LEVELS OF EVIDENCE AND STUDY DESIGNS

Borisova EO ✉, Eremina JuN, Gulbekova OV

Pirogov Russian National Research Medical University, Moscow, Russia

In this article, various designs of clinical trials used to obtain new scientific knowledge in the field of clinical medicine are considered from the position of their evidential value in studying the cause-and-effect relationship between the influencing factor and result of its potential effect on human health. Basic differences between observational and experimental trials, their limitations due to peculiarities of design of clinical trials are being discussed. A conclusion is made that validity of results of clinical trials should be assessed taking into account the limitations that are typical of various designs. Accuracy of clinical trials depends on many factors that can distort the obtained results as compared with true values. It is noted that observational trials are subject to systematic and accidental errors to a greater extent than experimental ones. It occurs because design characteristics do not allow observational trials to control the mistakes associated with possible incompatibility of comparison groups. They can detect a statistical relation between the phenomena, but only randomized clinical trials can prove that there is a causal relationship. Accuracy of a randomized clinical trial can be increased using systematic reviews and meta-analysis.

Keywords: designs of clinical trials, observational controlled trials, experimental clinical trials, systematic review, meta-analysis, systematic errors, substantiation of conclusions, evidence reliability levels, recommendation strength levels

ДОКАЗАТЕЛЬНОСТЬ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ РАЗЛИЧНЫХ ДИЗАЙНОВ

Е. О. Борисова ✉, Ю. Н. Еремина, О. В. Гульбекова

Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва, Россия

В данной статье различные дизайны клинических исследований, которые используются для получения новых научных знаний в области клинической медицины, рассматриваются с позиций их доказательной ценности в изучении причинно-следственных взаимоотношений между воздействующим фактором и результатом его потенциального влияния на здоровье людей. Обсуждаются основные различия между наблюдательными и экспериментальными исследованиями, а также их ограничения, обусловленные особенностями дизайнов клинических исследований. В заключение делается вывод о том, что к оценке достоверности результатов клинических исследований надо подходить с пониманием тех ограничений, которые характерны для различных дизайнов. Точность клинических исследований зависит от влияния многих факторов, которые способны приводить к искажению получаемых результатов по сравнению с их истинными значениями. При этом отмечается, что наблюдательные исследования подвержены систематическим и случайным ошибкам в большей степени, чем экспериментальные. Это объясняется тем, что в силу особенностей дизайна наблюдательные исследования не могут контролировать ошибки, связанные с возможной несопоставимостью групп сравнения. Они способны выявлять наличие статистической связи между явлениями, но доказать, что связь носит причинно-следственный характер, могут только рандомизированные клинические исследования. Точность рандомизированного клинического исследования может быть повышена с помощью систематических обзоров с метаанализом.

Ключевые слова: дизайны клинических исследований, наблюдательные контролируемые исследования, экспериментальные клинические исследования, систематический обзор, метаанализ, систематические ошибки, доказательность выводов, уровни достоверности доказательств, уровни убедительности рекомендаций

Getting new scientific knowledge in the field of modern clinical medicine is mainly based on the results of clinical epidemiological trials. They enable to detect the factors leading to occurrence and progression of diseases, estimate the quantitative input of these factors into development and subsequent course of diseases, stratify a population by the extent of risk and determine prognosis, monitor the level of risk factors and estimate effectiveness of preventive programs, plan clinical trials, formulate and check hypotheses.

The role of dyslipidemia, arterial hypertension, smoking and diabetes mellitus in development of atherosclerosis and associated diseases was mainly revealed owing to epidemiological trials. Clinical trials (CT) were conducted and treatment and prevention recommendations of these diseases were developed both at the population and individual levels [1].

In clinical epidemiology, several types of CT are used. They have different structures and are aimed at searching answers regarding some clinical issues about assessment of prevalence of pathological conditions, searching and studying the reasons or factors of risk of diseases, assessment of frequency, relative risk and prognosis of morbidity. The principal clinical issues

include assessment of effectiveness of preventive, diagnostic and therapeutic medical interventions.

Every task can be solved using a CT with a certain logical structure that includes methods of enrolling people into a trial, formation of comparison groups, collection of data, and methods of its analysis and interpretation. Design is a trial form created to search for answers to the set clinical questions. The design reveals a degree of accuracy for the result obtained during the trial, which shows real connections between the events.

In this article, attention is given to the factors that limit the degree of trial reliability associated with survey design; the structure and degree of reliability of various designs in comparative terms are considered. Recommendations regarding determination of evidence reliability levels and evidence strength levels are provided.

ACCURACY OF RESEARCH

Reliability of the trial is determined by its accuracy, which consists of the extent with which the results can be applied to other groups (external validity or generalizability), extent to which the trial can exclude the alternative explanation of the obtained results (internal validity) and extent of exact assessment of measured results (confidence) [2].

External sample validity is determined by the extent of representativity as related to its population [3]. Scientific clinical trials involve not the entire population suffering from the studied pathology or to whom the assumed risk factor is applied, but a part of this population (sample). If characteristics of these participants completely correspond to those of the population, i. e., the populations are representative, the obtained results can be applied to all people from this population. However, the sample can be representative only if it was formed using the random selection method. The random selection method deals with selection of all patients with this pathology and subsequent accidental or equally probable inclusion of representatives of all types of patients from the general population into the sample.

In medical trials, this is almost impossible. Thus, patients from clinical trials can differ from all the patients with the studied disease by age, gender and nationality, social status, material wealth, attitude to health, location, condition severity and many other characteristics. It makes the sample not accidental and not quite representative. In this case, the external sample validity is insufficient.

Conclusions associated with non-random samples can be applied to the general population with a certain proportion of errors (bias). This error occurs during sample formation and is called a systematic sample error.

In statistics, **a systematic error** means unintentional, but regular, non-accidental and unidirectional deviation of the calculated indicators from their actual values [4].

The less is the sample representative, the less exact is the trial, the more likely it is that other factors (errors) that distort conclusions influenced the trial results. The sample representativeness can be increased with numbers, thus, our trust in trials with a higher number of participants is stronger.

Internal validity is determined by how well the trial design can exclude alternative explanations of these conclusions. Differences in the results of the compared groups are not a mere consequence of the studied factor. There are other explanations, too. We can't exclude an effect on the result of other functions, which the researcher didn't plan to trial, failed to take into account or which he wasn't aware of, but which can also influence the outcome.

In case of irregular distribution of these factors among groups of comparison and control, effect of these factors will displace true results of the intervention and lead to inaccurate and erroneous conclusions. The factors cause unilateral bias (distortion) of trial results and are called systematic selection errors. The selection errors include all the factors that lead to incomparability of the studied group and control group.

The results of clinical trials can be influenced by other systematic errors such as errors obtained while collecting information, memory errors, withdrawal-related errors, errors that occur while assessing and analyzing the results, and some others [5, 6]. All systematic errors can make the differences visible, though they do not exist in reality, or, on the contrary, the real existing differences can be hidden. A systematic error can arise in any observations and at any stage of the trial. The sample size does not influence the systematic error value.

To be sure that the observed result is a consequence of the studied factor but not systematic errors, their significance should be excluded or reduced. This is achieved during sampling through increasing its representativeness or at the stage when comparison groups are formed during randomization. They can also be partially taken into account while analyzing the trial results. The principal method that minimizes the effect of the majority of systematic errors is represented by randomization, i. e., accidental distribution of patients among comparison groups. Meanwhile, systematic errors are also regularly distributed among comparison groups and fail producing the bias effect if the groups are large enough.

RELIABILITY OF RESEARCH

An accidental error is another explanation of differences in the results among the compared groups. **An accidental error** is a deviation of a single observation (or measurement) from its true value that occurs while processing accounting documents, during measurement or registration of data due to an accidental combination of circumstances. There is an equal probability that an accidental error can result in overestimation or underestimation of research results. Any observations are exposed to accidents. Complete exclusion of accidental errors is not possible though they can be minimized using more exact methods of trial parameter estimation, for instance, standardized ones, or by increasing a number of patients in a trial.

An accidental error can be estimated and accounted at the stage of statistical analysis of results, which allows to answer a question about the probability of obtaining the results in an accidental way. In medical research, the accessible level of probability of getting an accidental result arises when p is less than 0.05 [7].

The marked accidental error is commonly observed in small samples with highly non-homogenous characteristics (both inhabitants of cities and villages, men and women, those with and without bad habits of a wide range of ages are included). The higher the sample heterogeneity, the greater the probability of an accidental error and the more people should be included into comparison groups to increase the reliability of conclusions. Even a marked accidental error does not provide for a bias (does not distort the research result), but may prevent revealing statistical reliability of the obtained results.

The level of systematic errors is controlled by strict fulfillment of design requirements. Owing to design characteristics, clinical trials can control the effect of systematic errors to a different extent and can have certain limitations regarding the degree of reliability. It should be noted that some factors such as the use

of improper statistical methods of analysis, lack of adjustment for systematic and accidental errors, negligent data handling can distort the trial results irrespective of the selected design.

Clinical trials with various designs are used in scientific medicine. Three basic designs can be found among them. Their task is to find and examine the causal relationships. They include case-control trials, cohort trials and randomized clinical trials [8, 9].

DESIGN AND EVIDENTIAL VALUE OF CASE-CONTROL TRIALS

These are observational trials when researchers do not only interfere in the natural course of disease occurrence and distribution. They only observe how the situation that doesn't depend on them is developed, collect data on the examined issue and make conclusions [10].

Case-control trials are used to reveal unknown risk factors of known diseases. To detect the relation between the clinical outcome and preceding effect of the assumed factor, two groups of people are included into the trial. The main group includes those with a disease or condition that seems interesting to the researchers. The group is called 'cases'. The control group involves people without such a disease or condition. A history of all trial participants includes presence or absence of certain factors that could be the reason for development of the studied disease. The both groups are then compared by the rate of potential risk factors for this outcome, and the statistical significance of these differences is determined.

The feature of the case-control trial means that this design doesn't mean randomization while making comparison groups, leading to incomplete comparison of the main and control groups due to systematic errors.

A number of 'cases' is selected among patients with the studied disease or condition and to whom the researcher would like to disseminate the conclusions he is determined to obtain. The group of 'cases' should always be representative of the studied population. Insufficient representativity of the group of 'cases' (sampling error) can result in unproper generalization of the trial results.

The researcher selects a group of 'controls' based on characteristics of the 'case' group but not in the result of randomization, which is a source of systematic selection failure. When selecting the control persons, the main condition consists in their maximum comparison with a group of cases based on all the basic characteristics, except for the studied disease. To obtain a more reliable result, a group of 'controls' should be comparable with a group of 'cases' to the greatest extent [11]. For this, 'controls' should be selected from the same population as the 'cases', preferably during the same period of time. For instance, both the 'cases', and the 'controls' should be selected from among the people admitted at the same hospital, receiving treatment at the same outpatient clinic, living in the same district or working at the same enterprise.

In case of insufficient comparison, cases and controls can differ by the condition severity, concomitant pathology, social status, bad habits, and use of medicines influencing one's health, etc. [12].

To reduce the selection error, a paired design is used. It ensures an individual selection of 'cases' for every group participant that corresponds by a set of characteristics to a control group participant [13]. As a result, researchers obtain almost similar groups of comparison with the only difference: presence or absence of the studied disease.

One of the systematic selection errors, when a true result is displayed in a wrong way, can be due to an effect of an unknown

or unaccounted factor. The factor can produce a simultaneous influence both on the outcome, and on the studied factor of the disease. The factor is called 'a confounding factor' or 'a confusing variable' (confounder) [14].

A trial that examined a link between a birth order (1st, 2nd, 3rd child, etc.) and presence of Down's disease can serve as an example. In this trial, maternal age will be a confusing variable as it influences both the outcome (a higher maternal age is directly associated with a possible development of Down's syndrome in a child) and a birth order when every next child, except for twins, is born when the mother is older than she was when she gave birth to the 1st child.

The presence of confusing factors can be clear or not. Thus, the conclusions obtained based on observational trials can fail to display a real effect of using the examined intervention.

Retrospective trials have typical systematic **errors at the stage of data collection and memory errors**. During case-control trials, a search for causal relationships always moves from a consequence to the assumed reason, i. e., retrospectively. At the initial stage of a retrospective trial, a researcher has already been informed of an interesting outcome and collects data about the events (possible risk factors) that took place in the past. Medical records or outpatient cards stored at healthcare organizations (i. e., secondary information), recollections of patients, interviews with their relatives or questioning results constitute a source of information. This is associated with occurrence of information systematic errors and memory errors. Data registered in medical documentation were collected for other purposes and tasks, the researcher failed to participate in their collection and frequently doesn't know who and when collected the data.

Archive information may not correspond to the purpose of the conducted trial to the full extent, it may not be collected properly and some data can be lacking. Data collected from people can insufficiently reflect the events of the past. Selective memory of a patient and healthy person can make a difference.

For instance, a sick person can recollect the events potentially related to occurrence of this disease better than a healthy one, and fail to recollect certain facts that can seem interesting to a researcher. Memory failures are particularly true if they relate to data about the effect of the studied risk factor, which is a principal shortcoming of all retrospective trials [5].

Data registered in medical documents were collected for other purposes and tasks, whereas the researcher didn't participate in their collection and frequently does not know who and when gathered the data.

Along with sampling errors, selection errors and data collection errors, case control trials are not protected from accidental errors, providing many alternative explanations to the obtained results. Substantiation of this type of trials is not very high.

Statement of hypotheses about disease risk factors and conditions form the result of the trial. The hypotheses should be confirmed during more exact cohort trials.

Though the case-control trial doesn't prove there is a causal relationship, such trials are the only suitable ones to study the risk factors of rare diseases [7].

DESIGN AND EVIDENTIARY VALUE OF COHORT TRIALS

Cohort trials are also observational. The data are collected by observing events without a researcher's intervention [8].

The purpose of the trial is to search and detect unknown consequences of effects by assumed risk factors on human health and examination of interrelations. For the study purposes,

a group of people (cohort) that should be a representative sample of the population is selected from a general set (population). A cohort is a group of people with common characteristics or experience during a certain period of time when new disease cases are expected to occur. People living in the same city, exposure to hazardous substances, undergoing a certain medical procedure, belonging to representatives of the same profession or social group, being born at a certain period of time, etc. belong to a unifying feature.

The examined cohort is represented by people influenced by the examined risk factor, whereas the control cohort includes people not influenced by the examined factor [15]. The control group is selected from the same population the cohort is composed of or another cohort that was affected little or not affected at all, with all the other characteristics being most similar to those of the studied group. These cohorts are observed for some period of time to understand, which outcomes can lead to this risk factor. An obligatory condition of inclusion of these people into the examined and control cohorts is represented by a lack of the studied disease at study enrollment.

Then the both groups are compared by the rate of disease development, the value of relative risk that confirms the relation between a risk factor and outcome probability is determined and the statistical significance of differences is estimated.

Cohort trials are called prospective if the search for the causal relationship moves from the reason to the assumed effect. In other words, the cohort is being observed from initiation of the trial when the disease is still lacking; the observation is being continued for a period enough for the assumed outcome to develop. Meanwhile, the researcher can't know the outcomes beforehand excluding subjectivity while selecting those analyzed. In this case, the source of data is represented by data assumed to obtain during a trial and independently registered by the investigators, that's why they are more reliable and correspond to the study purposes to a greater extent.

Cohort trials can be retrospective when at the beginning a researcher has information at his disposal and collects data about the events that took place in the past. However, the groups are formed depending on the presence or absence of risk factors. Like in other retrospective trials, data are collected using archival documents (case history, questionnaires, results of participants' survey, etc.). The researcher analyzes the past data by tracing morbidity and mortality for all members of the studied groups until now [15].

Cohort trials are not exempt from systematic and accidental errors. **Errors related to cohort representativeness** can occur if its composition does not completely correspond to the population it was selected from [16]. The situation is possible when the cohort includes visitors of a certain medical center where the patients can enter not accidentally but because they live nearby or where they can be referred because of a severe condition or because they can pay for the medical services, whereas the general population includes not only patients of medical centers but also those from municipal hospitals and outpatient clinics. The differences can relate to the age, gender, social and economic status, living conditions, health, etc. It is sometimes difficult to generalize the results even of large clinical trials.

For instance, it is difficult to determine the rate of representativeness of a rich American city Framingham (Framingham trial of IHD risk factors) at least for the USA, or that of the trial on British doctors at least for representatives of other professions in Great Britain (trial of the association between cigarettes and lung cancer).

Correct cohort assessment influences the possibility to transfer the data obtained during the trial to the initial population and population with similar characteristics. The larger is its size, the more exact are the obtained data, the more they correspond to the general population [17].

Data and memory errors while conducting cohort trials with retrospective collection of data show that it is difficult to reconstruct the events of the past without distortions. Some documents recording the effect (for instance, a harmful factor in the past) can be lost, whereas recollections of relatives are not exact. Data collection and memory errors result in masking the influencing effect and distorted conclusions.

Another error observed during prospective cohort trials is represented by the **error of withdrawal from the study**. Depending on the examined disease, prospective cohort trials can last for a long time — for years or even for decades. In such duration of observations, some patients can withdraw from the study due to their shift to another place of residence, refusal to participate, death, loss of contacts, etc. A decreased number of cohorts is associated with reduced statistical power and, as a consequence, less reliability of the study. It is believed that when over 10% of the cohort is lost, the study results are doubtful, whereas dropout of over 20% of participants displays its uncertainty [9, 18].

Cohort trials can be associated with **selection errors** that include all the factors except for the examined ones, which, in case of irregular distribution between the studied and control cohorts, can result in the lack of their compatibility and influence the study results.

Examples can include differences in treatment, number of visits to doctors or any other values. Inclusion of patients into trials at different times can result in significant differences among the compared groups. For instance, during mixed retrospective and prospective trials no difference in terms and exact diagnosis, past approaches to therapy (say, 15 years ago) and today can be taken into account. In this case, changed outcomes can rather be explained by a difference in assessment of disease severity than by treatment effect.

Undocumented or unknown confounding factors are found among the factors that can be a source of a systematic error. The confounding factors produce such an effect that the effect of the studied factor can be overestimated or underestimated. To exclude the effect of known confounders, the both groups should be comparable to the greatest extent by the largest number of parameters, except for the examined ones [19]. While analyzing data, there are methods enabling to consider the effect of all factors we are aware of. But even after all amendments the confounding factors not known to us can be left unaccounted. The balance of unknown confounders is achieved through randomization. Randomization in cohort trials is impossible, as the observational approach to studying the relations between the events excludes accidental distribution of people into the compared groups.

Impossible control over unknown confounders is a serious disadvantage, that makes observational trials different from a randomized experiment. Unfortunately, it is impossible to get rid of this shortcoming of observational trials. That's why we get an uncomplete level of evidence of observational trials, and cohort trials, in particular [14].

Though prospective cohort trials do not exclude all the possible mistakes, they are the most evidence-based among observational trials and reflect a causal relationship in a more precise way. Cohort design is considered the best when it is necessary to examine the effect of potentially harmful risk factors on disease occurrence, i. e., when human experiments are not possible.

DESIGN AND EVIDENTIAL VALUE OF RANDOMIZED CLINICAL TRIALS

A randomized clinical trial (RCT) is an experimental study where a researcher simulates a clinical situation which suits the best to examine the causal relations between the studied phenomena. As a rule, experimental trials are conducted to check the cause-and-effect hypotheses while examining effectiveness of various methods of treatment and prevention, both drug-induced, and not.

In experimental trials, it is ethically acceptable to examine only the effects of factors, which, as assumed, deliver benefit to a patient. Thus, artificial intervention into the natural history of events occurs at the expense or with the elimination of suspected factors that cause diseases or while administering medicinal agents, using methods or performing activities able to produce a favorable effect on the studied disease [20–22].

Design of a RCT is much like design of cohort trials. A group of people which is a representative sample is selected from the general set (population) based on strict criteria of inclusion and exclusion. Then the included patients are accidentally (irrespective of a researcher's will) distributed into the study group (obtain the studied intervention) and control group (obtain placebo or known intervention with known effectiveness). During the trial, the participants are under a planned observation with registration of their subjective and objective condition. At the end of the trial, the differences in the results of the both groups are assessed along with their statistical significance.

Experimental trials can be prospective, retrospective and mixed (historical control study). During a prospective trial, the researcher should collect and register data about a patient; during a prospective trial, data are collected using archival medical documentation or interviews of patients, decreasing the reliability.

Design of the RCT differs from other types of trials by the possible procedure of randomization. It is the randomization that allows to neutralize the significance of the majority of systematic errors occurring during a CT. They involve **systematic errors creating a disbalance between the comparison groups including confounding.** Thus, there is a low probability that the obtained results are not due to the studied intervention, but have an alternative explanation. However, it is true only when the researcher fails to violate the basic randomization principle. According to it, every sampling member should have equal chances to be included both in the studied group, and in the control group [23].

The reason for incorrect randomization is inclusion of patients into the group of comparison by indications, order of selection, days of the week, case history numbers, insurance policy or date of birth. These grounds introduce a systematic error into the process of formation of comparison groups. It is better to use a table of accidental figures, methods of envelopes or centralized computer distribution of treatment options.

When the principle of equal changes is violated, no regular distribution of the effect of systematic errors occurs and the evidence level of this trial goes down reaching the level of cohort observation [20].

Randomized historic control trials are less exact as compared with prospective ones due to **errors that occur during collection of data and memory errors**, and because of possible differences in diagnostic criteria and accuracy of the examination of patients from the control group. **A systematic error associated with withdrawal** of patients from the

long-term study requires correction at the stage of result assessment.

The randomized trials do not completely exclude **sampling errors** that reduce the possibility to apply the obtained results to a wider population of patients. For instance, the majority of RCT are conducted with relatively young patients without concomitant diseases, whereas the medicinal agents studied under these conditions are consumed by elder patients suffering from many diseases. Randomized trials performed on selective groups have low representativity. The use of selective groups is justified while studying a novel medicinal product to confirm its pharmacological activity and determining its safe doses during the first stages of CT.

It is desirable to detect and eliminate some **systematic errors associated with positive expectations** of patients related to their participation in a CT (placebo effect) at the stage of selection. It is necessary because different expectations of patients in the compared groups can influence the study results to the greatest extent. Psychological patterns and expectations arise not only among patients, but also among medical personnel who conducts the study. It is due to a prejudiced attitude of an investigator while selecting patients and subjectivity when assessing the borderline results of the study. To exclude these psychological phenomena, it is necessary to limit awareness of researchers concerning the provided medicinal agents in the comparison groups (blind, double-blind trials). It is shown that a lack of double blinding can increase effectiveness of medicinal agents by 15–20% in average [21].

The use of a blind method regarding to patients, doctors and researchers estimating the clinical outcomes and statisticians enables significantly reduce the probability of a systematic error of that type.

In spite of the randomization, the compared groups can be heterogenous due to insufficient sampling size and associated increased effect of an **accidental error**. The probability of an accidental error is increased in case of high heterogeneity (nonuniformity) of the population that constitutes the sample.

Thus, small RCT or RCT held in one center only, have insufficient representativity (non-homogenous sample), reduced internal validity (disbalance of compared groups) and insufficient reliability (increased probability of an accidental error). As the accidental error and sample heterogeneity are decreased with size, trust in large multi-centered RCT is always higher. To ensure better reliability, it is necessary to perform multiple checks of RCT results to prove the causal relationship. It is desirable that the study should be repeated by various researchers on many differentiated samples, at different time and under various conditions. It is impossible to completely exclude the effect of an accidental error, that's why there is always 5-percent probability that the result obtained during the study is due to accidental occurrence of circumstances [24].

In spite of possible problems, properly planned and conducted RCT enable to obtain highly significant conclusions and are a golden standard of evidence-based CT.

STRUCTURE AND EVIDENTIAL VALUE OF SYSTEMATIC REVIEW AND META-ANALYSIS

As even RCT are not very exact, methods of evidence-based medicine such as systematic reviews including or excluding meta-analysis have been developed.

A systematic review (SR) is an analytical study of analytical observational and experimental trials presented in literature and serves as a tool of secondary analysis of scientific publications.

The study begins with the formulation of a clinical issue that requires an answer. It is about effectiveness of treatment, prevention or diagnostic methods. The best works that are used to investigate the same problem and that have a similar structure possess the most powerful design and are conducted in the most scrupulous manner. The trials are selected based on distinct inclusion and exclusion criteria, which should be substantiated and determined beforehand. Then the results of all trials that passed the selection are generalized. An answer to the clinical question is provided based on these generalized results. It can be expressed as a confirmed causal relationship, its denial or when qualitatively conducted primary trials are not enough to give a definite answer to the question [25].

The source of data for SR is constituted by all discovered published analytical observational and experimental trials about the examined clinical issue. The data are searched through electronic information data bases, which include only materials that correspond to certain criteria of methodological quality. These are Medline, Embase, Cochrane Library, and eLibrary.ru.

However, not all trials can be included into SR, as SR generalizes results of relatively homogenous trials only. Generalization of study results significantly different by patient's characteristics, various aspects of using the compared medicinal products, assessment criteria of the studied outcome is considered illegal, as these differences increase the non-homogeneity (heterogeneity) of the generalized data and reduce the significance of conclusions.

SR can include the use of the statistical method generalizing the results of several primary trials as if this could be one large study and make a common statistical conclusion on its basis. The method is called meta-analysis. United trials provide for a larger sample for analysis and greater statistical power. This increases exactness of assessing the effect of the analyzed intervention and improves substantiation of systematic review data with metanalysis as compared to separate experimental or descriptive trials.

The metanalysis can detect the effect failed to be detected during other experiments due to insufficient statistical power (a small number of participants in every experiment), it also enables a general conclusion based on several trials with various and even contradicting results [26, 27].

In spite of all advantages, meta-analysis is also not free from the effect of systematic errors and can contain false conclusions. It includes systematic errors such as errors of inclusion into SR and publication bias [28].

Inclusion errors reflect a low quality of systematic review. It is known that quality of meta-analysis significantly depends on quality of included initial trials and articles, i. e., on quality of the systematic review it is based on. The meta-analysis carried systematic errors of all primary works it consists of. When the published scientific literature reflects false assertions, meta-analysis also confirms false results.

Publication bias occurs when certain conducted stud trials without statistically significant differences in results between the groups of comparison or with results not different from the

known data remained unpublished and weren't included into the meta-analysis. Then proportion of publications with positive results exceeds the real value resulting in overestimation of the averaged effect.

Disturbed methodology of SR is an insufficiently complete search of data, non-compliance with strict selection criteria and inclusion of low-quality trials leads to accumulation of systematic errors and reduces the veracity of SR results. Thus, a large high-quality RCT can provide more reliable results as compared with meta-analysis of some small ones.

Thus, systematic reviews and high-quality meta-analysis form the basis of evidence-based medicine analytical base and a very valuable tool while taking decisions about the choice of the most effective and safe methods of treatment and prevention.

HIERARCHY OF EVIDENCE OBTAINED IN CLINICAL TRIALS

Results of CT with various designs are currently used to develop clinical recommendations on prevention, diagnostics, treatment and rehabilitation. To understand the relative force of their substantiation, a hierarchy of evidence defined as ranking of CT with various designs by the degree of their liability to systematic errors was suggested [29]. At the top of the hierarchy, a method with the largest freedom from the systematic bias is located. It means that the true effect is close to the one obtained in the trial. At the lowest level of the hierarchy are types of trials not free from many systematic errors, which significantly reduces confidence in truthfulness of the obtained results.

Classification of the levels of evidence with some differences in CT assessment protocols are developed and utilized in various countries and large medical organizations. In the Russian Federation, the evidence levels of CT included into clinical recommendations are assessed based on the results of one or several CT of the highest rank in accordance with a single scale along with requirements approved by the order of the Ministry of Health of the Russian Federation as of Febr. 28, 2019 Np. 103н.

Level of evidence (LE) is a level of confidence indicating that the found effect related to the medical intervention is true [30]. Five levels of evidence reliability are provided (table 1).

Recommendations made using CT results are also ranged based on the **evidence level** (EL), which is determined as the rate of confidence in validity of the intervention effect and that following recommendations will do more good than it does harm.

The evidence level is determined based on assessment of methodological quality of CT, consistency of results of CT used to assess the EL, and importance of outcomes.

Methodological quality of CT is estimated using the respective point questionaries developed separately for SR, RCT, cohort trials and case-control trials. The CT results are considered as agreed if all the CT have effects of the same direction and if, as a consequence, the same conclusions are made. It means that there is an advantage of intervention A over

Table 1. Scale for determining the levels of evidence (LE) for therapeutic, rehabilitation and preventive interventions

LE	Hierarchy of designs of clinical trials (in descending order of the evidence level from 1 to 5)
1	Systematic review of RCT using meta-analysis
2	Other RCT and systematic reviews of trials of any design (except for RCT using meta-analysis)
3	Non-randomized comparative trials, including cohort ones
4	Non-comparative trials, description of a clinical case or set of cases, a case-control trial
5	Preclinical studies (substantiating the mechanism of action for this intervention) or experts' opinion

Table 2. Scale determining the evidence strength levels

LE	Interpretation
A	**Strong recommendation** (simultaneous fulfillment of two conditions): • all considered criteria of effectiveness (outcomes) are important, • all trials are of high or satisfactory methodological quality, • their conclusions on interesting outcomes are agreed
B	**Conditional recommendation** (if at least one condition is met): • not all considered effectiveness criteria (outcomes) are important, • not all trials have a high or satisfactory methodological quality, • their conclusions on interesting outcomes are not agreed
C	**Weak recommendation** means a lack of evidence of proper quality (if at least one condition is met): • all considered criteria of effectiveness (outcomes) are not important, • all trials are of low methodological quality, • their conclusions on interesting outcomes are not agreed

intervention B in all CT with a higher design [31]. Based on CT results, importance (significance) of outcomes is determined as important and not important. Important outcomes include all clinical outcomes ('solid end points'), surrogate outcomes estimated by validated scales, surrogate outcomes with proven associated clinical outcomes based on CT results.

Not important outcomes include surrogate outcomes in the lack of CT that confirm association with clinical outcomes ('solid end points'). These are values of non-validated clinical scales, laboratory values, subjective assessments of patients (including using the visual analogue scales), and duration of symptoms.

Assessment of the level of evidence of recommendations for diagnostic, therapeutic and preventive interventions and rehabilitation activities is also carried out in accordance with a single scale and requirements approved by order of the Ministry of Health of Russia as of February 28, 2019 No. 103н. As far as evidence goes, there are strong, conditional and weak recommendations denoted using Latin letters A, B, C (table 2).

Proper assessment of evidence levels of recommendations and levels of confidence of CT, on which recommendations are based, should ensure their high scientific validity, which corresponds to requirements of medicine based on evidence.

CONCLUSIONS

Various clinical epidemiological trials intended to achieve different purposes and tasks are applied as a tool to obtain new knowledge in the field of medicine. CT differ by their structure and exactness used to estimate the cause-and-effect relations between the phenomena. Thus, while estimating accuracy of these conclusions, we need to be patient about the limitations typical of various designs. Exactness of CT depends on many factors, which can distort the obtained results as compared with their true values. The influence of these factors (systematic and accidental errors) enables to make alternative conclusions about the reasons for the discovered differences.

Designs of various CT admit the influence of a greater or a smaller number of these factors. It is reflected on the reliability of results of CT. Neither study is free from systematic and accidental errors. However, observational trials are subject to them to a greater extent than experimental ones. This is explained by the fact that due to design characteristics observational trials can't be used to control errors associated with the possible non-correspondence of comparison groups. They can be used to detect a statistical relation between the phenomena but only RCT can prove that this is about a causal relation. Exactness of RCT can be increased with systematic reviews and meta-analyses.

References

1. Martsevich SIu, Drozdova LIu, Voronina VP. Zdorov'e i obrazovanie vracha: dve sostavliaiushchie uspekha. Ratsional'naia Farmakoterapiia v Kardiologii. 2010; 6 (1): 73–76. Russian.
2. Glants S. Mediko-biologicheskaia statistika. M.: Izd-vo «Praktika», 1999; 459 p.
3. Begrambekova Iu L. Metodologiia analiza klinicheskikh issledovanii i opredeleniia ikh sviazei s klinicheskoi deistvitel'nost'iu. Chast' I. Kardiologiia: Novosti. Mneniia. Obuchenie. 2016; 4 (11): 69–75. Russian.
4. Everitt BS, Howell DC. Encyclopedia of Statistics in Behavioral Science. Chichester: John Wiley & Sons Ltd., 2005; 1451–62.
5. Vandenbroucke JP, Von Elm E, Altman DG, et al. Povyshenie kachestva otchetov o nabliudatel'nykh issledovaniiakh v epidemiologii (STROBE): raz"iasneniia i utochneniia. Voprosy sovremennoi pediatrii. 2022; 21 (3): 173–208. DOI 10.15690/vsp.v21i3.2426. Russian.
6. Martsevich SIu, Kutishenko NP, Lukina IuV, Lukianov MM, Drapkina OM. Klinicheskie issledovaniia po otsenke lekarstvennoi terapii. Vidy, dostovernost' rezul'tatov, mesto v sovremennoi dokazatel'noi meditsine. Ratsional'naia Farmakoterapiia v Kardiologii. 2021; 17 (2): 249–255. Russian.
7. Bavrina AP. Nabliudatel'nye epidemiologicheskie issledovaniia i osobennosti predstavleniia rezul'tatov v nauchnom otchete. Meditsinskii al'manakh. 2021; 2 (67): 83–89. Russian.
8. Narkevich AN, Vinogradov KA. Dizain meditsinskogo issledovaniia. Sotsial'nye aspekty zdorov'ia naseleniia. 2019; 5 (65): 13. Available from URL: http://vestnik.mednet.ru/content/view/1108/30/lang,ru. DOI: 10.21045/2071–5021–2019–65–5–13. Russian.
9. Simpson MR. Kohortstudier [Cohort studies]. Tidsskr Nor Laegeforen. 2021 Oct 25; 141. Norwegian. DOI: 10.4045/tidsskr.21.0511. PMID: 34726047.
10. Grjibovski AM, Ivanov SV, Gorbatova MA. Case-control studies in health sciences. Nauka i Zdravoohranenie [Science & Healthcare]. 2015; 4, 5–17. Russian.
11. Pearce N. Analysis of matched case-control studies. BMJ. 2016 Feb 25; 352: i969. DOI: 10.1136/bmj.i969. PMID: 26916049; PMCID: PMC4770817.
12. Koterov A N. Kriterii prichinnosti v medikobiologicheskikh distsiplinakh: istoriia, sushchnost' i radiatsionnyi aspekt. Soobshchenie 1. Postanovka problemy, poniatie o prichinakh i prichinnosti, lozhnye assotsiatsii. Radiatsionnaia biologiia. Radioekologiia. 2019; 59 (1): 5–36. DOI: org/10.1134/S0869803119010065.Russian.
13. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 2018 Jan; 33 (1): 5–14. DOI: 10.1007/s10654–017–0325–0. Epub 2017 Nov 3. PMID: 29101596; PMCID: PMC9188848.
14. Thoresen M. Konfundering — et tilbakevendende problem [Confounding — a recurring problem]. Tidsskr Nor Laegeforen. 2018 Feb 5; 138 (3). Norwegian. DOI: 10.4045/tidsskr.17.0795. PMID: 29411587.

15. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract. 2009; 113 (3): c214–7. DOI: 10.1159/000235241. Epub 2009 Aug 18. PMID: 19690438.

16. Kholmatova KK, Khar'kova OA, Grzhibovskii AM. Osobennosti primeneniia kogortnykh issledovanii v meditsine i obshchestvennom zdravookhranenii. Ekologiia cheloveka. 2016; 4: 56–64. Russian.

17. Cortegiani A, Absalom AR. Importance of proper conduct of clinical trials. Br J Anaesth. 2021 Feb; 126 (2): 354–356. DOI: 10.1016/j.bja.2020.09.030. Epub 2020 Oct 26. PMID: 33121749.

18. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 407–416. DOI: 10.1111/aogs.13319. Epub 2018 Mar 5. PMID: 29415329.

19. Pripp AH. Matching i kohortstudier [Matching in cohort studies]. Tidsskr Nor Laegeforen. 2020 Jan 27; 140 (2). Norwegian. DOI: 10.4045/tidsskr.19.0308. PMID: 32026854.

20. Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. 2020 Jul; 158 (1S): S79–S87. DOI: 10.1016/j.chest.2020.03.013. PMID: 32658656; PMCID: PMC8176647.

21. Kennedy-Martin T, Curtis S, Faries D, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015; 16: 495. DOI 10.1186/s13063–015–1023–4.

22. Petrovskaya YuA, Ogorodova LM, Petrovskiy FI. Principles of Decision Making on the Prescription of Medicines. What else Do You Need to Know when Interpreting the Results of Clinical Trials? Pediatric pharmacology. 2018; 15 (4): 343–348. DOI: org/10.15690/pf.v15i4.1949. Russian.

23. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials. 2017; 18 (1): 498–510. DOI:10.1186/s13063–017–2240–9.

24. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 380–387. Available from URL: DOI: 10.1111/aogs.13309. Epub 2018 Feb 27. PMID: 29377058.

25. Donato H, Donato M. Etapas na Condução de uma Revisão Sistemática [Stages for Undertaking a Systematic Review]. Acta Med Port. 2019 Mar 29; 32 (3): 227–235. Portuguese. DOI: 10.20344/amp.11923. Epub 2019 Mar 29. PMID: 30946795.

26. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. Urol Oncol. 2008 May-Jun; 26 (3): 320–9. DOI: 10.1016/j.urolonc.2006.12.001. Epub 2007 Nov 7. PMID: 18452828.

27. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018 Sep; 74 (3): 785–794. DOI:10.1111/biom.12817. Epub 2017 Nov 15. PMID: 29141096; PMCID: PMC5953768.

28. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008 Aug 28; 3 (8): e3081. DOI: 10.1371/journal.pone.0003081. PMID: 18769481; PMCID: PMC2518111.

29. Stegenga J. Down with the Hierarchies. Topoi. 2014; 33 (2): 313–322.

30. Prikaz Ministerstva zdravookhraneniia Rossiiskoi Federatsii ot 28.02.2019 № 103n "Ob utverzhdenii poriadka i srokov razrabotki klinicheskikh rekomendatsii, ikh peresmotra, tipovoi formy klinicheskikh rekomendatsii i trebovanii k ikh strukture, sostavu i nauchnoi obosnovannosti vkliuchaemoi v klinicheskie rekomendatsii informatsii". Available from URL: http://www.consultant.ru/document/cons_doc_LAW_324276/.

31. Omelianovskii VV, Avksenteva MV, Zhelezniakova IA, Sukhorukikh OA, Khachatrian GR, Galeeva ZhA, et al. Klinicheskie rekomendatsii kak instrument povysheniia kachestva meditsinskoi pomoshchi. Onkopediatriia. 2014; 4 (4): 246–259. Russian.

Литература

1. Марцевич С. Ю., Дроздова Л. Ю., Воронина В. П. Здоровье и образование врача: две составляющие успеха. Рациональная фармакотерапия в кардиологии. 2010; 6 (1): 73–76.

2. Гланц С. Медико-биологическая статистика. М.: Изд-во «Практика», 1999; 459 с.

3. Беграмбекова Ю. Л. Методология анализа клинических исследований и определения их связей с клинической действительностью. Часть I. Кардиология: Новости. Мнения. Обучение. 2016; 4 (11): 69–75.

4. Everitt BS, Howell DC. Encyclopedia of Statistics in Behavioral Science. Chichester: John Wiley & Sons Ltd., 2005; 1451–62.

5. Vandenbroucke JP, Von Elm E, Altman DG и др. Повышение качества отчетов о наблюдательных исследованиях в эпидемиологии (STROBE): разъяснения и уточнения. Вопросы современной педиатрии. 2022; 21 (3): 173–208. DOI 10.15690/vsp.v21i3.2426.

6. Марцевич С. Ю., Кутишенко Н. П., Лукина Ю. В., Лукьянов М. М., Драпкина О. М. Клинические исследования по оценке лекарственной терапии. Виды, достоверность результатов, место в современной доказательной медицине. Рациональная фармакотерапия в кардиологии. 2021; 17 (2): 249–255.

7. Баврина А. П. Наблюдательные эпидемиологические исследования и особенности представления результатов в научном отчете. Медицинский альманах. 2021; 2 (67): 83–89.

8. Наркевич А. Н., Виноградов К. А. Дизайн медицинского исследования. Социальные аспекты здоровья населения. 2019; 5 (65): 13. [Электронный ресурс]. Режим доступа: URL: http://vestnik.mednet.ru/content/view/1108/30/lang, ru. DOI: 10.21045/2071–5021–2019–65–5–13.

9. Simpson MR. Kohortstudier [Cohort studies]. Tidsskr Nor Laegeforen. 2021 Oct 25; 141 c. Norwegian. DOI: 10.4045/tidsskr.21.0511. PMID: 34726047.

10. Гржибовский А. М., Иванов С. В., Горбатова М. А. Исследования типа «случай-контроль» в здравоохранении. Наука и здравоохранение. 2015; 4: 5–17.

11. Pearce N. Analysis of matched case-control studies. BMJ. 2016 Feb 25; 352: i969. DOI: 10.1136/bmj.i969. PMID: 26916049; PMCID: PMC4770817.

12. Котеров А. Н. Критерии причинности в медикобиологических дисциплинах: история, сущность и радиационный аспект. Сообщение 1. Постановка проблемы, понятие о причинах и причинности, ложные ассоциации. Радиационная биология. Радиоэкология. 2019; 59 (1): 5–36. DOI: org/10.1134/S0869803119010065.

13. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 2018 Jan; 33 (1): 5–14. DOI: 10.1007/s10654–017–0325–0. Epub 2017 Nov 3. PMID: 29101596; PMCID: PMC9188848.

14. Thoresen M. Konfundering — et tilbakevendende problem [Confounding — a recurring problem]. Tidsskr Nor Laegeforen. 2018 Feb 5; 138 (3). Norwegian. DOI: 10.4045/tidsskr.17.0795. PMID: 29411587.

15. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract. 2009; 113 (3): c214–7. DOI: 10.1159/000235241. Epub 2009 Aug 18. PMID: 19690438.

16. Холматова К. К., Харькова О. А., Гржибовский А. М. Особенности применения когортных исследований в медицине и общественном здравоохранении. Экология человека. 2016; 4: 56–64.

17. Cortegiani A, Absalom AR. Importance of proper conduct of clinical trials. Br J Anaesth. 2021 Feb;126 (2): 354–356. DOI: 10.1016/j.bja.2020.09.030. Epub 2020 Oct 26. PMID: 33121749.

18. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 407–416. DOI: 10.1111/aogs.13319. Epub 2018 Mar 5. PMID: 29415329.

19. Pripp AH. Matching i kohortstudier [Matching in cohort studies]. Tidsskr Nor Laegeforen. Norwegian. 2020 Jan 27; 140 (2). DOI: 10.4045/tidsskr.19.0308. PMID: 32026854.

20. Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. 2020 Jul; 158 (1S): S79–S87. DOI: 10.1016/j.chest.2020.03.013. PMID: 32658656; PMCID: PMC8176647.

21. Kennedy-Martin T, Curtis S, Faries D, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015; 16: 495. DOI 10.1186/s13063–015–1023–4.

22. Петровская Ю. А., Огородова Л. М., Петровский Ф. И. Принципы принятия решений о назначении лекарственных средств. Что еще надо знать при интерпретации результатов клинических исследований? Педиатрическая фармакология. 2018; 15 (4): 343–348. DOI:.org/10.15690/pf.v15i4.1949.

23. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. Trials. 2017; 18 (1): 498–510. DOI:10.1186/s13063–017–2240–9.

24. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018 Apr; 97 (4): 380–387. DOI: 10.1111/aogs.13309. Epub 2018 Feb 27. PMID: 29377058.

25. Donato H, Donato M. Etapas na Condução de uma Revisão Sistemática [Stages for Undertaking a Systematic Review]. Acta Med Port. 2019 Mar 29; 32 (3): 227–235. Portuguese. DOI: 10.20344/amp.11923. Epub 2019 Mar 29. PMID: 30946795.

26. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. Urol Oncol. 2008 May-Jun; 26 (3): 320–9. DOI: 10.1016/j.urolonc.2006.12.001. Epub 2007 Nov 7. PMID: 18452828.

27. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018 Sep; 74 (3): 785–794. DOI:10.1111/biom.12817. Epub 2017 Nov 15. PMID: 29141096; PMCID: PMC5953768.

28. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008 Aug 28; 3 (8): e3081. DOI: 10.1371/journal.pone.0003081. PMID: 18769481; PMCID: PMC2518111.

29. Stegenga, J. Down with the Hierarchies. Topoi. 2014; 33 (2): 313–322.

30. Приказ Министерства здравоохранения Российской Федерации от 28.02.2019 № 103н «Об утверждении порядка и сроков разработки клинических рекомендаций, их пересмотра, типовой формы клинических рекомендаций и требований к их структуре, составу и научной обоснованности включаемой в клинические рекомендации информации». [Электронный ресурс]. Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_324276/.

31. Омельяновский В. В., Авксентьева М. В., Железнякова И. А., Сухоруких О. А., Хачатрян Г. Р., Галеева Ж. А. [и др.]. Клинические рекомендации как инструмент повышения качества медицинской помощи. Онкопедиатрия. 2014; 4 (4): 246–259.