

Planning Clinical Trials: The Main Step in Developing New Drugs

Michel Bourin*

Neurobiology of mood disorders, University of Nantes, France

Received: October 11, 2017; Accepted: November 30, 2017; Published: December 04, 2017

*Corresponding author: Michel Bourin, Neurobiology of mood disorders University of Nantes, 98, rue Joseph Blanchart, 44100, Nantes, France, E-mail: michel.bourin@univ-nantes.fr

Abstract

A therapeutic trial is generally conducted to determine a new treatment or improve therapeutic applications. Several crucial questions should be considered before the trial is undertaken. The planning of a clinical trial involves a succession of steps intended to develop a protocol providing a clear and unbiased answer to a precise therapeutic question. What treatments are to be compared? What type of population should be studied? What are the evaluation criteria? What is the most suitable experimental scheme? How many patients should be included? What statistical procedures should be implemented (single analysis, intermediary analysis, serial analysis)? This paper considers briefly the main points to be resolved at each planning stage.

Keywords: Clinical Trials; Experimental Plan; Evaluation Criteria; Statistical Procedures

Introduction

Planning a clinical trial supposes two prerequisites: the formulation of a clear, precise and relevant therapeutic question and the certainty that the reply to this question has not already been given in earlier studies.

First of all, this step requires careful reflection about the wording of the question. Experience has shown that an initial therapeutic question often includes a series of questions that must be analyzed to determine the most important one. The objective at this stage is to obtain a sufficiently precise question, i.e. which clearly defines the pathology and the type of patients studied, the therapeutic modalities envisaged, the modes of comparison, and the judgment criterion on which the evaluation will be based.

Secondly, exhaustive bibliographic study is needed to be sure that the question has not already been resolved. If the question is truly pertinent, it is likely that other investigators have already considered it and attempted to find an answer. If the literature does not indicate that the question has been resolved, it is desirable at this stage, in the light of published works on the same subject, to reconsider its pertinence. If it still seems of interest, then the feasibility of the study must be ensured. A maximum of data must be obtained to understand the most recent developments in the subject, the methodological choices, the problems encountered,

etc.

It is important at this stage to consider that a clinical trial can answer only one question precisely. As the number of therapeutic trials that can be performed is limited, they should relate to the most important questions to be resolved. The protocol must define the means needed to reply to the question, and particularly the number of subjects to be included in the study. The carrying out of the trial implies the means to recruit the number of subjects needed. Otherwise, it is preferable not to attempt it [1].

The Experimental Plan

Three simple principles govern the procedure ensuring an unbiased reply to the question:

The Need for Comparison

When a patient receives a treatment, the effect observed after a given period of time depends on the natural course of the disease, the fact of being treated (placebo effect) and the pharmacologic activity of the drug. The spontaneous variability of the course of most diseases and the often relative results of therapy mean that the effect of any new treatment should be evaluated by comparing a group receiving treatment with one that is not. Moreover, it is necessary to use a control group receiving a placebo to evaluate the contribution of the pharmacologic activity of the treatment to the therapeutic effect observed.

The Need for Initial Comparability of Groups

If an observed difference is to be attributed to the therapy studied, the treatment and control groups must be comparable before the beginning of the trial. Otherwise, it would be impossible to distinguish between the effect of the treatment and what is due to the initial difference between groups. Several procedures have been proposed to determine the treatment to be attributed to a patient. All involve the constitution of groups that are more or less comparable. The only means of obtaining truly comparable groups relative to all the individual characteristics of the patients is to draw lots (randomization). In practice, this means that each patient receives the treatment assigned by a random procedure. This method of composing the groups has the added advantage of ensuring the validity of the statistical tests used to compare the groups [2].

The Need to Maintain the Comparability Resulting From Randomization throughout the Study

If an observed difference is to be attributed to the therapy studied, the treatment and control groups must remain comparable throughout the trial. Otherwise, it would be impossible to distinguish between the effect of the treatment and what is due to external factors. The awareness of the physician and/or the patient as to the type of treatment administered would be likely to introduce a bias. A physician who knows that he is giving a patient a placebo or a supposedly active treatment may have a very neutral or highly enthusiastic attitude that could influence the response of the patient negatively or positively. Likewise, the response of a patient may be influenced negatively or positively if he knows that he is receiving a placebo or a supposedly active treatment. In this context, the only means of maintaining the comparability resulting from randomization throughout the study is to perform a double-blind trial. This procedure requires that the different treatments be totally indiscernible and that the modalities of administration be absolutely identical. When a double-blind trial is not possible, a single-blind procedure can be considered in which only the physician is aware of the treatment administered. In this case, and even when the single-blind approach is impossible, the judgment criterion should be evaluated by a physician unaware of the treatment the patient is receiving. This blind procedure avoids a possible bias in evaluating the effect of the treatment administered.

These three simple principles (comparison, randomization and double-blind procedure) can be applied systematically in most therapeutic trials. Though the first two principles are intangible, the double-blind procedure must sometimes be abandoned because it raises ethical or technical problems or involves high additional costs. The bias potentially induced by the physician's and/or patient's awareness of the treatment should be estimated according to the question posed, the judgment criterion chosen, etc.

The second phase is to define the experimental scheme used to compare treatments, which can be either a trial with parallel groups or a crossover trial [3]. The former consists in assigning each patient only one of the two treatments (as determined by randomization), so that the two resulting groups are composed of different patients. The latter depends on each patient serving as his own control, i.e. each patient receives both treatments successively. In this case, randomization determines the order in which the treatments are received rather than assignment to a specific treatment group. Both groups are composed of the same patients, thereby reducing the number of subjects to be included. The crossover trial supposes that the treatments do not allow the disease to be cured definitively (for instance, treatment of symptoms) and are rapidly effective and of short duration. These conditions suppose that the basal disease state at the beginning of the second sequence is comparable to that at the beginning of the first sequence and that the judgment criterion may be readily obtained. Finally, for purposes of simple and unequivocal interpretation of the results of the comparison, there must be no interaction between the treatments studied and the two different

periods, i.e. that the difference between the two treatments is the same whether they be administered in the first or second period [4].

As a result of all these constraints, the crossover scheme is rarely used in phase III trials, but rather in phase I or II [5]. However, it is possible to consider it for phase III in certain specialties such as ophthalmology, when the pathology concerned involves both eyes, and dermatology, when several lesions exist simultaneously, provided that the treatments studied involve no systemic route.

Experimental Schemes Allowing Comparison of More Than Two Treatments

Among the experimental schemes allowing comparison of more than two treatments, the 2 x 2 factorial approach is of particular interest since it provides evaluation of the distinctive effects of two treatments (versus placebo) and of the possible interaction of these two treatments (i.e. that the overall effect of the two treatments is greater or lesser than that obtained simply by adding up the particular contribution of each treatment). In the absence of interaction between the treatments, this type of experimental scheme allows an evaluation (versus placebo) of the distinctive effects of two treatments using the same number of subjects as would be necessary for the comparison of just one treatment with a placebo [6].

The 2 x 2 factorial approaches define 4 modalities:

- One group receives neither A nor B (placebo group),
- One group receives only treatment A,
- One group receives only treatment B,
- and one group receives both treatments A and B

At the end of this study, and in the absence of interaction between treatments A and B, the distinctive effect of A is evaluated by comparing the patients who received A (those of groups A and A+B) with those who did not receive A (those of the placebo and B groups). The distinctive effect of B is evaluated by comparing the patients who received B (those of groups B and A+B) with those who did not receive B (those of the placebo and A groups). Interaction is evaluated by comparing the patients of the placebo and A+B groups with those of groups A and B.

Choice of the Population

The population included in the trial should be representative of that intended to benefit from the new therapy but uniform enough to demonstrate its efficacy. These two requirements are apparently contradictory since representativeness implies a broad range of differences in the population included [7]. In fact, the selection criteria for patients must be defined to provide the best compromise.

Inclusion criteria define which patients are ideal for the trial, i.e. those who would best demonstrate the effects of the treatment studied. They include demographic criteria (sex, age, weight, etc.), nosologic criteria defining the diagnosis of the disease, criteria indicative of the severity and progressive nature

of the disease, criteria defining the conditions for informing and observing patients, and possibly criteria concerning treatments previously administered [8].

Exclusion criteria define patients who correspond to the inclusion criteria but whose presence in the trial would make it more difficult to demonstrate the effects of the treatment studied. These patients may show poor understanding of the protocol or poor compliance with treatment or present a difficulty for evaluation of the judgment criterion. They may run a greater risk in undergoing the treatment because of renal, hepatic or respiratory insufficiency, heart failure, pregnancy, breast-feeding, etc. They may have a particularly severe clinical form of the pathology studied, or associated pathologies that introduce particular risks. There may be contraindications to the explorations in the protocol, or to one of the treatments studied. Certain therapeutic indications may be forbidden, or particular dangers may be involved relative to the treatments studied.

The following points are important to recall at this stage:

1. The selection criteria should be precise and unequivocal, so that each investigator can determine whether a given patient is capable of being included in the trial.
2. The selection criteria should define patients who are representative of the eventual treatment population to which the results of the trial can be extrapolated.
3. The results of the trial can in fact be extrapolated to the population from which the participating patients are derived. When the selection criteria are determined, the relative frequencies of given patients presenting particular characteristics should be considered.
4. Less restrictive selection criteria allow simpler and more rapid recruitment and easier extrapolation of results but can also increase the number of subjects required and make it difficult to demonstrate efficacy when the treatment is not the same for different disease classes.
5. More restrictive selection criteria make recruitment more difficult and slower and limit the scope of the results but can also decrease the number of subjects required and make it easier to demonstrate the efficacy of the treatment.

Treatments

The treatments considered are those studied in the trial as well as previous and associated ones.

Treatments Studied In the Trial

Treatment to be evaluated. Phase I and II trials should have defined the galenic form of the drug, the administration route, the dosage and the optimal modes of administration required to obtain the expected therapeutic effect or the best compromise between efficacy and tolerance [9]. The choices made are indicated in the protocol. In addition, it is necessary to specify the length of treatment (if for a set period) or the criteria allowing discontinuance of treatment (if for a variable period). In certain cases, an adaptation of the dosage should be considered either

from the start of treatment, if the dosage is based on the particular characteristics of the patient (age, weight, body surface area, creatinine clearance, etc.), or during treatment, depending on the patient's response as evaluated by precise criteria of efficacy and/or tolerance [10].

Treatment given to the control group. In practice, the control group can be given no treatment or receive a placebo or a reference treatment [11]. The first procedure should only be chosen if the other two are impossible since it does not allow the performance of a double-blind trial. The choice between a placebo and a reference treatment depends on the therapeutic purpose (evaluation of the "absolute" or the "relative" efficacy of a new treatment) [12]. Ethical reasons sometimes require the choice of a reference treatment which should be clearly effective versus placebo and recognized by the scientific community. Moreover, the galenic form, administration route, dosage, modes of administration and length of treatment should be optimal to obtain the expected therapeutic effect or the best compromise between efficacy and tolerance. When a blind procedure (single or double) is planned, the treatments compared should be totally indiscernible as to presentation, form, color and taste and be administered according to the same modalities. When the control group receives a reference treatment, it is often simpler to produce a "placebo" in galenic form for each of the treatments studied and give each patient the treatment assigned by randomization and the placebo of the other treatment (double-placebo technique).

Previous treatment

Previous treatments can be considered in determining inclusion and exclusion criteria and may affect the conduct of the trial by inducing a wash-out period. This period, which can be obtained without treatment, or under placebo, or sometimes by symptomatic treatment, allows evaluation of the state of patients without active treatment and can also be used, on the basis of specific criteria, to select the pre-included patients who will finally be included in the trial [13].

Associated treatments

Associated treatments intended for the treated patient. It is essential to have a precise definition of the associated treatments which may be approved and the circumstances and their administration modes. This is a delicate problem in all cases since it is difficult at the time of evaluation to distinguish between the effect due to the studied drug and that due to the associated treatment [14].

Associated treatments intended to combat an intercurrent affection. It may be necessary to have a precise definition of the associated treatments approved for treatment of an intercurrent affection and of their administration modes.

Randomization

The purpose of randomization is to compose groups comparable for all factors (known and unknown) characteristic of the disease, in particular those which could influence response [15].

Two types of randomization can be considered: fixed randomization, which gives each included patient the same likelihood of receiving one of the treatments, regardless of the time of inclusion in the trial; and adaptive randomization, which gives each included patient the likelihood of receiving one of the treatments, depending on the time of inclusion [16]. The latter allows the correction of a possible imbalance in the number of patients or in the initial disease characteristics compared. Only fixed randomization will be considered here.

The Simplest Form of Randomization

This approach consists in determining the treatment to be assigned to the patient at each inclusion by the toss of a coin. If the coin is correctly weighted, the repetition of a great number of tosses inevitably leads to the composition of two groups of very similar size. To avoid the toss of a coin, random number tables can be used in which 0 to 9 appear at random regardless of how the tables are read (by lines or columns). One treatment can then be assigned to certain numbers (e.g. even ones) and the other treatment to other numbers (e.g. odd ones). The amount of numbers used depends on the number of patients to be included [17]. Moreover, this approach makes it easy to create randomization lists to obtain a different ratio between treatments (e.g. to include twice as many patients in treatment A as in treatment B, numbers 1 to 6 are attributed to A, 7 to 9 to B, and 0 is excluded). More than two treatments can also be distributed in this manner. For example, if an equal number of patients is to be included in 3 treatments A, B and C, A can be attributed to 1-2-3, B to 4-5-6, C to 7-8-9, and 0 is excluded. The major drawback in this randomization method is that it can cause imbalances between the groups, especially when the number of patients is low [18].

Block randomization

This approach involves the use of random number tables in which series of numbers appear (e.g. 1 to 6) according to the sequences defined (by lines or columns). The order of these numbers within a sequence is completely random. The rules for assigning treatments from successive numbers are the same as above. The advantage of this method over the preceding one is the creation of balanced groups at the end of each sequence. For example, if a permutation table with 6 elements is used and A is to be assigned for numbers 1-2-3 and B for numbers 4-5-6, it is certain that 3 out of 6 successively randomized patients will receive each of the treatments. In the worst of cases, there would be an imbalance of 3 patients between the two treatments at the end of the study. To make the balance perfect between the treatments studied, it is only necessary to choose a rule that provides an exact division of the number of subjects to be included. However, this method used in this fashion does not ensure a perfect balance of patients between different centers (in the case of a multicenter study) and of the essential prognostic factors of the disease between the treatment groups [19].

Stratified randomization

This approach is most often a block randomization (though the method can also be applied to the simple form) in which the randomization list is by level of the stratification variable. For example, in a multicenter study, the center is always a stratification variable. Or a prognostic factor which supposedly has an important influence on response can be a stratification variable. The advantage of stratification applied to block randomization is to create balanced groups for each level of the stratification variable [20].

Finally, it is possible to stratify the randomization with respect to several variables (e.g. the center and 1 or 2 essential prognostic factors), although it is apparent that the number of resulting randomization lists depends on the number of levels of each variable. For example, the desire to stratify randomization on the center, in the case of 4 centers and 2 prognostic factors with respectively 2 and 3 different levels, would lead to the creation of 6 distinct randomization lists per center for a total of 24. Besides the great complexity involved in the practical achievement of this type of randomization (a complexity which may also cause mistakes), certain combinations of stratification factors may introduce difficulties in recruitment possibilities. In practice, the best compromise should be found between the benefit of creating balanced groups for one or two essential prognostic factors and the feasibility of the procedure [21].

Evaluation Criterion

The choice of the evaluation criterion is crucial in a therapeutic trial since it allows an unambiguous reply to the question posed and determines the number of subjects required, thereby partially affecting the feasibility of the study [22].

Conditions determining the choice of an evaluation criterion

The criterion must be medically pertinent. For example, the choice of measuring transcutaneous oxygen pressure after exercise to evaluate the supposed efficacy of a treatment in arteritis of the lower limbs is certainly less pertinent than measuring the walking perimeter [23]. The criterion must also be currently accepted by the international scientific community. A therapeutic trial is intended to convince all physicians of the validity and benefit of the results, which is much easier if the methodology is irrefragable and the criterion used is widely accepted for the pathology studied. The criterion must be measured in the same manner regardless of the center and the investigator but also, and especially, of the treatment administered. It must be sensitive (allowing detection of slight improvements or worsening) and specific (varying only through the effect of the treatment).

Moreover, the measuring instrument used to evaluate the criterion should possess metrologic qualities such as sensitivity (the aptitude to reveal minimal variations induced

by the treatment), specificity (the aptitude to measure only the evaluation criterion), repeatability (the aptitude to give the same result if the measurement is repeated by the same observer in the same conditions within a short period of time), and reproducibility (the aptitude to give the same result if the measurement is repeated by two different observers in the same conditions at two different times) [24].

Intermediary criteria

In certain cases, the most pertinent judgment criterion cannot be evaluated until after a considerable delay, e.g. the efficacy of an antihypertensive treatment of cardiovascular mortality. It is then possible to use other types of evaluation criteria known as intermediary criteria, provided that there is a demonstrated relation between the intermediary criterion and the most pertinent evaluation criterion and that therapeutic action on the intermediary criterion leads to a demonstrated therapeutic effect on the pertinent criterion. In the example just cited, blood pressure could serve as an intermediary criterion [25].

Different types of evaluation criteria

Objective or subjective criterion?

The choice between an objective criterion (which, like blood pressure, can be determined directly with a measuring device) and a subjective criterion (which, like pain or the moment for resorting to a final treatment, can be left to the discretion of the patient or physician) must be based above all on the answer to the question posed [26]. If the choice is possible and the two criteria are of equal pertinence, preference will probably be given to the objective criterion which may be assumed to have a lesser spontaneous variability [27].

Qualitative, quantitative or censured criterion?

The choice between a qualitative criterion (which evaluates the quality of the patient's response, e.g. success or failure of the treatment), a quantitative criterion (which evaluates the patient's response on a graded scale, e.g. blood pressure) and a censured criterion (which evaluates the occurrence or not of an event and the delay in its occurrence, e.g. death) must be based above all on the answer to the question posed. If the choice is possible and two or three criteria show the same pertinence (which is unlikely), preference will probably be given to the evaluation criterion for which the number of required subjects is lowest [28].

Main or secondary criterion?

The answer to the question posed requires a definition of the preferential judgment criterion which will allow determination of the number of subjects to be included in the trial. This preferential criterion is the main one in the trial. As a therapeutic trial is often a complicated and prolonged procedure, it is desirable to evaluate treatment response relative to other criteria than the main one. These secondary criteria must be clearly identified as such, and their number limited (less than 10 to 15) to avoid poor evaluation and, more importantly, any deterioration in the assessment of the main criterion. Their choice depends on the same rules as those for the main criterion. The possibility of detecting differences

in these criteria is not controlled a priori since the number of subjects required depends on the main criterion alone [29].

The moment for measuring the criterion

The evaluation criterion can be measured at the end of the treatment period, or before and after, or before and then repeatedly during the treatment period. When the criterion is measured before and at least once after the beginning of treatment, the variation calculated is either absolute (the difference in parameter values before and after treatment) or relative (the percentage of variation between the values before and after treatment)[30]. Once again, the choice depends on the question posed. However, it is always beneficial to remain as close as possible to the measured value, particularly in order to convince the scientific community of the pertinence of the results observed. Repeated measurements should only be considered in phase III trials when there is good evidence that the effect and therefore the difference between the treatments is likely to change with time, even though a state of pharmacokinetic equilibrium has been reached.

The Number of Subjects Required

Calculating the number of subjects required is a determinant step after the choice of the main judgment criterion since it has a partial influence on the feasibility of the trial [31]. This simple step involves establishing the value of four parameters: the minimum benefit of clinical interest relative to the main judgment criterion, or the benefit that has the best chance of being detected if it exists; the variability of the judgment criterion s in the selected population; the first species of risk α of the statistical test, i.e. the risk of finding the study treatment more effective than the control treatment, whereas efficacy is actually the same; and the second species of risk β of the statistical test, i.e. the risk of finding the study treatment less effective than the control treatment, whereas efficacy is the same or better.

The value of α must be determined realistically. As a result of an obvious overestimation, the trial would very likely show no difference between the treatments studied, whereas an underestimation would produce too slight a benefit to prove of clinical interest and convince the scientific community. Moreover, this second misestimating would lead to an unnecessary increase in the number of subjects included in the study [32].

The value of β must be carefully determined on the basis of values in the literature for patients similar to those in the protocol and for a measurement device identical to that used in the protocol. Precise knowledge of the variability of the judgment criterion is all the more likely if a currently used classical criterion is chosen. The estimation of s determined from the literature must be compared with estimations provided by the investigators drafting the protocol on the basis of experience with their own patients.

The value of α and β risks is defined by the statistician participating in the drafting of the protocol. In practice, α is 5% and β can range between 5 and 20%. Accepted a β risk above 5% allows the number of subjects in the trial to be reduced, although

this is likely to decrease the probability of demonstrating the efficacy of the treatment. When $\beta = 20\%$, there is one chance in 5 of not detecting a difference in efficacy between the treatments equal to.

Statistical Analysis

The use of statistical analysis must be considered from the planning phase in order to avoid any bias that could be introduced by the choices made in terms of the results of the study [33].

The classical approach for a therapeutic trial is to perform a single statistical analysis for all randomized patients once the number of expected subjects has been reached. The statistical tests used depend on the nature of the judgment criterion: the chi-square test for a qualitative criterion (comparison of percentages), Student's t-test for a quantitative criterion (comparison of means), and the Log rank test for a censored criterion (comparison of survival distributions). If the statistical test, when performed in these conditions, shows a significant difference between the treatment groups relative to the main judgment criterion, and if the trial has been carried out according to the methodological principles discussed in section 2 (randomized, comparative double-blind trial), the difference observed can be attributed to the treatment.

In practice, statistical analysis of a therapeutic trial is often much more complicated than the classical approach would suggest because of deviations from the protocol, the possibility of imbalances in certain prognostic factors between the treatment groups which require adjustment strategies, and possible interactions between treatment results and certain co variables (e.g. time). Moreover, the possibility of analyzing the data during the trial in order to discontinue it if a difference or a lack of difference in efficacy is clearly apparent leads to the implementation of sequential-type statistical methods which have their own limitations.

The strategy for statistical analysis of a therapeutic trial is almost necessarily complex, which implies the participation from the planning stage of a statistician experienced in the methodology of clinical research.

Conclusion

The planning of a therapeutic trial involves a succession of steps from the formulation of the question to the choice of statistical analysis to be performed once the trial is finished. This series of steps, which leads to the elaboration of a protocol, is often long and requires the participation of various individuals including specialists in the pathology considered, clinicians, biologists, a pharmacologist and a biostatistician. It is important to take the time needed to develop a good protocol since the improvised resolution of problems occurring during the trial by the investigators themselves is likely to bias the results of the study.

References

- Amato A, Aringhieri E, Boccia S, Buccella F, Gorini B, Gramaglia D, et al. A harmonized and efficient clinical research environment would benefit patients and enhance European competitiveness. Commentary. *Ann Ist Super Sanita*. 2017;53(2):104-107. doi: 10.4415/ANN_17_02_05
- McDougall J, Tsonis M. Quality of life in survivors of childhood cancer: a systematic review of the literature (2001-2008). *Support Care Cancer*. 2009;17(10):1231-1246. doi: 10.1007/s00520-009-0660-0
- Curtin F, Heritier S. The role of adaptive trial designs in drug development. *Expert Rev Clin Pharmacol*. 2017;10:727-736. doi.org/10.1080/17512433.2017.1321985
- Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Stat Med*. 2016;35:325-347. DOI: 10.1002/sim.6472
- Bourin M, Chagraoui A. Prerequisites for phase I and II clinical drug trials in human. *SOJ Pharm Pharm Sc*. 2016; 3:1-3.
- Elsässer A, Regnstrom J, Vetter T, Koenig F, Hemmings RJ, Greco M, et al. Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. *Trials*. 2014;15:383. doi: 10.1186/1745-6215-15-383
- Bernard A, Vaneau M, Fournel I, Galmiche H, Nony P, Dubernard JM. Methodological choices for the clinical development of medical devices. *Med Devices (Auckl)*. 2014;7:325-34. doi: 10.2147/MDER.S63869
- Bourin M. Criteria of patient selection in a clinical trial. *SOJ Pharm Pharm Sc*. 2017; 4(3):1-4. doi.org/10.15226/2374-6866/4/3/00164
- Bourin M, Chagraoui A. Methodological issues of Phase II Clinical Trials in Psychopharmacology. *KJ Pharmacol*. 2016;1:100107.
- Travessa AM, Rodrigues FB, Mestre TA, Ferreira JJ. Fifteen Years of Clinical Trials in Huntington's Disease: A Very Low Clinical Drug Development Success Rate. *J Huntingtons Dis*. 2017;6(2):157-163. doi: 10.3233/JHD-170245
- Khan O, Ferriter M, Huband N, Powney MJ, Dennis JA, Duggan C. Pharmacological interventions for those who have sexually offended or are at risk of offending. *Cochrane Database Syst Rev*. 2015;(2):CD007989. doi: 10.1002/14651858.CD007989
- Wang RS, Hall KT, Giulianini F, Passow D, Kaptchuk TJ, Loscalzo J. Network analysis of the genomic basis of the placebo effect. *JCI Insight*. 2017;2(11):93911. doi: 10.1172/jci.insight.93911
- Bourin MS, Severus E, Schronen JP, Gass P, Szamosi J, Eriksson H, et al. Lithium as add-on to quetiapine XR in adult patients with acute mania: a 6-week, multicenter, double-blind, randomized, placebo-controlled study. *Int J Bipolar Disord*. 2014;2:14. doi: 10.1186/s40345-014-0014-19
- Yatham LN, Vieta E, Goodwin GM, Bourin M, de Bodinat C, Laredo J, et al. Agomelatine Study Group Agomelatine or placebo as adjunctive therapy to a mood stabiliser in bipolar I depression: randomised double-blind placebo-controlled trial. *Br J Psychiatry*. 2016;20:78-86. doi: 10.1192/bjp.bp.114.147587
- Flecha OD, Douglas de Oliveira DW, Marques LS, Gonçalves PF. A commentary on randomized clinical trials: How to produce them with a good level of evidence. *Perspect Clin Res*. 2016;7:75-80. doi: 10.4103/2229-3485.179432

16. Dalton BG, Gonzalez KW, Millsbaugh DL, Desai AA, Sharp SW, St Peter SD. Pectus excavatum: Benefit of randomization. *J Pediatr Surg.* 2015;50:1937-1939. doi: 10.1016/j.jpedsurg.2015.05.009
17. Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials.* 1988;9:312-26. doi.org/10.1016/0197-2456(88)90046-3
18. Okeh UM, Ugwu AC. Randomized controlled trials: the most powerful tool in modern clinical research. *East Afr J Public Health.* 2009;6(1):8-10.
19. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials.* 1988;9(4):327-344.
20. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. *Trials.* 2015;16:405. doi: 10.1186/s13063-015-0920-x
21. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol.* 1999;52:19-26.
22. Bourin M, Garay RP. Choice of evaluation criteria in a clinical trial? *SOJ Pharm Pharm Sci.* 2016;3(3):1-4.
23. Orloff S. Some new criteria for the evaluation of disease activity in rheumatoid arthritis. *Acta Rheumatol Belg.* 1977;1:263-278.
24. Gruen ME, Griffith EH, Thomson AE, Simpson W, Lascelles BD. Criterion Validation Testing of Clinical Metrology Instruments for Measuring Degenerative Joint Disease Associated Mobility Impairment in Cats. *PLoS One.* 2015;10(7):e0131839. doi: 10.1371/journal.pone.0131839
25. Redon J. Antihypertensive treatment: should it be titrated to blood pressure reduction or to target organ damage regression? *Curr Opin Nephrol Hypertens.* 2005;14:448-452.
26. Parving A, Ostri B. On objective criteria for hearing impairment and hearing disability. *Scand Audiol.* 1983;12:165-169.
27. Overall JE, Beller SA. The Brief Psychiatric Rating Scale (BPRS) in geropsychiatric research: I. Factor structure on an inpatient unit. *J Gerontol.* 1984;39:187-193.
28. Ratner J, Mullins D, Buesching DP, Cantrell RA. Pragmatic clinical trials: U.S. payers' views on their value. *Am J Manag Care.* 2013;19(5):e158-165.
29. Kisely SR, Campbell LA. Compulsory community and involuntary outpatient treatment for people with severe mental disorders. *Schizophr Bull.* 2015;41:542-543. doi: 10.1002/14651858.CD004408
30. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med.* 1989;19:61-70. doi.org/10.1016/0010-4825(89)90036-X
31. Heo M, Xue X, Kim MY. Sample size requirements to detect a two- or three-way interaction in longitudinal cluster randomized clinical trials with second-level randomization. *Clin Trials.* 2014;11:503-507. DOI: 10.1177/1740774514532724
32. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med.* 2011;30:3267-3284. DOI: 10.1002/sim.4102
33. Thompson PA, Wright DE, Counsell CE, Zajicek J. Statistical analysis, trial design and duration in Alzheimer's disease clinical trials: a review. *Int Psychogeriatr.* 2012;24:689-697. DOI: 10.1017/S1041610211001116