



Statistical hypothesis testing: general approach in medical research

Alexander Yu. Suvorov, Nikolay M. Bulanov✉, Anastasia N. Shvedova, Ekaterina A. Tao,
Denis V. Butnaru, Maria Yu. Nadinskaia, Alexey A. Zaikin

*Sechenov First Moscow State Medical University (Sechenov University)
8/2, Trubetskaya str., Moscow, 119991, Russia*

Abstract

Statistical hypothesis testing is one of the key steps in modern medical research. Initially, scientists formulate a research hypothesis based on which the statistical hypothesis is then developed and statistically tested. This review provides the null and alternative hypotheses' compiling examples for different research questions and the general algorithm for their testing using t-test. The authors also describe type I errors, which are necessary to interpret *p-values* estimated from statistical tests, and type II errors, which are used to assess study power. The article focuses on effect size and its calculation methods, and the difference between statistically significant and clinically significant effects. The associations between effect size, sample size, and type II error are also discussed.

Keywords: statistics as topic; null hypothesis; type I error; type II error; effect size; *t*-test

MeSH terms:

STATISTICS AS TOPIC
BIOMEDICAL RESEARCH
RESEARCH DESIGN

For citation: Suvorov A.Yu., Bulanov N.M., Shvedova A.N., Tao E.A., Butnaru D.V., Nadinskaia M.Yu., Zaikin A.A. Statistical hypothesis testing: general approach in medical research. *Sechenov Medical Journal*. 2022; 13(1): 4–13. <https://doi.org/10.47093/2218-7332.2022.426.08>

CONTACT INFORMATION:

Nikolay M. Bulanov, Cand. of Sci. (Medicine), Associate Professor, Department of Internal, Occupational Diseases and Rheumatology, Sechenov First Moscow State Medical University (Sechenov University)

Address: 8/2, Trubetskaya str., Moscow, 119991, Russia

Tel.: +7 (919) 100-22-79

E-mail: bulanov_n_m@staff.sechenov.ru

Conflict of interests. The authors declare that there is no conflict of interests.

Financial support. The study was not sponsored (own resources).

Received: 05.04.2022

Accepted: 11.05.2022

Published Online: 16.06.2022

Date of publication: 23.06.2022

УДК [61:57.084];311.1

Проверка статистических гипотез: общие подходы в практике медицинских исследований

А.Ю. Суворов, Н.М. Буланов✉, А.Н. Шведова, Е.А. Тао, Д.В. Бутнару,
М.Ю. Надинская, А.А. Заикин

*ФГАОУ ВО «Первый Московский государственный медицинский университет им. И.М. Сеченова»
Минздрава России (Сеченовский Университет)
ул. Трубецкая, д. 8, стр. 2, г. Москва, 119991, Россия*

Аннотация

Проверка статистических гипотез – один из ключевых этапов современных исследований в области медицины. На начальном этапе ученые выдвигают исследовательскую гипотезу, на основе которой формулируют статистическую гипотезу, которая поддается проверке с помощью статистических тестов. В руководстве представлены примеры составления нулевой и альтернативной статистических гипотез для разных исследовательских вопросов, представлен общий алгоритм их проверки на примере t -теста. Авторы разбирают концепции ошибок I рода, которые необходимы для интерпретации p -значений, полученных в статистических тестах, и ошибок II рода, которые используют для расчета мощности исследования. Существенное внимание уделено понятию размера эффекта и его оценке, различиям между статистически и клинически значимыми эффектами. Продемонстрирована взаимосвязь между размером эффекта, численностью выборки и величиной ошибки II рода.

Ключевые слова: статистика как раздел; нулевая гипотеза; ошибка I рода; ошибка II рода; размер эффекта; t -тест

Рубрики MeSH:

СТАТИСТИКА КАК ТЕМА

БИМЕДИЦИНСКИЕ ИССЛЕДОВАНИЯ

НАУЧНЫХ ИССЛЕДОВАНИЙ ПЛАНИРОВАНИЕ

Для цитирования: Суворов А.Ю., Буланов Н.М., Шведова А.Н., Тао Е.А., Бутнару Д.В., Надинская М.Ю., Заикин А.А. Проверка статистических гипотез: общие подходы в практике медицинских исследований. Сеченовский вестник. 2022; 13(1): 4–13. <https://doi.org/10.47093/2218-7332.2022.426.08>

КОНТАКТНАЯ ИНФОРМАЦИЯ:

Буланов Николай Михайлович, канд. мед. наук, доцент кафедры внутренних, профессиональных болезней и ревматологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет)

Адрес: ул. Трубецкая, д. 8, стр. 2, г. Москва, 119991, Россия

Тел.: +7 (919) 100-22-79

Е-mail: bulanov_n_m@staff.sechenov.ru

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Финансирование. Исследование не имело спонсорской поддержки (собственные ресурсы).

Поступила: 05.04.2022

Принята: 11.05.2022

Дата публикации онлайн: 16.06.2022

Дата печати: 23.06.2022

List of abbreviations

BP – blood pressure

RR – relative risk

SBP – systolic blood pressure

H_0 – null hypothesis

H_1 (H_A) – alternative (null) hypothesis

Hypothesis testing is one of the essential steps in modern medical research. Many studies aim not only to describe the data but also to determine the differences in study subjects' characteristics and assess their significance. Study subjects may be patients, animal models or cell cultures. Any existing subjects' observation, with or without active intervention, can be called an experiment. In medicine, one way to conduct an experiment is by performing a designed study [1]. Strict rules and firm conditions must be defined and implemented before initiating a designed study:

- the aim of the study is accompanied by a specific *research question*;
- *the objectives* are formulated through which the aim can be achieved;

- one or several research hypotheses are formulated that will be tested to answer the research question;
- the study design is determined to achieve the goal in the most effective and valid way, and to obtain reliable and reproducible results by minimizing the risk of making a mistake;
- the inclusion and exclusion criteria are formulated without ambiguity;
- the statistical analysis methods, which are used for achieving the objectives mentioned before and answering the research question, are described in detail. Thus, when designing an experiment, the researchers test a model under ideal conditions which allow them to answer the research question. In addition, designed

studies have restricted sample size and time to perform all the steps.

An example of a slightly different approach are observational exploratory studies; the main goal of which is not to confirm pre-formulated hypotheses about the influence of certain factors on the outcome but to look for any relationships and generate hypotheses. Thus, observational studies focus on determining whether there are any significant interactions between factors within the study population. In observational studies the restricted main research question that requires confirmation might be missing. However, in exploratory studies the researchers must also determine the goal, the objectives, and the study population, which is generally larger compared to the designed studies samples, in advance.

The aim of this review is to explain the main aspects of the designed research and the relationship between the research question and its answer in terms of research methodology, a key aspect of which is statistical hypotheses testing.

RESEARCH AND STATISTICAL HYPOTHESES

Every study, or experiment, begins with the formulation of a hypothesis (scientific assumption). After that, researchers try to confirm or reject the hypothesis. For example, they can assume that a particular region's bad ecology affects people's health, or that smoking is associated with a high risk of cardiovascular events. These assumptions are *research hypotheses*, that should be distinguished from *statistical hypotheses*. Statistical hypothesis represents the research hypothesis that can be analyzed by using statistical methods through an experiment with specific design.

A statistical hypothesis is a judgment about the parameters that describe a statistical population (general population), but not a sample. And the statistical population is a homogeneous group of people, for example, those at risk or of interest in the present study (experiment). Such a group may be a city population or all patients in a hospital over a certain period of time. Initially, researchers claim two hypotheses about the possible relationship of the observed phenomena (potential risk factors and outcomes): null and alternative.

The *null hypothesis*, commonly shown as H_0 , claims that observed effects, phenomena, and relationships happen randomly; meaning they are not connected. On the contrary, the *alternative hypothesis* (H_1 or H_A) claims that observed phenomena do not occur randomly and are connected. It should be noted that connection here means any association, not only causal effect. For example, a minor study assesses the differences between mean values (μ_1 and μ_2) of empirical value – total cholesterol level – in patients with or without a history of myocardial infarction (group 1 and 2,

respectively). The research hypothesis can state that groups differ or do not differ. In this case, the null hypothesis would claim that total cholesterol level is not connected with the risk of myocardial infarction, i.e. there is no difference between the two mean values (observed differences are random):

$$H_0 : \mu_1 = \mu_2.$$

The alternative hypothesis claims that the observed differences are significant and not random:

$$H_1 : \mu_1 \neq \mu_2.$$

The null and alternative hypotheses are *mutually exclusive*, meaning that true H_0 excludes H_1 being false and otherwise. Therefore, accepting the alternative hypothesis (existing difference between two groups) implies rejecting the null hypothesis.

STATISTICAL TESTS

Statistical tests are statistical proof methods, or inferential statistical analysis, which are used to determine whether the null hypothesis (H_0) might be rejected. It should be noted that statistical tests do not formally accept the null hypothesis but only demonstrate if we can reject it in favor of the alternative one.

Each statistical test is a mathematical function which calculates test statistics. Test statistics show how far an observed value differs from the expected value distribution, given that the null hypothesis is not rejected. The larger the test statistic value the bigger the difference between the observed and expected distributions.

Test statistics set evaluated from specific test follows defined distribution law (for more on distributions see [2]). For example, *t*-test evaluates *t*-statistics which follows Student's *t*-distribution. To reject the null hypothesis, the test statistic distribution threshold must be determined, and that threshold is called *critical value*, and corresponding probability – *significance level*.

Significance level (α) shows the probability of the true null hypothesis being incorrectly rejected. This probability is known as Type I error. Significance level and corresponding critical value are initially proclaimed in order to assess the probability of making Type I error as the lower the α the less the risk of rejecting the true null hypothesis. However, there is always a chance that this might happen, thus researchers can only minimize the risk by choosing the suitable threshold. Depending on the field of study, various α values can be chosen, for instance, medical researchers commonly use $\alpha = 0.05$. When it is highly crucial to avoid Type I error (i. e. two study groups are comparable) α would be set as low as possible ($\alpha = 0.01$ or lower).

After the test statistic is determined, researchers also obtain the *p-value* (p). *P-value* is the probability of acquiring the same results or more extreme than

imagined, with the null hypothesis being true. Extreme results are generally understood to mean results which differ greater from the expected distribution. Small *p-value* means that if the null hypothesis is true, then the chance of obtaining the same or more extreme results is rather low, therefore it is highly likely that the null hypothesis is false and can be rejected. If $p \leq \alpha$, i. e. test statistic equals critical value or exceeds it, the result is statistically significant (Fig. 1). That is why in the “Materials and Methods” section it is always stated at which *p-value* the results are considered to be statistically significant. The phrase “*P-values* below 0.05 were considered statistically significant” means that researchers chose Type I error threshold $\alpha = 0.05$.

Let us imagine that we are conducting a parallel study with two groups: one group is given the antihypertensive treatment A, the other group is given a placebo. Our null hypothesis states that the drug A effect on blood pressure (BP) is equal to placebo if we estimate the mean blood pressure at the end of the study in both groups:

- μ_{TRT} – mean BP value in patients who received treatment A;
- μ_{PLC} – mean BP value in patients who received placebo;
- $H_0 : \mu_{TRT} = \mu_{PLC}$ or $\mu_{TRT} - \mu_{PLC} = 0$;
- $H_1 : \mu_{TRT} \neq \mu_{PLC}$ or $\mu_{TRT} - \mu_{PLC} \neq 0$.

Using the Student’s *t*-test, we verify if we can reject the null hypothesis claiming $\mu_{TRT} - \mu_{PLC} = 0$. We can build a graph, which demonstrates *t*-statistics distribution according to the difference (Fig. 1: green lines). After conducting the experiment, we have the difference $\mu_{TRT} - \mu_{PLC}$, which corresponds with *t*-statistic = 2.5, given the null hypothesis is true (Fig. 1: Red line). We see that the observed result exceeds critical values, therefore the null hypothesis can be rejected. Corresponding *p*-value is 0.0126, thus, if the chosen α is equal to 0.05 the null hypothesis can be rejected, and the differences are statistically significant.

EFFECT SIZE

Research hypotheses are most commonly generated to identify differences or associations between certain parameters. However, not only are the differences essential but also their significance. For example, one study compared the proportion of patients who achieved therapeutic effect in the treatment group and the placebo group, and they were 10% and 80%, respectively. We see the remarkable difference equal to 70%. In another study, these proportions were equal to 45% and 55%. We see that the results differ as well, but only by 10%. In the third study, the treatment efficacy in two groups reached 75% and 80% differing only by 5%. Another example of the new antihypertensive treatment illustrates this point clearly. The new treatments X, Y, and Z lower systolic blood pressure (SBP) by 15, 8,

and 1 mmHG on average, respectively. Conventional antihypertensive treatment lowers SBP by 1 mmHg. All examples demonstrate that a certain effect exists, but it is *different* and more or less prominent.

Effect size is a wide statistical concept that refers to a certain statistic or parameter which shows the quantity of differences or associations between the distributions in different groups. In medicine, the effect size is important as it refers not only to the statistical significance but also to the clinical significance of the results observed in the study. For example, we use the average decrease in SBP as the effect size. The average decrease in SBP by 1 mmHg, most likely, is not clinically significant – for the patient, the treatment Z will not be more effective than conventional therapy, thus, from a practical point of view, its administration provides no advantages. On the contrary, too great an effect (treatment X) may be associated with complications caused by hypotension. Treatment Y is probably the best choice amongst the new drugs in clinical practice as it has a significant but not excessive antihypertensive effect.

It should be emphasized that we try to statistically determine the effect size when testing statistical hypotheses. Absolute difference rarely equals null, however it can be small and clinically insignificant and both positive and negative. On the other hand, when testing statistical hypotheses, it is necessary to

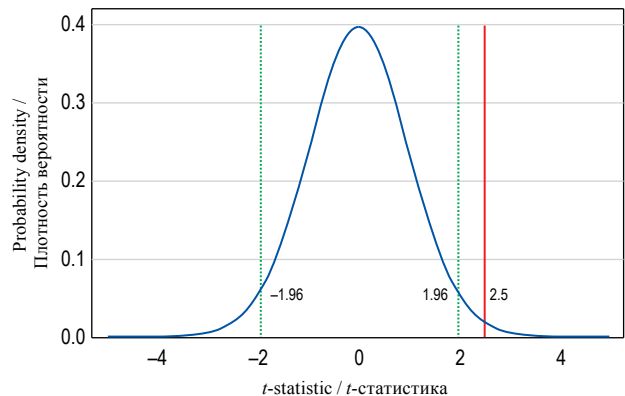


FIG. 1. *T*-distribution for 50 degrees of freedom
РИС. 1. *T*-распределение для 50 степеней свободы

Notes: green dotted lines indicate critical values of 1.96 and -1.96 corresponding to the 97.5th percentile and the 2.5th percentile, respectively, i.e. two-sided significance level $\alpha = 0,05$. B. Red solid line indicates *t*-statistics = 2.5, which exceeds critical value of 1.96 using two-tailed *t*-test ($p = 0,0126$ with specified $\alpha = 0,05$), thus the null hypothesis can be rejected, and we can accept the alternative one.

Примечания: зелеными пунктирными линиями указаны критические значения -1,96 и 1,96, соответствующие 2,5-му и 97,5-му перцентиллям – двустороннему уровню значимости $\alpha = 0,05$. Красной сплошной линией обозначена *t*-статистика = 2,5, которая превышает критическое значение 1,96 при двустороннем *t*-тесте ($p = 0,0126$ при заданном $\alpha = 0,05$). Таким образом нулевая гипотеза может быть отклонена и принята альтернативная.

define the criteria for the detection of plausible clinical effect. In the antihypertensive treatment example, an average decrease in SBP of ≤ 2 mmHg can be defined as the absence of the clinical effect, 3–10 mm Hg – as a moderate effect, and >10 mm Hg – large effect. For testing research hypothesis, we want to determine whether the average decrease in SBP will at least reach a moderate effect size after treatment with the drugs X, Y, and Z.

How to measure the effect size?

Psychologist and statistician Jacob Cohen made a significant contribution to the conceptualization of effect size, writing in his later work: “The primary product of a research inquiry is one or more measures of effect size, not p values” [3]. Nowadays, there are many statistics to measure the effect size, testing any hypotheses.

Effect size measuring methods can be standardized and unstandardized. Standardized methods allow measuring the effect size for variables with both similar and different dimensions (for example, correlation coefficient for variables in different units), for combining results of different studies (meta-analysis and meta-regression), and for comparing variables varying in metrics (e. g., g/l in one study and mmol/l in another one) [4].

Methods to measure the effect size are the following (Tables S1–4 in the supplementary):

- Effect size assessing connections between numeric variables’ distributions, or how one variable’s distribution influences another (correlation coefficient, coefficient of determination);
- Effect size assessing the difference between statistics (Cohen’s d, Glass’ Δ , risk difference);
- Effect size assessing the connections between categorical variables (Cohen’s h, odds ratio).

Research hypotheses and the effect size

Effect size plays a crucial role in research and the forming of statistical hypotheses. In the beginning, researchers put a question if the observed effect occurs randomly. For example, do total cholesterol levels in exposed and control groups differ? Observed differences may be random. In order to verify if the effect is random or not, pilot studies are conducted which aim to determine the presence or absence of the effect. These studies are hypothesis-generating (for more on preferred study design for different purposes see [5]). If the effect is confirmed, pilot studies can assess its observed or hypothetical size. However, pilot studies are often restricted and do not consider the effect size validly (Fig. 2).

The pilot studies are followed by confirmatory studies that determine a certain effect size. For example, a pilot study established that smoking in men aged 35–45, who live in urban area, increases the risk of cardiovascular diseases over 10 years (relative risk

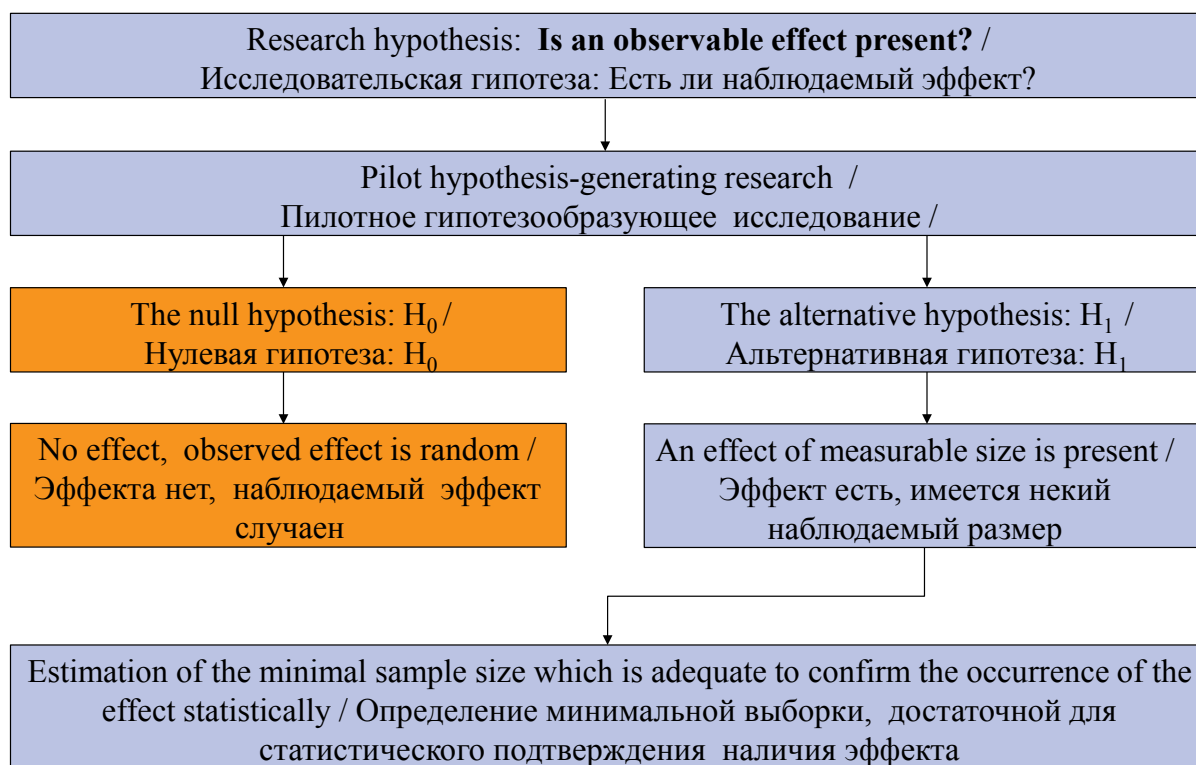


FIG. 2. Flowchart of the hypothesis-generating studies
РИС. 2. Схема проведения гипотезообразующих исследований

(RR) is X). Researchers then have to question whether this is true for men living in the rural area. To conduct such a study, we can rely on the data from the city group and expect that the effect size (RR) will be at least X (Fig. 3). It is also possible that according to our research assumptions, the effect will be n -fold less, so we can conduct a study that will be able to detect the effect size of at least X/n .

The concept of the effect size allows avoiding pilot studies as the results of previous research can be used. One of the key concepts in medical research is the *effect size threshold* that should be plausible from a practical point of view. For example, in the study of a new drug for weight loss in patients weighing more than 200 kg, a statistically significant weight loss of 1 kg was obtained after one year. The effect is statistically significant but has no benefit for the patients: within 1 year, nutritionists (and patients themselves) would likely want to achieve a more noticeable weight loss. It would probably be more justifiable to introduce an effect size threshold of 10 or 15 kg.

ONE-SIDED AND TWO-SIDED STATISTICAL TESTS

Two-sided tests

Let us return to the antihypertensive treatment example. There is a promising new treatment (TRT) that is supposed to lower SBP; however, we do not know how

effective it will be in clinical research. The conventional antihypertensive treatment (CTRL) is administered to the control group. Estimated average decrease in SBP will be μ_{TRT} and μ_{CTRL} in the treatment group and in the control group, respectively.

If the aim of our study is to determine which treatment is more effective, the null hypothesis asserts $H_0 : \mu_{TRT} = \mu_{CTRL}$ and the alternative one asserts the opposite $H_1 : \mu_{TRT} \neq \mu_{CTRL}$. Hence, we can divide the latter into two simple statements:

$$H_1 = \begin{cases} \mu_{TRT} > \mu_{CTRL} \\ \mu_{TRT} < \mu_{CTRL} \end{cases}$$

In other words, we test alternative hypotheses for two scenarios: when TRT is more effective than CTRL and when TRT is less effective. Tests used to check both statements are called two-sided tests. For example, if we analyze t -statistic probability distribution (Fig. 1), one will notice two opposite critical values of 1.96 and -1.96 (two-sided test with a given $\alpha = 0.05$). In two-sided tests, the significance level is divided by half and critical values of each side are $\alpha/2$:

$$\begin{cases} \alpha_{right} = \alpha/2 \\ \alpha_{left} = \alpha/2 \end{cases}$$

Given α equals 0.05 (5%), the null hypothesis will be rejected, if observed t -statistic exceeds any of the two critical values (2.5% and 97.5%) indicated by

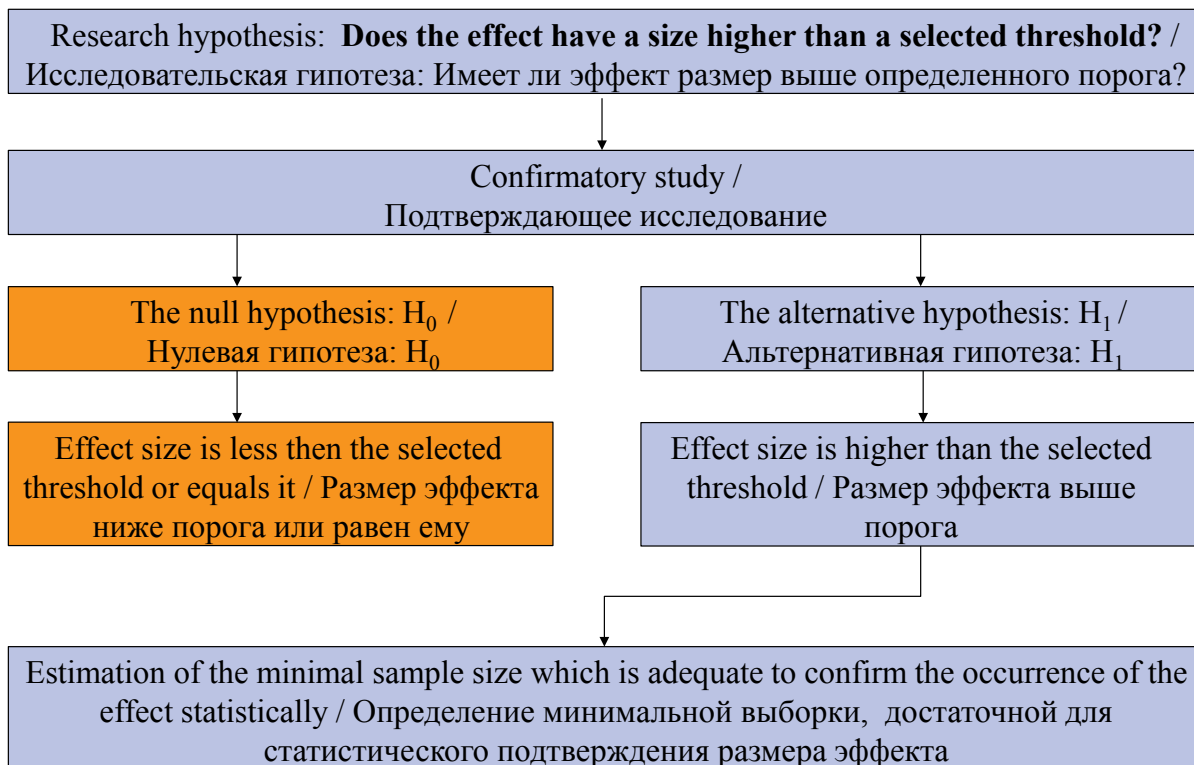


FIG. 3. Flowchart of confirmatory studies
 РИС. 3. Схема проведения подтверждающих исследований

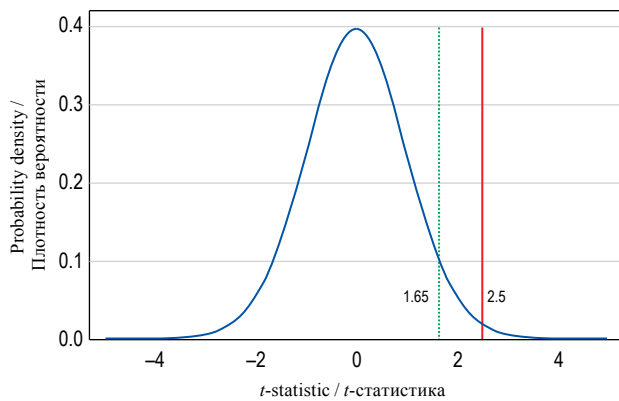


FIG. 4. One-sided *t*-test illustration
РИС. 4. Иллюстрация одностороннего *t*-теста

Notes: the green line indicates critical value of 1.65 which corresponds with one-sided significance level $\alpha = 0.05$; *p*-value is 0.006 for *t*-statistic $t = 2.5$ (red line): the result is clinically significant, given the chose type I error value as $p < \alpha$.

Примечания: зеленая линия обозначает критическое значение 1,65, соответствующее одностороннему уровню значимости $\alpha = 0,05$. Для статистики $t = 2,5$ (красная линия) *p*-value = 0,006: результат статистически значим при выбранном значении ошибки I рода, поскольку $p < \alpha$.

green lines (Fig. 1). So why do researchers use two-sided tests?

As the TRT effectiveness has yet to be discovered, the two-sided test is appropriate for all scenarios: TRT and CTRL have approximately the same effectiveness, TRT is better or worse than conventional treatment.

One-sided tests

On the other hand, if the main question is whether the new treatment is more effective than the conventional one, the hypotheses would change. The actual goal is to determine the significant effect size $\mu_{TRT} > \mu_{CTRL}$.

In this case, the null and alternative hypotheses assert the following:

$$\begin{cases} H_0 = \mu_{TRT} \leq \mu_{CTRL} \\ H_1 = \mu_{TRT} > \mu_{CTRL} \end{cases}$$

To test this hypothesis, one-sided tests are used not only to estimate the *certain effect size* but also its *direction*. In this case, it must be verified if the test

statistic exceeds the critical value located on the right side of the distribution on the graph (Fig. 4).

TYPE I AND II ERRORS

We discussed type I errors but for valid interpretation of the results we should consider another type of errors (Table).

Type I error (α , false-positive result) occurs when the true null hypothesis is rejected and the false alternative hypothesis is accepted. For example, a researcher assumes that the differences are significant between two groups, however in fact they are random.

Type II error (β , false-negative result) occurs when neither the false null hypothesis is rejected nor the true alternative hypothesis is accepted. For example, researchers consider the differences to be random, however in fact they are significant and occur not randomly.

To conduct a successful experiment, the risk of type I and II errors should be minimized. Type I error, as mentioned previously, is the significance level of the test which is then compared to the *p-value*. The low risk of type I error means that with high probability, the true null hypothesis will not be rejected.

Type II error shows the probability of rejecting the incorrect null hypothesis equal to $(1 - \beta)$. This probability is known as *statistical test power* = $1 - \beta$. In medical studies, the minimal acceptable is usually set to 80%, i. e. the maximal type II error is 20%.

ADEQUATE SAMPLE SIZE AND THE CONCEPT OF STATISTICAL HYPOTHESIS TESTING

The last step in statistical hypothesis testing is to choose the adequate sample size required for accepting or rejecting the null hypothesis [6]. We can summarize the statistical hypothesis testing in the flowchart shown in Fig. 5.

After formulating the research question and assuming the expected effect size, the researcher should do the following:

- choose the most suitable statistical test which relates with the effect size distribution law in statistical population;

Table. Type I and II errors
Таблица. Ошибки I и II рода

In statistical population / В статистической популяции	In the study / В ходе исследования	After testing H_0 / Результат проверки H_0	Probability / Вероятность
H_0 true / H_0 верна	H_0 not rejected / H_0 не отклонена	Decision not to reject is correct / Решение не отклонять верное	$P = 1 - \alpha$
H_0 true / H_0 верна	H_0 rejected / H_0 отклонена	Incorrect (false) rejection, Type I error / Ошибочное отклонение, ошибка I рода	$P = \alpha$
H_0 false / H_0 неверна	H_0 not rejected / H_0 не отклонена	Decision not to reject is incorrect (false), Type II error / Решение не отклонять ошибочное, ошибка II рода	$P = \beta$
H_0 false / H_0 неверна	H_0 rejected / H_0 отклонена	Correct rejection / Верное отклонение	$P = 1 - \beta$

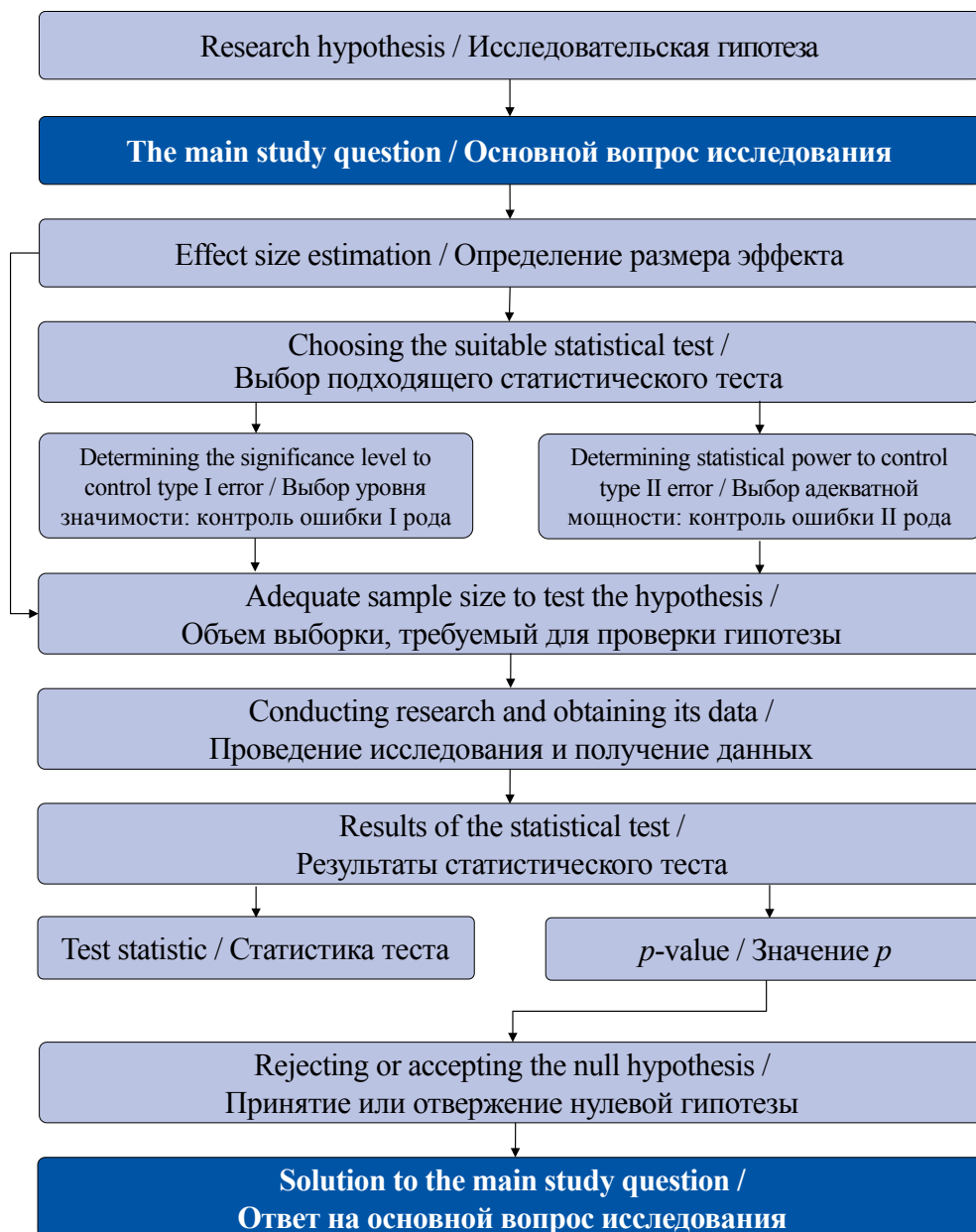


FIG. 5. Flowchart of the statistical hypothesis testing
РИС. 5. Схема проверки статистических гипотез

- choose the suitable significance level and study power;
- calculate the adequate sample size.

Only after performing all those steps, can we conduct the statistical test. But what is the connection between study power, significance level, effect size, and sample size? Let us assume that we are trying to statistically determine the effect size E with power $1 - \beta$, significance level α (Fig. 6A), and sample size n [7]. We note in advance that the test significance level α must remain fixed under any circumstances. If the researcher wants to keep the sample size the same but to increase the power, the simplest solution is to

assume that we will observe a larger effect size, for example, twofold (or $E \times 2$). In this case, the power will actually increase (Fig. 6B). However, in reality, the researcher cannot voluntarily observe a larger or smaller effect, moreover, the assumption of the effect size is only a research hypothesis. In this case, the effect size should not increase either; however, as the sample size increases, the study power also increases (Fig. 6C). Thus, if one wants to obtain a certain effect size with the strict significance level, the only way to minimize the risk of falsely accepting an incorrect null hypothesis (β) is to increase the sample size.

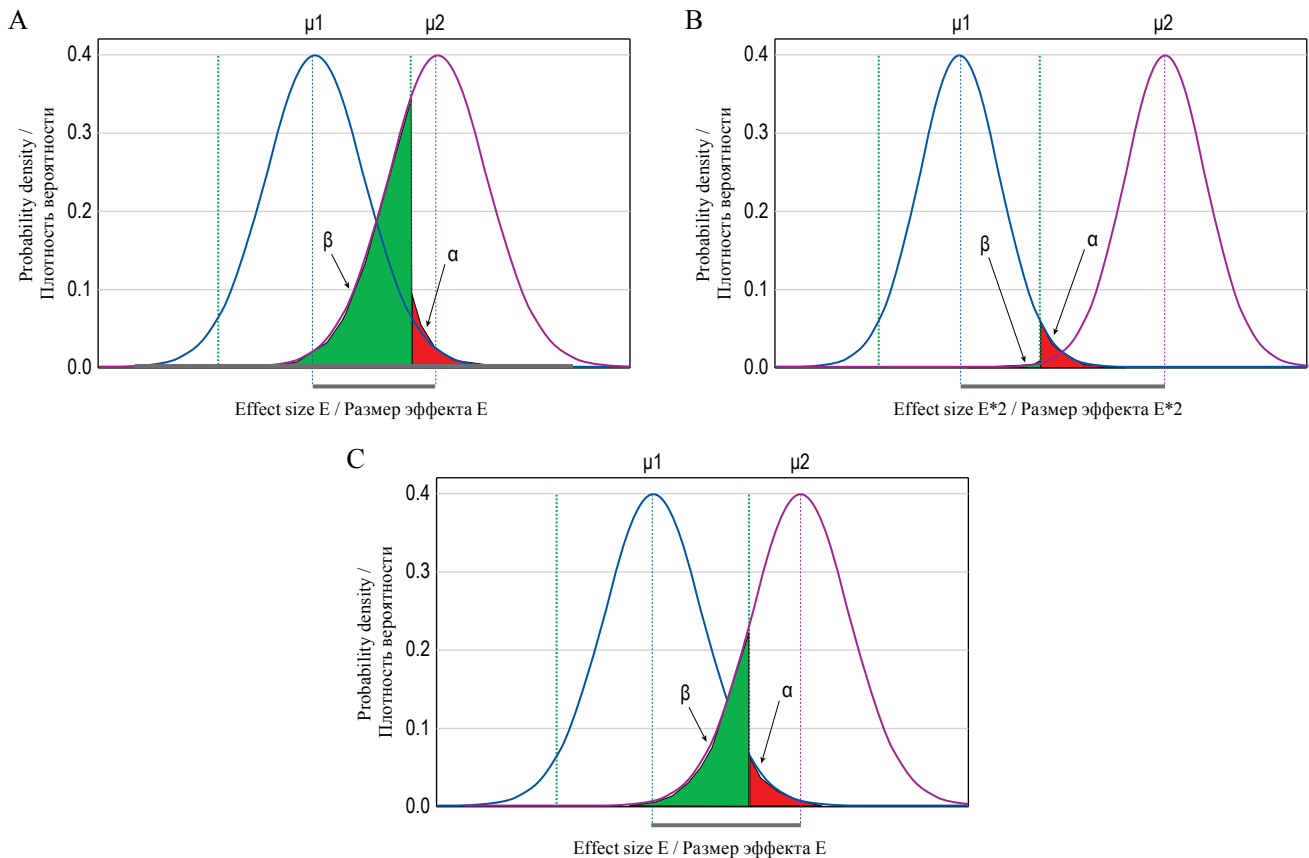


FIG. 6. The relationship between effect size, type I and II errors, and sample size, when testing statistical hypotheses:

- A. Effect size and type I and II errors.
- B. Type II error changing after the effect size increasing.
- C. Type II error changing after sample size increasing.

РИС. 6. Взаимосвязь между размером эффекта, ошибками I и II рода и размером выборки при проверке статистических гипотез:

- A. Взаимосвязь размера эффекта, ошибок I и II рода.
- B. Изменение величины ошибки II рода при увеличении размера эффекта.
- C. Изменение величины ошибки II рода при увеличении размера выборки.

Note: μ_1 – mean in group 1, μ_2 – mean in group 2.

Примечание: μ_1 – среднее в группе 1, μ_2 – среднее в группе 2.

CONCLUSION

Formulating adequate research and statistical hypotheses is an important skill for all medical researchers, without which successful planning and conducting research in medicine is almost impossible. In addition, the concepts of effect size, type I and II

errors are essential for interpreting the results of a scientist’s own research, as well as those of published studies. These ideas are universal and applicable to any statistical tests, and moreover, they are of much greater importance for a scientist than the skill of applying any specific method.

AUTHOR CONTRIBUTIONS

Alexander Yu. Suvorov, Nikolay M. Bulanov, and Anastasia N. Shvedova contributed equally to this work and should be considered as co-first authors. Alexander Yu. Suvorov, Nikolay M. Bulanov, Anastasia N. Shvedova, Ekaterina A. Tao, Alexey A. Zaikin and Maria Yu. Nadinskaia, participated in writing the text of the manuscript. Alexander Yu. Suvorov, Nikolay M. Bulanov, and Anastasia N. Shvedova searched and analyzed the literature on the review topic. Alexander Yu. Suvorov and Denis V. Butnaru developed the general concept of the article and supervised its writing. All authors participated in the discussion and editing of the work. All authors approved the final version of the publication.

ВКЛАД АВТОРОВ

А.Ю. Суворов, Н.М. Буланов, А.Н. Шведова в равной степени внесли вклад в эту работу и должны считаться первыми соавторами. А.Ю. Суворов, Н.М. Буланов, А.Н. Шведова, Е.А. Тао, А.А. Заикин и М.Ю. Надинская участвовали в написании текста рукописи. А.Ю. Суворов, Н.М. Буланов и А.Н. Шведова выполняли поиск и анализ литературы по теме обзора. А.Ю. Суворов и Д.В. Бутнару разработали общую концепцию статьи и осуществляли руководство ее написанием. Все авторы участвовали в обсуждении и редактировании работы. Все авторы утвердили окончательную версию публикации.

SUPPLEMENTARY MATERIALS

Supplementary materials associated with this article can be found in the online version at doi:
<https://doi.org/10.47093/2218-7332.2022.426.08.S>

REFERENCES / ЛИТЕРАТУРА

1. *Kestenbaum B.* Epidemiology and biostatistics: An introduction to clinical research. Springer, Cham, 2019, 246p. <https://doi.org/10.1007/978-3-319-96644-1>
2. *Bulanov N.M., Suvorov A.Yu., Blyuss O.B., et al.* Basic principles of descriptive statistics in medical research. *Sechenov Medical Journal.* 2021; 12(3): 4–16. <https://doi.org/10.47093/2218-7332.2021.12.3.4-16>
3. *Cohen J.* Things I have learned (so far) *Am Psychol.* 1990; 45(12): 1304–1312. <https://doi.org/10.1037/0003-066X.45.12.1304>
4. *Kirkwood B.R., Sterne J.A.C.* Essential medical statistics.

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Дополнительные материалы, прилагаемые к этой статье, можно посмотреть в онлайн-версии по адресу:
<https://doi.org/10.47093/2218-7332.2022.426.08.S>

- 2nd edition Blackwell Science. 2003, 512p. ISBN: 978-0-86542-871-3
5. *Bulanov N.M., Blyuss O.B., Munblit D.B., et al.* Studies and research design in medicine. *Sechenov Medical Journal.* 2021; 12(1): 4–17. <https://doi.org/10.47093/2218-7332.2021.12.1.4-17>
6. *Sawilowsky S.S.* New effect size rules of thumb. *Journal of modern applied statistical methods.* 2009; 8 (2): 467–474. <https://doi.org/10.22237/jmasm/1257035100>
7. *Cohen J.* Statistical power analysis for the behavioral sciences. 2nd edition. Routledge. <https://doi.org/10.4324/9780203771587>

INFORMATION ABOUT THE AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

Alexander Yu. Suvorov, Cand. of Sci. (Medicine), Chief Statistician, Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0002-2224-0019>

Суворов Александр Юрьевич, канд. мед. наук, главный статистик Центра анализа сложных систем ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0002-2224-0019>

Nikolay M. Bulanov✉, Cand. of Sci. (Medicine), Associate Professor, Department of Internal, Occupational Diseases and Rheumatology, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0002-3989-2590>

Буланов Николай Михайлович✉, канд. мед. наук, доцент кафедры внутренних, профессиональных болезней и ревматологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0002-3989-2590>

Anastasia N. Shvedova, 6th year student, Institute of Child's Health named after N. F. Filatov, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0001-7518-2273>

Шведова Анастасия Никитична, студентка 6-го курса клинического института детского здоровья им. Н.Ф. Филатова ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0001-7518-2273>

Ekaterina A. Tao, Cand. of Sci. (Medicine), Assistant Professor, Department of Internal, Occupational Diseases and Rheumatology, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0002-3989-2590>

Тао Екатерина Александровна, канд. мед. наук, ассистент кафедры внутренних, профессиональных болезней и ревматологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0002-0621-7054>

Denis V. Butnaru, Cand. of Sci. (Medicine), Vice-rector for Research, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0003-2173-0566>

Бутнару Денис Викторович, канд. мед. наук, проректор по научной работе ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0003-2173-0566>

Maria Yu. Nadinskaia, Cand. of Sci. (Medicine), Associate Professor, Department of Internal Medicine Propaedeutics, Gastroenterology and Hepatology, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0002-1210-2528>

Надинская Мария Юрьевна, канд. мед. наук, доцент кафедры пропедевтики внутренних болезней, гастроэнтерологии и гепатологии ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0002-1210-2528>

Alexey A. Zaikin, Cand. of Sci. (Phys. and Math.), Deputy Director, Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University (Sechenov University).
 ORCID: <https://orcid.org/0000-0001-7540-1130>

Заикин Алексей Анатольевич, канд. физ.-мат. наук, заместитель директора Центра анализа сложных систем ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России (Сеченовский Университет).
 ORCID: <https://orcid.org/0000-0001-7540-1130>

✉ Corresponding author / Автор, ответственный за переписку