

Review

Manuel Martínez-Sellés¹ 
Luis Prieto-Valiente² 

Clinical study types and guidance for their correct post-pandemic interpretation

¹Servicio de Cardiología, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, CIBERCV. Universidad Europea, Universidad Complutense. Madrid. Spain
²Cátedra Internacional de Análisis Estadístico y Big Data. Universidad Católica de Murcia (UCAM). España. ADA.rt Advanced Data Analysis. Research and Teaching. SCIB

Article history

Received: 11 January 2024; Accepted: 10 February 2024; Published: 26 February 2024

ABSTRACT

Randomized clinical trials (RCTs) are key to the advancement of medicine and microbiology, but they are not the only option. Observational studies provide information on long-term efficacy and safety, are less expensive, allow the study of rare events, and obtain information more quickly than RCTs. On the other hand, they are more vulnerable to confounding factors.

Prospective exploratory pilot studies share many aspects with RCTs but are not subject to supervision by external commissions or mandatory registration. Multitesting can pervert the balance of publications in favor of the desired effect. Bonferroni's reasoning shows that if 10 studies are performed with an ineffective antibiotic, the probability that at least one will show $P < 0.05$ might be 40%. Scenarios in which there is intensive pressure to perform research, such as the recent pandemic, might result in many research teams trying to study the effect of an antimicrobial. Even if the drug has no efficacy, if 100 research teams conduct a study to assess its usefulness, it might be virtually certain that at least one will get a P value < 0.05 . If the other studies (with $P > 0.05$) are not published, the scientific community would consider that there is strong evidence in favor of its usefulness.

In conclusion, RCTs are a very good source of clinical information, but are not the only one. The systematic registration of all research can and should be applied to all types of clinical studies.

Keywords: methodology, statistics, clinical trials.

Correspondence:
Manuel Martínez-Sellés.
Cardiology Service, Hospital General Universitario Gregorio Marañón
C/ del Dr. Esquerdo 46, 28007, Madrid. Spain
E-mail: mmselles@secardiologia.es

Tipos de estudios clínicos y pautas para su correcta interpretación post-pandemia

RESUMEN

Los ensayos clínicos aleatorizados (ECA) son claves en el avance de la medicina y la microbiología, pero no son la única opción posible. Los estudios observacionales proporcionan información sobre la eficacia y la seguridad a largo plazo, son menos costosos, permiten estudiar eventos poco frecuentes y obtener información más rápido que los ECA. En su contra, son más vulnerables a factores de confusión.

Los estudios prospectivos piloto exploratorios, comparten con los ECA muchos aspectos, pero no están sujetos a la supervisión por parte de comisiones externas ni a la obligatoriedad de su registro. El *multitesting*, puede pervertir el balance de publicaciones a favor del efecto buscado. El razonamiento de Bonferroni muestra que, si se realizan 10 estudios con un antibiótico no efectivo, la probabilidad de que al menos uno arroje $P < 0,05$ puede ser del 40%. Escenarios en los que hay mucha presión para investigar, como la reciente pandemia, son propicios para que muchos equipos traten de estudiar el efecto de un antimicrobiano. Aunque el fármaco no tenga ninguna eficacia, si 100 equipos de investigación realizan un estudio para valorar su utilidad, puede ser prácticamente seguro que aparecerá al menos uno con $P < 0,05$. Si los demás estudios (con $P > 0,05$) no se publican, la comunidad científica consideraría que hay fuerte evidencia a favor su la utilidad.

En conclusión, los ECA son una muy buena fuente de información clínica, pero no la única. El registro sistemático de todas las investigaciones iniciadas puede y debe aplicarse a todo tipo de estudios clínicos.

Palabras clave: metodología, estadística, ensayos clínicos.

INTRODUCTION

Clinical studies (with human beings) can be divided, simply, into observational and prospective interventional studies (Figure 1). Although there are situations where aspects of these categories overlap, this outline helps to clarify the advantages and disadvantages of each type of design.

In observational studies, the researcher does not perform any type of action on the subjects, and only observes the clinical situation, personal, and environmental circumstances, looking for possible relationships between conditions and risk or preventive factors. In prospective interventional studies, the physicians undertake an action on the subjects and observe the response that occurs in the short- or medium-term. For example, an antimicrobial is given to patients with an infection and its possible benefits and side effects are investigated. Two large categories of prospective interventional studies are distinguished: exploratory pilot studies and randomized clinical trials (RCTs) that, in addition, are controlled, blinded, and registered. RCTs are key to the advancement of medicine and microbiology. However, they are not the only possible option, but rather they share importance with other types of studies. In this article, we will try to show: 1) the error of disregarding all information not coming from RCTs, 2) that the different

types of design do not compete with each other but rather are complementary, 3) the precautions to take when interpreting studies results.

OBSERVATIONAL/CASE-CONTROL STUDIES VS. PROSPECTIVE STUDIES

Observational studies can provide information on long-term efficacy and safety that is typically lacking in RCTs. In addition, they are less expensive, allow to study rare events and to obtain information more quickly. New and ongoing developments in data and analytics technology, such as artificial intelligence and big data, offer a promising future for these studies.

We should not question its usefulness, since many treatments and interventions aimed at preventing or solving health problems, with proven effectiveness, have not been subjected to rigorous evaluation using RCTs [1]. From the so-called evidence-based medicine, it is not correct to demand the adoption of interventions evaluated using only data from RCTs [2].

Observational studies include case reports and case series, ecological studies, cross-sectional studies, case-control (CC) studies, and cohort studies. The latter include clinical registries that are gaining increasing importance as a method to moni-

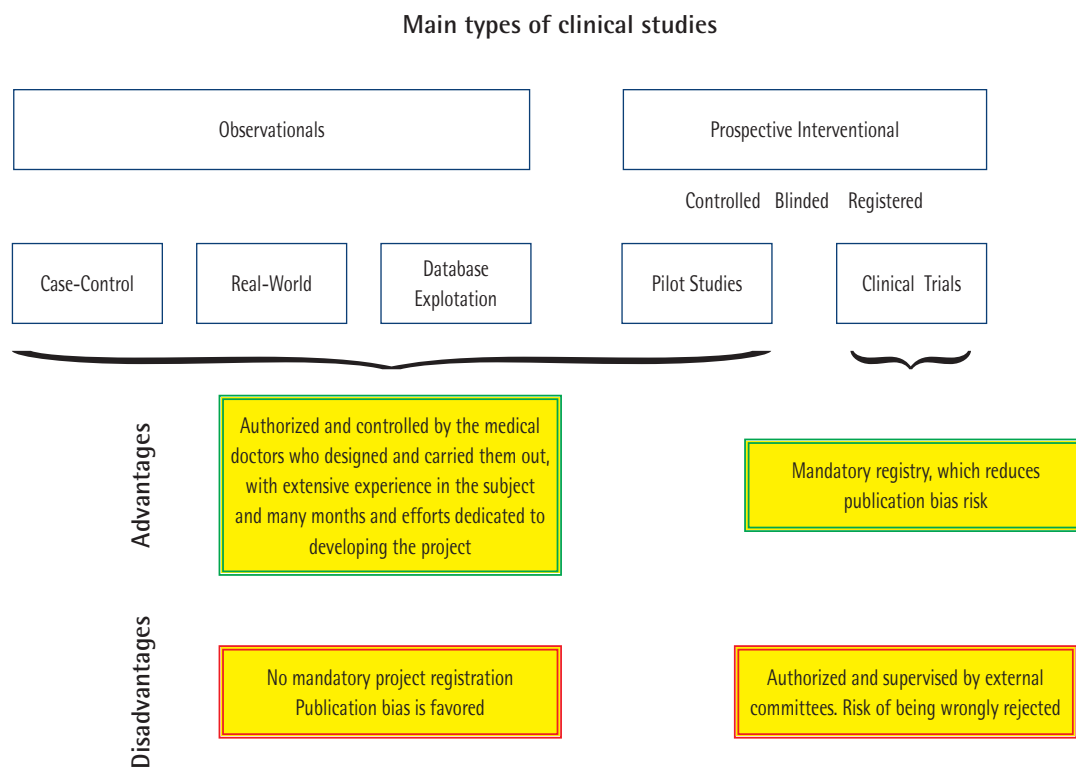


Figure 1 | Most important types of clinical studies with their main advantages (in the green box) and disadvantages (in the red box).

Type of study	Aim	Advantages	Disadvantages
Case reports and case series	To describe new or extraordinary events	Easy, detailed information to generate hypotheses	No generalizable
Ecological studies	To describe data about a population	Easy if you use routinely collected data	No subject data
Cross-sectional studies	To describe population profiles or outcome of interest at a single point in time	Relatively easy	Absence of temporality
Case-control studies	To identify risk factors for an event or disease	You can explore rare events	Limited to a single event
Cohorts	To estimate the incidence of events and their determinants	Longitudinal. Study multiple events and risk factors	Relatively difficult, expensive and time-consuming

tor and improve the quality of healthcare [3]. The main types of observational studies used in health research, their purpose and main strengths and limitations are shown in Table 1. Its purpose may be descriptive, analytical or both [4]. Descriptive studies are designed primarily to describe the characteristics of a studied population while analytical studies seek to address questions of cause and effect or, at least, generate hypotheses in that sense [4].

Among the various types of observational studies, CC designs occupy a relevant place. A recent survey carried out among medical doctors attending a course on "Expansion of basic concepts of statistical analysis in medicine", carried out at the Chair of Statistical Analysis and Big Data of the Catholic University of Murcia, showed that more than 80% were not able to identify the difference between the term "*controlled*" applied to prospective studies and the noun "*control*" used to designate CC studies (unpublished data). Although the word is the same, in prospective studies the control group does not receive the treatment we are studying, while in CC design the control group does not have the disease we are studying.

CC studies take a random sample of patients with a certain disease and another random sample of people without that disease (healthy or with health problems not related to the condition). In both groups, the frequency of individuals who had been exposed to a certain factor will be evaluated. To be more specific, supposing that we ask ourselves if following a certain diet is associated with the frequency of meningitis. A prospective study would consist of taking a sample of, for example, 180 people, randomizing them into two groups, with and without that diet, following them for twenty years and comparing the percentage of meningitis that appears in each group. In the CC approach, we take a sample of 90 cases of meningitis and another of 90 people who have not suffered that infection and we find out the percentage of people who followed that diet during the last 20 years. We see that the CC approach requires much less time, because the 20 years that patients must be followed with the prospective design, in the CC "have already passed" when the research is done. Another

advantage of CC designs is that, depending on how frequent meningitis is and the consumption of that diet, the statistical power can be much greater than with a prospective study. For instance, if 30% of people follow that diet and meningitis affects 1% of the population that does not use it and 4% of the population that uses it, (taking that diet multiplies the risk of suffering from the disease by 4), the statistical power of the study, for a P value of the test 0.05, is 0.92. The probability of finding a P value <0.05 is 0.92, that is, of every hundred studies that were carried out with this design, 92 would yield a P value <0.05. With a prospective study (RCT or pilot), the statistical power of the study, for a P value of the test 0.05, is 0.16. The probability of finding a P value <0.05 is 0.16, that is, of every hundred studies that were carried out with this design, only 16 would show a P value <0.05.

It must be considered that CC studies do not allow, in principle, to estimate relative risks (RR) that indicate how much greater the risk of disease is in patients on the diet than in those who do not follow it. CC studies allow us to estimate the odds ratio (OR), which is another way to quantify the harmful effect (in this example) of the diet. In this assumption it would be $OR = 3.13$, very close to the $RR = 3$. This proximity occurs whenever the risks involved are small, but with greater risks the value of the OR can be very different from the value of the RR. For instance, if the risk of a certain infection were 60% in those who follow the diet and 20% in those who do not follow it, the RR is 3, which tells us that following the diet multiplies the percentage of patients by 3. But, in this case, the OR is 6, which tells us that following the diet multiplies by 6 the number of sick people for every healthy person. With a CC study we could estimate this OR value, but not the RR value 3. Supplementary material depicts the concept of odd and OR.

Moreover, it must also be said that CC studies are much more vulnerable to the action of confounding factors. Part of this confusion is controlled with different types of multivariate analysis, regarding factors that have been recorded in each study. Another part can be controlled by sampling in a stratified and even paired manner. However, it will rarely be possible

to control as effectively as randomization does in prospective studies. In observational studies, adjustment for potential confounders can be performed and techniques such as propensity score, but only for a limited number of confounders and only those that are known and have been collected. Randomization in RCTs minimizes selection bias, while blinding controls information bias. Therefore, to, for example, know the effectiveness of a drug, RCTs provide the strongest evidence. However, this reality should not lead us to disregard the information provided by observational studies; information that is increasingly more precise and refined with new analysis methods. In fact, technological advances are changing the way in which observational studies are carried out [5].

NEW TECHNIQUES FOR ANALYSIS AND REPORT OF OBSERVATIONAL STUDIES

Even without randomizing subjects, methods have emerged in recent years that allow for less biased comparisons of two or more subgroups. Propensity score is a way of bringing together two or more groups for comparison, so that they appear to have been randomly assigned to an intervention or a comparator. In summary, the method involves logistic regression analysis to determine the probability (propensity) of each person within a cohort who was in the intervention, and then matching subjects who were in the intervention with those who were not, based on those propensity scores. The results are then compared between the two groups [6].

Increasing sophistication in data collection techniques, artificial intelligence, and the use of big data is also enabling continued improvements in the ability to conduct observational studies. Automatic linking of multiple data already offers a convenient way to capture results, even retrospectively. However, ethical considerations must be considered, such as whether informed consent may be required before data is linked, or who can access it. Machine learning already enables the capture, processing and analysis of unstructured text. New statistical techniques allow the imputation of missing data [7], a frequent problem in observational studies.

Unfortunately, the reporting of observational study research is often inadequate, making it difficult to evaluate their strengths and weaknesses and extrapolate results. The Strengthening The Reporting of OBServational studies In Epidemiology (STROBE) initiative [8] has allowed us to develop recommendations on how an observational study should be designed and reported. This initiative focuses on the three main types of observational studies: cohort, CC, and cross-sectional. It is based on a checklist with 22 items (the STROBE statement) that relate to the title, abstract, introduction, methods, results, and discussion sections of the articles. The vast majority of the items (18 of them) are common to the three types of study and four are specific for cohort, case-control or cross-sectional studies. STROBE is to observational studies what the Statement of Consolidated Standards of Reporting Trials (CONSORT) is for clinical trials. Both initiatives have helped improve the quality of studies and, above all, the way they are reported.

Similar initiatives exist for other areas of research, for example, for reporting meta-analyses or diagnostic studies. In any case, STROBE is not a panacea nor does it free us from the biases and limitations inherent to observational studies. Yes, it is a useful tool that provides guidance on how to report the results of observational research correctly. Although these recommendations are not prescriptions for designing or conducting studies, they must be considered when planning any observational study. Furthermore, while clarity of reporting is a prerequisite for evaluation, the checklist is not an instrument for evaluating the quality of observational research. This means that an observational study that meets all the STROBE items may still have important gaps. The STROBE statement is a key checklist before writing a paper with the results of an observational study. Its sections include the title and summary of the article (item 1), the introduction (items 2 and 3), the methods (items 4 to 12), the results (items 13 to 17) and the discussion sections (items 18 to 17). 21), and other information (point 22 on financing).

PROSPECTIVE PILOT EXPLORATORY STUDIES

Prospective exploratory pilot studies, hereinafter "pilot", share many aspects with RCTs, but they differ from them in two important issues:

- a) They are not subject to supervision, authorization, and possible veto by external commissions, in which the authors of the study do not participate.
- b) There is no systematic record of each of them when it is launched, which is why it is much more difficult to have a record of studies with "negative" results. Therefore, it is practically impossible to avoid publication bias.

The advantages of submitting a research project to the supervision of an external commission are evident. If the external commission that evaluates a research project is made up of truly qualified professionals who can dedicate the necessary time to this evaluation, they would undoubtedly make observations that could improve the project. But here are three points that require detailed reflection:

I) Committees cannot include experts in all fields. Most likely, no member of the committee is an expert on the topic to be evaluated and, if there is one, he or she may not know the topic better than the researchers may. In cases where this commission decides to enlist the help of an expert in the field, that person will know the topic under study, but probably will not know it better than the parents of the project.

II) In general, the people who make up that commission – or the expert consulted – will be able to dedicate many fewer hours to the evaluation than those the authors dedicated to preparing it.

III) It is necessary to evaluate very carefully when it makes sense for the evaluation commission to reject a project. In this aspect, three types of situations must be distinguished: a) when the evaluation committee is a doctoral thesis committee or another academic degree, or has to decide if a project de-

serves financial aid, it is obvious that for each project it has to accept it or not; b) when the committee is the editorial board of a scientific journal, it must also accept it for publication or not; c) in other situations the evaluation commission does not perform these two functions. So, it is questionable whether that commission has the power to veto a project. There are frequent cases in which the objections made by these committees to justify their veto of a project are of a methodological nature (not of a clinical nature) and, upon closer inspection, turn out to be wrong. This error would not have major negative consequences if the rejection were advisory, because the authors, after verifying with an expert that the objection is wrong, would ignore it. However, if the objection implies a veto and appealing that "sentence" is difficult, the most common thing is that the authors of the project give up continuing with it and abandon an initiative that could have provided useful knowledge.

An argument frequently put forward is that these commissions ensure patient safety, defending their rights and – it is assumed – protecting them from possible abuse by researchers. However, there is little reason to think that the members of the external commission will take better care of patients than the doctors who are in personal and direct contact with them. Attending physicians do not ask any committee for authorization every morning to explore and treat each of their patients. For all these reasons, it does not seem appropriate for the evaluation committee to have veto power when this is not an essential part of their mission [9].

REGISTRATION OF MEDICAL STUDIES AND PUBLICATION BIAS

In principle, all RCTs that are launched are registered and their authors undertake to publish the results, since it is assumed – with good judgment – that, if the study is well planned and developed, its result, whatever it may be, provides useful information. However, in practice it is very difficult to get studies published that do not show some "statistically significant" effect. Even the most serious and honest authors when they find poor results tend to give the study as "failed", that is, it was not possible to carry it out properly, instead of "with a negative result" and assume that it is not necessary to publish their work. Thus, we have the problem of multitesting, which can seriously pervert the final balance of publications towards the desired effect. Bonferroni's reasoning helps us quantify the probabilities of finding results with a test P value lower than a certain limit, if several attempts are made and only those that yield a sufficiently small P value are considered worthy of publication.

Let us consider the case in which the pathogen "A", causing a serious infection with very high mortality, was sensitive to a certain antibiotic, but has recently developed resistance, resulting in a strain resistant to all currently known antibiotics. A study aimed at verifying whether it is sensitive to the new drug "B", that has given very promising "provisional" results, creates high expectations, which lead a large number of teams

to carry out studies that can confirm it. A pertinent design would be to randomize 100 patients with this infection into two groups of 50, one with placebo and the other with "B". If the antibiotic is not effective, the study will most likely return a result with a relatively large P value, say greater than 0.05. But if several studies of the same type are carried out with ineffective drugs, the probability that at least one of them will yield a P value < 0.05 increases significantly with the number of studies carried out. Specifically, if N studies are performed, we have: probability (at least one study with P < 0.05) = 1 - (1 - 0.05)^N.

Thus, if 10 studies are carried out with an ineffective antibiotic, the probability that at least one will show P < 0.05 is 1 - (1 - 0.05)¹⁰ = 1 - 0,95¹⁰ = 1 - 0,60 = 0,40. And if 40 studies are carried out, the probability that at least one shows P < 0.05 is 1 - 0.95⁴⁰ = 0.83. There are scenarios in which there is a lot of pressure to research the treatment of an infection, the pandemic is possibly the best example. It is therefore plausible that the number of teams that decide to do this research reaches 100 and in this case the probability that at least one returns P < 0.05 is 0.994, that is, with a useless drug, it is practically certain that it will appear at least one with P value < 0.05. In fact, one study with P < 0.01 and others around 4 studies with P < 0.05 are very likely. If all other studies (with P > 0.05) are not published, the scientific community would consider that there is strong evidence in favor of the usefulness of that drug. To quantify this evidence, a meta-analysis could be carried out with the five published ones and a P value < 0.001 would be obtained.

A possible resource to partially alleviate the problem of multitesting would be to ask for a smaller test P value, for example 0.01, to consider that there is sufficient evidence in favor of the new drug being effective. However, the accumulation of studies has an effect that is only partially counteracted by this lowering of the value that is set as a threshold. Applying the previous formula, we find that the probability that at least one study yields P < 0.01 is 0.10 for N = 10; 0.33 for N = 40 and 0.63 for N = 100. We see that the probability of finding studies with "significant" results when there is not really the desired effect is high and if the medical scientific community does not have access to studies with negative results that have appeared simultaneously with the positive ones, is committed to considering good products and procedures that are useless. In some cases, these treatments can even have a harmful effect, as we have seen in the case of hydroxychloroquine used for COVID-19 infection [10].

The problem of publication bias arises in the most general case, where there is no promising previous data regarding a specific product. It is enough that there is – as is usual in medicine – a problem to which a solution is sought and many research groups are mobilized for it. If each group tests a different product or procedure, the probability that one of them that is truly useless will be declared useful also increases according to the Bonferroni formula and therefore the need to faithfully report all clinical studies that are launched. It is imperative to reduce publication bias.

RCTs have this component – the systematic recording of all studies initiated – that is very positive, but it is obvious that the same praxis can be applied to all types of clinical studies. The entire administrative infrastructure that enables the registration of RCTs can and should accommodate the registration of prospective, pilot and exploratory clinical studies. The logistics involved are complicated and expensive, but the medical science community has no choice. If there are no rigorous systematic prior records that allow us to know at least a good part of the studies with negative results, the publication bias is very high. Only with these records and the publication of negative results is it possible to carry out reliable meta-analyses.

In conclusion, RCTs are a very good source of clinical information, but not the only one. The systematic registration of all research initiated can and should be applied to all types of clinical studies.

FUNDING

None to declare

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Yeh RW, Valsdottir LR, Yeh MW, Shen C, Kramer DB, Strom JB, et al; PARACHUTE Investigators. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ*. 2018;363:k5094. doi: 10.1136/bmj.k5094. Erratum in: *BMJ*. 2018 Dec 18;363:k5343.
2. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Int J Prosthodont*. 2006;19(2):126-8.
3. Gilmartin-Thomas JF, Liew D, Hopper I. Observational studies and their utility for practice. *Aust Prescr*. 2018;41(3):82-85. doi: 10.18773/austprescr.2018.017.
4. Diamond GA. Randomized trials, observational registries, and the foundations of evidence-based medicine. *Am J Cardiol*. 2014;113(8):1436-41. doi:10.1016/j.amjcard.2014.01.420.
5. Ricotta EE, Rid A, Cohen IG, Evans NG. Observational studies must be reformed before the next pandemic. *Nat Med*. 2023;29(8):1903-1905. doi: 10.1038/s41591-023-02375-8.
6. Ligthelm RJ, Borzi V, Gumprecht J, Kawamori R, Wenying Y, Valensi P. Importance of observational studies in clinical practice. *Clin Ther*. 2007;29 Spec No:1284-92.
7. Mainzer R, Moreno-Betancur M, Nguyen C, Simpson J, Carlin J, Lee K. Handling of missing data with multiple imputation in observational studies that address causal questions: protocol for a scoping review. *BMJ Open*. 2023;13(2):e065576. doi:10.1136/bmjopen-2022-065576.
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenberghe JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-7. doi: 10.1016/S0140-6736(07)61602-X. PMID: 18064739.
9. Carazo Díaz C, Prieto L, Martínez-Sellés M. Errores y malentendidos que entorpecen la investigación médica y dificultan el avance de la medicina. *REC Interv Cardiol*. 2024 (in press). DOI: <https://doi.org/10.24875/RECIC.M23000436>
10. Pradelle A, Mainbourg S, Provencher S, Massy E, Grenet G, Lega JC. Deaths induced by compassionate use of hydroxychloroquine during the first COVID-19 wave: an estimate. *Biomed Pharmacother*. 2024;171:116055. doi: 10.1016/j.biopha.2023.116055.