистической значимости»), с которой были получены результаты исследования.

Перечень критериев, необходимых для расчета репрезентативной выборки [1–6]:

- показатель вариабельности основного показателя исследования в числовой шкале измерений. Таким показателем, как правило, является стандартное отклонение, если не обоснована другая статистическая характеристика вариабельности;
- размер эффекта, или «наименьший клинически значимый эффект». Обычно выражается как разность средних значений основного показателя эффекта или разность величин долей ответа на терапию (%). Реже вместо разности используется отношение данных показателей. При отсутствии результатов предыдущих исследований или данных литературы размер ожидаемого эффекта может выражаться в значениях стандартного отклонения в виде стандартизированной разности интересующего показателя;
- уровень статистической значимости (α) — пороговое значение 0,05 при двустороннем и 0,025 при одностороннем тестировании гипотезы. Ниже указанного значения отклоняют H_0 и принимают H_1 при условии подтверждения наличия разницы между группами, соответствующей пороговой величине «превосходства», «эквивалентности» или «не меньшей эффективности», обоснованной при планировании соответствующего дизайна КИ;
- статистическая мощность ($P=1-\beta$) — вероятность обнаружения разницы между группами терапии, если она действительно существует. При допущении 20%-ной вероятности ошибки ($\beta=0,20$) мощность устанавливается не менее 80% ($P\geq 0,80$) [2–6].

Вышеуказанные критерии должны быть определены и обоснованы на этапе разработки дизайна и статистической модели КИ (табл. 1 «Классификация и дизайны клинических исследований», опубликована на сайте журнала³, табл. 2) [1, 2, 10–18, 20–32].

Таким образом, критериями определения репрезентативной выборки для КИ являются [2, 4, 5]:

- изменчивость основного параметра оценивания (чем разнообразней показания, заявляемые для регистрации, тем больше наблюдений (участников) нужно включить в КИ);

- клинически важный размер эффекта (чем меньше величина эффекта основного критерия оценивания, тем больше наблюдений (участников) необходимо);
- условия тестирования H_0 и H_1 — определяют целями и задачами планируемого КИ;
- статистическая мощность теста (уровень вероятности, при которой следует отвергнуть H_0).

Изменчивость основного параметра оценивания в большинстве случаев характеризуется величиной стандартного отклонения, определенной на основании данных литературы, систематических обзоров на основе метаанализа (Кокрейновские обзоры) или собственных пилотных исследований, включающих результаты сравнения группирующих средних значений основных показателей эффективности терапии [2, 4, 6, 17]. При определении величины изменчивости основного параметра оценивания из данных источников литературы необходимо учитывать, что большое значение стандартного отклонения указывает на недостаточную значимость эффекта и приведет к формированию выборки большого размера.

Определение наименьшего размера клинически значимого эффекта должно быть основано на данных систематизированных источников, результатов аналогичных клинических исследований, Кокрейновских обзоров, экспертных заключений и клинических рекомендаций или данных собственных пилотных исследований, включающих статистически значимые результаты оценки основных показателей сравнительной фармакотерапии. В случае отсутствия информации о наименьшей клинически важной величине в подходящей для расчета выборки метрической шкале или для обоснования таковой для изучения нового оригинального ЛС могут потребоваться дополнительные перерасчеты опубликованных сходных результатов из доступных источников литературы и приведение их в подходящий формат с последующим обоснованием применимости для определения размера выборки участников планируемого дизайна или пилотного КИ [2, 4, 6–8, 17].

Еще одним способом определения клинически значимого размера эффекта, например при планировании исследования нового оригинального ЛС или при отсутствии какой бы то ни было информации, является его оценка на основе известных стандартизированных значений, названных J. Cohen «малый», «средний» и «большой» размеры эффекта [4, 8, 9]:

³ <https://doi.org/10.30895/1991-2919-2024-14-3-338-350-tabl1>

- при тестировании разницы между двумя средними размерами эффекта (d)⁴ составляют: малый — 0,20; средний — 0,50; большой — 0,80;
- при тестировании разницы между несколькими средними размерами эффекта (f)⁵ составляют: малый — 0,10; средний — 0,25; большой — 0,40;
- при тестировании разницы между долями с использованием критерия χ -квадрат и коэффициента корреляции Пирсона размеры эффекта (ω или r)⁶ составляют: малый — 0,10; средний — 0,30; большой — 0,50.

Следует учитывать, что размер эффекта зависит от дизайна КИ и критериев оценивания основного показателя, поэтому необходимо обосновать выбор объективной величины эффекта в соответствии со шкалой измерения, использующейся для тестирования исследовательских гипотез.

Условия тестирования нулевой и альтернативной гипотезы

Существуют два варианта тестирования нулевой (H_0) и альтернативной (H_1) гипотез (табл. 3, рис. 1) [2, 9, 17, 18]. В случае когда ожидается разница в одном направлении (правостороннее или левостороннее тестирование), используется одностороннее тестирование. Двустороннее тестирование подходит, когда ожидается разница в любом направлении [2, 17, 18]. Если критическое значение задается величиной δ , то характер распределения и положение критической области для H_0 и H_1 выглядят следующим образом:

$\mu \leq \delta$ $\mu > \delta$: критическая область справа — одностороннее тестирование;

$\mu \geq \delta$ $\mu < \delta$: критическая область слева — одностороннее тестирование;

$\mu = \delta$ $\mu = \delta$: две критические области — двустороннее тестирование.

При одностороннем тестировании исследовательской гипотезы требуется выборка меньших размеров, чем при двустороннем. Однако

возможность использования односторонних тестов для расчета должна быть обоснована, не допускается использовать такой тип тестирования только в целях уменьшения размера выборки [12].

Одностороннее направление тестирования не является достаточным основанием, чтобы использовать односторонний критерий для расчета численности субъектов исследования. Такой тип тестирования целесообразно использовать при планировании плацебо-контролируемых исследований или при сравнении комбинированной схемы лечения и монотерапии в КИ с дизайном превосходства [10, 11]. Если нет объективных причин для использования одностороннего тестирования, следует пользоваться двусторонней гипотезой.

При планировании дизайнов «не меньшей эффективности» и «эквивалентности» терапии двустороннее тестирование предполагает, что H_0 заключается в отсутствии различий, а H_1 предполагает, что различия между группами могут проявляться в любом направлении.

Определение уровня ошибки I рода. Альфа (α)-ошибка I рода — это ошибочное заключение о существовании различий, которых в действительности нет. За вероятность ошибки I рода принимается p -значение 0,05 или 0,025 в зависимости от дизайна КИ и условия тестирования гипотез (одно- или двустороннее тестирование). Нулевая гипотеза отклоняется в случаях, если величина p меньше 0,05 или 0,025. Величина p , определяемая как уровень значимости, является тем значением, ниже которого H_0 отклоняется и принимается H_1 о наличии доказательств эффекта. В большинстве случаев двустороннего тестирования уровень статистической значимости выбирается равным $\alpha=0,05$, или 5%, а при одностороннем тестировании $\alpha=0,025$, или 2,5%. Выбор большей величины альфа ($\alpha=0,05$) приводит к уменьшению размера численности субъектов для исследования.

⁴ d — стандартизованная средняя разница между двумя выборками, деленная на объединенное стандартное отклонение; $(\mu_1 - \mu_2)/\sigma$, где μ_1 и μ_2 — средние значения эффекта, σ — объединенная оценка дисперсии (s) выборок 1 и 2 соответственно $=\sqrt{(s_1^2 + s_2^2)/2}$.

⁵ f^2 — стандартизованная средняя разница между несколькими выборками при использовании F-теста для ANOVA или множественной регрессии;

f^2 для множественной регрессии: $f^2 = R^2/(1 - R^2)$, где R^2 — квадрат множественной корреляции;

f^2 для дисперсионного анализа (ANOVA) в сбалансированном дизайне с одинаковыми размерами выборок по группам: $f^2 = SS(\mu_1, \mu_2, \dots, \mu_k)/(K \cdot \sigma^2)$, где μ_j ($\mu_1, \mu_2, \dots, \mu_k$) — средние значения эффекта выборки в j -й группе из общего числа K групп, σ — объединенная оценка стандартных отклонений в каждой группе, SS — сумма квадратов в ANOVA.

⁶ r — коэффициент корреляции Пирсона, определяется как размер эффекта парных количественных данных при использовании анализа взаимосвязи между двумя переменными: диапазон величин от -1 до 1, где, 0 — отсутствие линейной связи, -1 — отрицательная линейная связь, 1 — положительная линейная связь между двумя переменными;

r^2 — коэффициент детерминации, определяется как квадрат корреляции Пирсона парных данных и используется для оценки доли дисперсии между двумя переменными: диапазон величин от 0 до 1, где r^2 имеет положительное значение;

ω — размер эффекта категориальных данных при использовании теста χ -квадрат: $\omega = \sqrt{\sum_{i=1}^m (p_{1i} - p_{0i})^2 / p_{0i}}$, где p_{0i} — доля i -й группы для H_0 , p_{1i} — доля i -й группы для H_1 , m — количество групп.

Таблица 3. Таблица значений критериев, необходимых для расчета выборки в зависимости от условия тестирования гипотезы [1–6, 18]**Table 3.** Tabulated test statistics required for sample size calculation based on the hypothesis testing conditions [1–6, 18]

Условие тестирования гипотезы <i>Hypothesis testing condition</i>	Доверительный интервал <i>Confidence interval</i>	Ошибка I рода (α), % <i>Type I error (α), %</i>	Ошибка II рода (β), % <i>Type II error (β), %</i>	α	z_α и $z_{\alpha/2}$ (для одно- и двухстороннего тестирования соответственно) z_α and $z_{\alpha/2}$ (for one- and two-sided tests, respectively)	β	z_β	Мощность теста ($P=1-\beta$), % <i>Test power ($P=1-\beta$), %</i>	Уровень статистической значимости, p <i>Statistical significance, p</i>
Двусторонний критерий <i>Two-sided test</i>	90	10	20	0,1	1,64	0,2	0,84	80	<0,1
	95	5	20	0,05	1,96	0,2	0,84	80	<0,05
	99	1	20	0,01	2,58	0,2	0,84	80	<0,01
Односторонний критерий <i>One-sided test</i>	95	5	20	0,1	1,64	0,2	0,84	80	<0,1
	97,5	2,5	20	0,025	1,96	0,2	0,84	80	<0,025
	99,5	0,5	20	0,005	2,58	0,2	0,84	80	<0,005

Таблица составлена авторами по данным литературы / The table is prepared by the authors on the basis of published data

Примечание. Примеры применения критериев, необходимых для расчета выборки, в концепции интервального оценивания.

А) Односторонний тест концентрирует 5% область отклонения гипотезы в одном хвосте нормального распределения (z -оценка 1,64 или выше). При двустороннем тестировании в случае такого же доверительного интервала z -оценка составит $\pm 1,96$, так как 5% составляют 2,5% в каждом из двух хвостов стандартного распределения (рис. 1).

Б) $z_{\alpha/2}$ для тестирования, уровень доверия⁷ которого соответствует 90%, определяется следующим образом: уровень значимости $\alpha/2$ составит $1-0,9=0,1$, $\alpha/2=0,1/2=0,05$. Критическое значение z определяется по таблице критических значений z [4], для абсолютной величины $\alpha/2=0,05$ составит 1,64 (при одностороннем тестировании гипотезы $z_{0,1/2}=1,64$). При условии выбора уровня доверия 95% α составит $1-0,95=0,05$, то есть $\alpha/2=0,05/2=0,025$ и соответствующее критическое значение z , для абсолютной величины $\alpha/2=1,96$ (при одностороннем тестировании гипотезы $z_{0,05/2}=1,96$).

В) При условии выбора уровня доверия 99% α составит $1-0,99=0,01$, то есть $\alpha/2=0,01/2=0,005$, соответствующее критическое значение z для абсолютной величины $\alpha/2=2,58$ (при одностороннем тестировании гипотезы $z_{0,01/2}=2,58$).

Note. Examples of applying the tests required for sample size calculation based on confidence interval estimation.

А) A one-sided test concentrates the 5% rejection region in one tail of a normal distribution (z -score of 1.64 or greater). A two-sided test for the same confidence interval would result in a z -score of ± 1.96 because the 5% rejection region is comprised of 2.5% regions in each of the two tails (Fig. 1).

Б) $z_{\alpha/2}$ is determined as follows: For a test that is using a 90% confidence level, the significance level α would be $1-0.9=0.1$, and $\alpha/2=0.1/2=0.05$. The corresponding critical value of z is determined using a table of critical values of z [4]. The critical value of z corresponding to the absolute value of $\alpha/2=0.05$ would be 1.64 (for one-sided hypothesis testing $z_{0,1/2}=1.64$). For a test that is using a 95% confidence level, the significance level α would be $1-0.95=0.05$, that is, $\alpha/2=0.05/2=0.025$. The critical value of z corresponding to this absolute value of $\alpha/2$ would be 1.96 (for one-sided hypothesis testing $z_{0,05/2}=1.96$).

С) For a test that is using a 99% confidence level, the significance level α would be $1-0.99=0.01$, that is, $\alpha/2=0.01/2=0.005$, and the critical value of z corresponding to this absolute value of $\alpha/2$ would be 2.58 (for one-sided hypothesis testing $z_{0,01/2}=2.58$).

Статистическая мощность исследования — критерий корректного отклонения H_0 в случаях подтверждения различия между группами сравнения. Этот критерий рассчитывается по формуле $P=1-\beta$ (или $100-\beta$ (%)), где вероятность ошибки II рода β устанавливается равной не более 20% (соответственно, величина статистической мощности должна быть не ниже 80%). При корректном учете статистической мощности вероятность случайной ошибки в установлении истинной клинически значимой разницы

в эффектах снижается и увеличивается вероятность получения статистически значимого результата исследования.

Размер выборки участников КИ может рассчитываться несколькими способами, при этом во всех случаях применяется одинаковый подход к выбору базовых элементов расчета [3–5, 8, 12–15]:

- общий набор формул для расчета размера выборки участников в разных дизайнах КИ [16, 17];

⁷ Уровень доверия $\beta=1-\alpha$ — это вероятность попадания значения тестовой статистики в доверительный интервал, который с заданной вероятностью покрывает оцениваемый показатель исследования. Уровни доверия 0,99; 0,95; 0,9 соответствуют уровням значимости 0,01; 0,05; 0,1; уровни значимости и доверия могут быть представлены в процентах и десятичных значениях.

⁸ Уровень значимости α — заданный уровень статистической значимости — «пороговое значение», относительно которого определяется вероятность (p -значение) непопадания в доверительный интервал значений, если значение тестовой статистики больше или меньше z -показателя альфа-уровня (или p -значение больше или меньше альфа-значения) в зависимости от условий тестирования гипотез исследования.

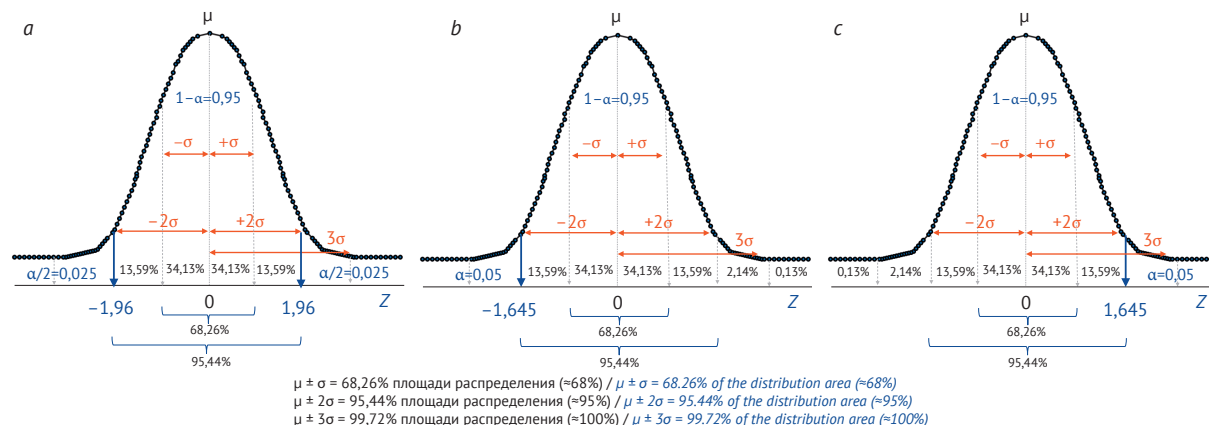


Рисунок подготовлен авторами по собственным данным / The figure is prepared by the authors using their own data

Рис. 1. Условия тестирования гипотезы: а – двустороннее тестирование; б – левостороннее тестирование; с – правостороннее тестирование

Fig. 1. Hypothesis testing conditions: a, two-sided testing; b, left-sided testing; c, right-sided testing

- возможность использования «быстрых формул»⁹ для расчета размера выборки участников при соблюдении требований по выбору уровня значимости, статистической мощности исследования и расчету стандартизированной разницы [13, 14];
- возможность использования общедоступных таблиц, содержащих стандартизированные значения для расчета размера выборки и проверки исследовательских гипотез, на основе работ J. Cohen [8] или сводных данных об ожидаемых размерах эффекта при использовании непарного *t*-критерия или критерия χ^2 -квадрат Пирсона [16, 17];
- использование общедоступной диаграммы «номограмма Альтмана», устанавливающей связь выборки, мощности статистического критерия, уровня значимости и стандартизированной разности, которая применима для различных статистических дизайнов КИ [3, 11].

Методологические подходы к разработке дизайна клинических исследований

Основные этапы определения необходимой численности участников клинических исследований включают [2–6]: 1) выбор дизайна, соответствующего фазе, цели и задачам КИ; 2) выбор гипотезы и метода расчета или обоснования размера выборки в соответствии с целью КИ, данными литературы и регистрационным статусом действующего вещества исследуемого ЛС; 3) обоснование достаточности размера выборки путем аргументации ожидаемой пороговой

величины различий в эффектах, выбора оценочных критериев, данных об экономической целесообразности (если применимо), особенностях популяции исследования и др.; 4) корректировку размера выборки в случаях проведения промежуточного анализа или в соответствии с правилами адаптивных дизайнов, предусматривающих увеличение размера выборки по результатам промежуточной оценки или учет выбывших из исследования субъектов.

Проведение рандомизированных клинических исследований является основным требованием при регистрации ЛС [15, 16, 18]. В практике проведения КИ наибольшее распространение получили три статистических дизайна (табл. 1 «Классификация и дизайны клинических исследований», опубликована на сайте журнала¹⁰, табл. 2): превосходства, не меньшей эффективности, эквивалентности, которые включают [12, 18–33]:

- неконтролируемые планы — дизайны, которые часто выбирают при проведении КИ I и II фаз и пострегистрационных исследований IV фазы;
- контролируемые планы — дизайны, которые часто выбирают при проведении КИ III фазы, в частности рандомизированные схемы параллельного и перекрестного группового анализа.

Исследование превосходства. Целью исследования является доказательство преимущества в терапевтическом эффекте сравниваемых групп лечения.

⁹ Быстрые формулы используются для ориентировочного расчета выборки участников при соблюдении определенных требований к выбору значений мощности, уровня значимости и величины клинически значимого эффекта, часто встречающихся в биофармацевтических исследованиях, и табулирования функции (пошаговое вычисление значений функции); например, допускается ориентировочная мощность 0,80 с уровнем альфа=0,05: для двустороннего *t*-теста, $n=16s^2/d^2$, где n — размер каждой выборки, s^2 — дисперсия популяции, d — разница в эффектах $=4sn$; для точного критерия Фишера или критерия χ^2 -квадрат: $s^2=pq$, где $p=(p_1+p_2)/2$ и $q=1-p$ и $d=|p_1-p_2|$.

¹⁰ <https://doi.org/10.30895/1991-2919-2024-14-3-338-350-tabl1>

Эталонным вариантом такого КИ является дизайн исследования терапевтической эффективности нового ЛС по сравнению с плацебо [17–24]. По ряду медицинских и этических причин такая схема не всегда применима, поэтому на практике часто осуществляют сравнение с группой пациентов, получающих стандартную терапию с использованием референтного препарата. Насколько терапия новым препаратом должна быть лучше плацебо или стандартной терапии, определяется пороговым значением, характеризующим наименьшую величину клинически значимой разницы терапевтического эффекта между сравниваемыми группами лечения (табл. 2).

На этапе расчета выборки необходимо предусмотреть визит оценки лечения, указать ожидаемую величину разницы (пороговое значение) для подтверждения превосходства, значение вариабельности основного показателя, величины ошибок I и II рода, приемлемые для исследования, с учетом экономической целесообразности и этических требований к количеству участвующих в исследовании пациентов.

Величина эффекта (δ) лечения подвержена значительным изменениям (вариабельности). Учитывая, что вариабельность основных показателей эффективности терапии может оказывать влияние на разницу в эффектах лечения в двух группах терапии, необходимо обосновать величину ожидаемого стандартного отклонения σ или дисперсии σ^2 основного параметра на основе результатов пилотного исследования или опубликованных данных аналогичных исследований [1, 2, 4].

Исследование эквивалентности. Целью исследования эквивалентности является установление идентичности эффектов сравниваемых методов лечения (нулевое значение разницы между группами) [25–30]. Поскольку величина разницы внесена в знаменатель формулы расчета выборки (табл. 2), а деление на ноль невозможно, целью исследования эквивалентности будет определение того, лежит ли разница в эффектах между двумя методами лечения в пределах определенного небольшого интервала от $-\delta$ до $+\delta$.

На этапе планирования важно указать величину разницы в эффектах лечения, которая будет являться пороговой величиной для принятия решения об эквивалентности [27–30]. В случае тестирования нового (оригинального) препарата лечения целесообразно максимально ограничить риски побочных эффектов от приема новой терапии, которая уступает стандартной терапии,

поэтому в качестве δ следует указывать допустимо малое значение, которое в то же время демонстрировало бы клинически значимый размер эффекта. По общему правилу δ может составлять не более половины величины, которая может использоваться в исследовании превосходства [26].

В исследовании эквивалентности H_0 состоит в том, что существует разница не менее δ , а цель исследования – в том, чтобы отклонить H_0 и принять H_1 для доказательства отсутствия разницы в эффекте между методами лечения [26]. В дизайнах исследования превосходства и эквивалентности формулировка исследовательских гипотез (табл. 2) и метод оценки размера выборки имеют сходство, однако интерпретация определяемой величины δ и выводы по результатам КИ различны:

Доказательство гипотезы эквивалентности предполагает двустороннее тестирование эффективности по основному показателю терапевтического ответа на лечение.

Исследование не меньшей эффективности. Целью исследования не меньшей эффективности является доказательство того, что новая терапия не менее эффективна, чем стандартное лечение, которое применяется при изучаемой патологии. Следует доказать, что разница в терапевтическом эффекте тестируемого и референтного препарата не выходит за пределы величины δ , установленной при планировании как пороговая величина, подтверждающая одинаковую эффективность лечения пациентов. По условию доказательства не меньшей эффективности новой терапии необходимо указать, что нижняя граница доверительного интервала для разницы в эффекте между сравниваемыми группами лечения находится выше установленной при планировании пороговой величины δ . Положение верхней границы доверительного интервала в большинстве случаев не рассматривается как пороговая величина. В отдельных случаях это может применяться, если предполагается, что новая терапия может продемонстрировать лучший эффект, чем стандартная (референтная) терапия, однако размера ожидаемого эффекта недостаточно для определения превосходства.

Доказательство гипотезы не меньшей эффективности предполагает одностороннее тестирование (табл. 3). По этой причине необходимое количество пациентов в таком исследовании будет меньше, чем в случае исследования эквивалентности [31].

Компьютерное программное обеспечение для расчета мощности и требуемого размера выборки

Для статистического анализа, оценки мощности и расчета размера выборки для различных дизайнов клинического исследования могут быть использованы следующие программные приложения [32]:

- пакеты прикладных программ для расчета мощности и размера выборки: GPOWER^{***}, NQUERY ADVISOR^{***}, PASS^{***}, PC-SIZE (CONSULTANT)*, POWER AND PRECISION^{***}, POWPAL*, POWERPACK^{**}, POWER PLANT^{**}, PS*, STATISTICA^{**}, POWER ANALYSIS^{**}, STAT POWER^{***}, STPLAN*;
- статистические пакеты прикладных программ общего назначения: DATASIM^{**}, JMP^{**}, MacANOVA*, MSUSTAT*, NCSS*, SHAZAM*, SigmaSTAT^{**}, STATISTICA*, STATA*, TRUE EPISTAT^{**}, WDIST*;
- пакеты специализированных прикладных программ для расчета размера выборки: EX-SAMPLE^{**}, POWER AND EFFECT^{**};
- пакеты специализированных прикладных программ для определения мощности: MONITOR^{**}, NSURV^{**}, TRENDS*;
- приложения для решения общих универсальных задач: DATASIM^{**}, MS EXCEL^{***}, GLIM*, SHAZAM*, SAS*, SIMSTAT, S-PLUS, XLISP-STAT.

Простота изучения и использования указанных программ варьирует от «сложно» (*), «удовлетворительно» (**) до «хорошо» (***) и «отлично» (****).

Программные пакеты привлекательны тем, что результаты расчета размера выборки и мощности можно представить в формате готовых отчетов (листингов), включающих формулировки исследовательских гипотез, значения параметров и статистические характеристики, влияющие на изменение объема выборки и требуемые для обоснования численности субъектов исследования (например, мощности, величины эффекта и другие взаимосвязанные критерии), что упрощает выполнение рутинных процедур планирования КИ.

Представление статистического плана в протоколе исследования

Расчет размера выборки субъектов осуществляется на этапе планирования дизайна и обобщается в протоколе исследования. Выбор

метода расчета размера выборки зависит от типа первичной конечной точки (непрерывной, категориальной и др.) и дизайна исследования (перекрестные, параллельные группы и др.).

Статистический план клинического исследования должен содержать следующую информацию:

- краткое описание целей исследования и дизайна исследования;
- первичные и вторичные конечные точки; гипотезы, подлежащие тестированию;
- уровень значимости и мощности исследования;
- подробное описание методов анализа, которые будут использоваться для анализа результатов и тестирования исследовательских гипотез;
- условия тестирования исследовательских гипотез (одно- или двустороннее);
- обоснование величины наименьшей клинически значимой разницы в эффектах и порогового значения для отклонения H_0 в пользу H_1 ;
- обоснование величин изменчивости (вариативности) основного эффекта, использованных в расчете размера выборки;
- обоснование продолжительности исследования и визита оценки терапевтического эффекта;
- обоснование коррекции потерь и методов восстановления данных выбывших субъектов исследования.

Теоретические аспекты теории вероятности (байесовский подход)

Согласно рекомендациям Коллегии ЕЭК¹¹, для обоснования ожидаемого эффекта терапевтического воздействия целесообразно использовать байесовский подход к анализу данных ранее проведенных исследований, то есть на основе априорной вероятности (вероятность правильности гипотезы до наступления события) осуществлять планирование проверки статистических гипотез и определение доверительного интервала вероятности апостериорного распределения интересующего эффекта вмешательства для принятия решения о доказательстве эффективности исследуемой терапии¹².

Байесовский подход к определению размера выборки учитывает подходящие результаты распределения эффекта лечения предшествующего

¹¹ Рекомендации Коллегии Евразийской экономической комиссии от 03.11.2020 № 19 «О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов».

¹² Рекомендации Коллегии Евразийской экономической комиссии от 03.11.2020 № 19 «О Руководстве по применению принципов биостатистики в клинических исследованиях лекарственных препаратов».

Методические рекомендации «Применение математического моделирования в доклинических исследованиях в целях проведения клинических исследований с участием особых популяций», разработанные в рамках рабочей программы НИР ФГБУ НЦЭСМП по этапу № 2 НИР «Разработка руководств по проведению научной экспертизы новых и зарегистрированных лекарственных препаратов для медицинского применения». НИР 121021800098-4 от 06.12.2021.

исследования и позволяет рассчитать прогнозируемые результаты распределения для тестовой статистики и, следовательно, ожидаемые потери для любого заданного размера выборки [33–35, 37, 40, 41]. Такой подход является процедурой двойной оптимизации, позволяющей определить оптимальное решение для заданной величины тестовой статистики и размера выборки, а затем определить размер выборки, который дает наименьшие минимальные потери [34, 35]. Подход D.V. Lindley предполагает использование функций потерь при неправильном принятии решений на основе наблюдаемых данных и байесовский алгоритм, но применяется, когда имеют сходные представления и предшествующие значения исходов [33, 34]. Система J. Gittins и Н. Pezeshk представляет собой гибридную байесовско-частотную систему, согласно которой предполагается, что будущие реализации лекарственного препарата зависят от того, насколько впечатляющими являются результаты клинических испытаний (измеряемые значимостью различий), но при планировании используются предшествующие результаты распределения по терапевтическому эффекту и байесовский подход [35, 36]. Метод расчета О'Наган и соавт. — это обеспечение гарантий на основе байесовской вероятности, что КИ даст значимый результат [37]. Такой подход исключает зависимость от конкретного предполагаемого клинически значимого различия, но при этом использует предшествующее распределение фармакологического эффекта, чтобы гарантированно получить ожидаемую мощность.

Байесовские подходы являются вариантами прогнозирования результатов распределения для тестовой статистики на основе предшествующих исходов, обеспечения гарантий надежности и точности оценок, включая результаты доклинических испытаний, и принятия решений для заданной величины тестовой статистики и размера выборки для планируемого клинического исследования [33–39].

ЗАКЛЮЧЕНИЕ

Расчет размера выборки участников является неотъемлемой частью планирования любого КИ. Обоснованный выбор основных критериев оценки и статистических моделей определяет необходимую мощность, качество проведения и последующей интерпретации результатов КИ. В обзоре подробно описаны критерии, термины и базовые статистические модели, определяющие корректность процедуры расчета размера выборки участников КИ в соответствии с рекомендациями Коллегии ЕЭК. Однако большое разнообразие заболеваний, особенностей патогенеза в различных популяциях, процедур диагностики, выбора терапевтических схем лечения, целевых оценочных шкал и целого ряда других характеристик вызывает необходимость применения специфических комбинаций планов КИ, следствием чего является востребованность в разработке дополнительных руководств и других информационных ресурсов, содержащих комментарии и примеры применения вероятностной статистики, и последующей гармонизации созданных национальных стандартов с международными.

ЛИТЕРАТУРА / REFERENCES

- Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272(2):122–4. <https://doi.org/10.1001/jama.1994.03520020048013>
- Thabane L. *Sample size determination in clinical trials*. HRM-733 Class Notes. Hamilton: St Joseph's Healthcare, 2004.
- Altman DG. Statistics and ethics in medical research. III. How large a sample? *BMJ*. 1980;281(6251):1336–8. <https://doi.org/10.1136/bmj.281.6251.1336>
- Петри А, Сэбин К. *Наглядная статистика в медицине*. М.: ГЭОТАР-МЕД; 2003.
- Petrie A, Sabin K. *Medical statistics at a glance*. Moscow: GEOTAR-MED; 2003 (In Russ.).
- Zodpey SP. Sample size and power analysis in medical research. *Indian J Dermatol Venereol Leprol*. 2004;70(2):123–8. PMID: 17642587
- Araujo P, Frøyland L. Statistical power and analytical quantification. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007;847(2):305–8. <https://doi.org/10.1016/j.jchromb.2006.10.002>
- Eng J. Sample size estimation: how many individuals should be studied? *Radiology*. 2003;227(2):309–13. <https://doi.org/10.1148/radiol.2272012051>
- Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Routledge; 1988. <https://doi.org/10.4324/9780203771587>
- Lenth RV. Some practical guidelines for effective sample size determination. *Am Stat*. 2001;55(3):187–93. <https://doi.org/10.1198/000313001317098149>
- Dubey SD. Some thoughts on the one-sided and two-sided tests. *J Biopharm Stat*. 1991;1(1):139–50. <https://doi.org/10.1080/10543409108835011>
- Bland JM, Altman DG. One and two sided tests of significance. *BMJ*. 1994;309(6949):248. <https://doi.org/10.1136/bmj.309.6949.248>
- Matthews JN. *Introduction to randomized controlled clinical trials*. New York: Chapman & Hall; 2006. <https://doi.org/10.1201/9781420011302>
- Lehr R. Sixteen S-squared over D-squared: a relation for crude sample size estimates. *Stat Med*. 1992;11(8):1099–102. <https://doi.org/10.1002/sim.4780110811>
- Machin D, Campbell MJ. *Statistical tables for the design of clinical trials*. Blackwell Scientific Publications; 1995. <https://doi.org/10.1002/sim.4780071211>
- Pocock SJ. *Clinical trials: a practical approach*. Wiley & Sons; 2013. <https://doi.org/10.1002/9781118793916>
- Armitage P, Berry G. *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications; 2002. <https://doi.org/10.1002/9780470773666>
- Fleiss JL. General design issues in efficacy, equivalency and superiority trials. *J Periodontol Res*. 1992;27(4 Pt 2):306–13. <https://doi.org/10.1111/j.1600-0765.1992.tb01684.x>
- Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*.

- 2007;46(5):947–54.
<https://doi.org/10.1016/j.jhep.2007.02.015>
19. Garrett AD. Therapeutic equivalence: fallacies and falsification. *Stat Med.* 2003;22(5):741–62.
<https://doi.org/10.1002/sim.1360>
 20. Blackwelder WC. “Proving the null hypothesis” in clinical trials. *Control Clin Trials.* 1982;3(4):345–53.
[https://doi.org/10.1016/0197-2456\(82\)90024-1](https://doi.org/10.1016/0197-2456(82)90024-1)
 21. Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med.* 2000;132(9):715–22.
<https://doi.org/10.7326/0003-4819-132-9-200005020-00006>
 22. Costa LJ, Xavier AC, Giglio A. Negative results in cancer clinical trials — equivalence or poor accrual? *Control Clin Trials.* 2004;25(5):525–33.
<https://doi.org/10.1016/j.cct.2004.08.001>
 23. Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg.* 2001;136(7):796–800.
<https://doi.org/10.1001/archsurg.136.7.796>
 24. Detsky AS, Sackett DL. When was a “negative” clinical trial big enough? How many patients you needed depends on what you found. *Arch Intern Med.* 1985;145(4):709–12.
<https://doi.org/10.1001/archinte.1985.00360040141030>
 25. Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA.* 2001;285(9):1206–8.
<https://doi.org/10.1001/jama.285.9.1206>
 26. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313(7048):36–9.
<https://doi.org/10.1136/bmj.313.7048.36>
 27. Lange S, Freitag G. Choice of delta: requirements and reality — results of a systematic review. *Biomed J.* 2005;47(1):12–27.
<https://doi.org/10.1002/bimj.200410085>
 28. Durrleman S, Simon R. Planning and monitoring of equivalence studies. *Biometrics.* 1990;46(2):329–36.
<https://doi.org/10.2307/2531438>
 29. Ebbutt AF, Frith L. Practical issues in equivalence trials. *Stat Med.* 1998;17(15–16):1691–701.
[https://doi.org/10.1002/\(sici\)1097-0258\(19980815/30\)17:15/16%3C1691::aid-sim971%3E3.0.co;2-j](https://doi.org/10.1002/(sici)1097-0258(19980815/30)17:15/16%3C1691::aid-sim971%3E3.0.co;2-j)
 30. Wiens BL. Choosing an equivalence limit for noninferiority or equivalence studies. *Control Clin Trials.* 2002;23(1):2–14.
[https://doi.org/10.1016/s0197-2456\(01\)00196-9](https://doi.org/10.1016/s0197-2456(01)00196-9)
 31. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291(20):2457–65.
<https://doi.org/10.1001/jama.291.20.2457>
 32. Thomas L, Krebs CJ. A Review of statistical power analysis software. *Bull Ecol Soc Am.* 1997;78(2):128–39.
 33. Lindley DV. The choice of sample size. *Statistician.* 1997;46(2):129–38.
<https://doi.org/10.1111/1467-9884.00068>
 34. Lindley DV. A statistical paradox. *Biometrika.* 1957;44(1/2):187–92.
<https://doi.org/10.2307/2333251>
 35. Gittins J, Pezeshk H. A behavioral Bayes method for determining the size of a clinical trial. *Drug Inf J.* 2000;34(2):355–63.
<https://doi.org/10.1177/009286150003400204>
 36. Gittins J, Pezeshk H. How large should a clinical trial be? *J R Stat Soc Series D Stat.* 2000;49(2):177–87.
<https://doi.org/10.1111/1467-9884.00228>
 37. O'Hagan A, Stevens JW, Campbell MJ. Assurance in clinical trial design. *Pharm Stat.* 2005;4(3):187–201.
<https://doi.org/10.1002/pst.175>
 38. Шредер ОВ, Бунятян НД, Горячев ДВ, Субаев РД, Енгальчева ГН, Кузнецова АД, Косенко ВВ. Математическое прогнозирование эффективности лекарственных средств в доклинических исследованиях. *Вестник Научного центра экспертизы средств медицинского применения. Регуляторные исследования и экспертиза лекарственных средств.* 2022;12(3):315–30.
 Shreder OV, Bunyatyan ND, Goryachev DV, Subaev RD, Engalicheva GN, Kuznetsova AD, Kosenko VV. Mathematical prediction of the efficacy of medicinal products in preclinical studies. *Bulletin of the Scientific Centre for Expert Evaluation of Medicinal Products. Regulatory Research and Medicine Evaluation.* 2022;12(3):315–30 (In Russ.).
<https://doi.org/10.30895/1991-2919-2022-12-3-315-330>
 39. Каркищенко НН. *Альтернативы биомедицины. Т. 1. Основы биомедицины и фармакомоделирования.* М.: Межакademическое издательство ВПК; 2007.
 Karkishchenko NN. *Biomedicine alternatives. Part 1. Basic principles of biomedicine and pharmacomodeling.* Moscow: Interacademic Publishing House of the Military Industrial Complex; 2007 (In Russ.).
 EDN: [SOPUYN](https://doi.org/10.30895/1991-2919-2022-12-3-315-330)
 40. Senn S. Determining the sample size. In: *Statistical issues in drug development.* John Wiley & Sons; 2007.
<https://doi.org/10.1002/9781119238614.ch13>
 41. Julious SA. Designing clinical trials with uncertain estimates of variability. *Pharm Stat.* 2004;3(4):261–8.
<https://doi.org/10.1002/pst.139>

Дополнительная информация. На сайте журнала «Регуляторные исследования и экспертиза лекарственных средств» размещена таблица 1.
<https://doi.org/10.30895/1991-2919-2024-14-3-338-350-tabl1>

Вклад авторов. Все авторы подтверждают соответствие своего авторства критериям ICMJE. Наибольший вклад распределен следующим образом: О.В. Шредер — разработка концепции, написание текста рукописи, редактирование, утверждение окончательного варианта рукописи для публикации; Д.В. Горячев — критический пересмотр текста рукописи, утверждение варианта для публикации; В.А. Меркулов — утверждение окончательной версии рукописи для публикации.

Additional information. Table 1 is posted on the website of *Regulatory Research and Medicine Evaluation*.
<https://doi.org/10.30895/1991-2919-2024-14-3-338-350-tabl1>

Authors' contributions. All the authors confirm that they meet the ICMJE criteria for authorship. The most significant contributions were as follows. Olga V. Shreder conceptualised the study, drafted and edited the manuscript, and approved the final version for publication. Dmitry V. Goryachev critically revised the manuscript and approved the final version for publication. Vadim A. Merkulov approved the final version of the manuscript for publication.

ОБ АВТОРАХ / AUTHORS

Шредер Ольга Васильевна, канд. биол. наук / Olga V. Shreder, Cand. Sci. (Biol.)

ORCID: <https://orcid.org/0000-0002-7926-6033>

Горячев Дмитрий Владимирович, д-р мед. наук / Dmitry V. Goryachev, Dr. Sci. (Med.)

ORCID: <https://orcid.org/0000-0001-8583-2372>

Меркулов Вадим Анатольевич, д-р мед. наук / Vadim A. Merkulov, Dr. Sci. (Med.)

ORCID: <https://orcid.org/0000-0003-4891-973X>

Поступила 11.12.2023

После доработки 08.04.2024

Принята к публикации 10.04.2024

Received 11 December 2023

Revised 8 April 2024

Accepted 10 April 2024