

# Methodology for Small Clinical Trials

Muriel Vray,<sup>1</sup> Danièle Girault,<sup>2</sup> Natalie Hoog-Labouret,<sup>3</sup> Raphaël Porcher,<sup>4</sup> Jean-Christophe Thalabard<sup>5</sup> and the Participants in Round Table No. 1, Giens XIX<sup>†</sup>

1 Institut Pasteur, Unité Epidémiologie Pathologies Emergentes, Paris, France

2 Laboratoires Takeda, Puteaux, France

3 Afssaps, Saint-Denis, France

4 Hôpital Saint-Louis, Paris, France

5 Hôpital Necker, Paris, France

## Abstract

Small clinical trials are trials in which the number of patients does not enable the objective of the study to be appropriately met with the usual methodological rules. This situation is common in the case of rare diseases, in paediatrics, in certain cancer pathologies or when the number of patients exposed to the treatment needs to be limited. The principal methodological problems are initially identified, and the classical methods (controlled, randomised, double-blind trial using parallel groups, crossover trial, factorial design, trial performed with several measures repeated over time, add-on design, randomised withdrawal design or early-escape design) and more uncommon methods (sequential approaches, meta-analyses, the 'N of 1' method and other methods that facilitate decision making or modelling) are then discussed. Subsequently, recommendations are made to ensure that the results obtained are not a matter of chance, and to increase the level of proof.

**Keywords:** small trials, statistical methods, level of proof

Small clinical trials are trials conducted and analysed using patient numbers lower than is required to meet the aims of the study according to the more common methodological rules. This paper therefore does not concern the standard phase I and II trials where the number of patients included in the study, although quantitatively small, is methodologically adapted to the aims of the study. It concerns all trials that aim to establish the proof of whether any given effect, be it of efficacy or tolerance, is present.

This situation is often the case with rare diseases, in paediatrics or in cancer treatment, when the external limitations of patient availability do not enable the recruitment of patient numbers that meet the usual assessment criteria. It is also the case when the number of patients exposed to the investigational treatment methods needs to be limited, either because of constraints regarding the particular protection of individuals (children, pregnant women), or because of the toxicity of the product (cancer treatment), or even its lack of efficacy (placebo, weak doses). The methodological issue in 'proof of concept' studies, the results of which lead to the decision on whether or not to continue developing a molecule, although based on a different context, can be related to this field.

The main issue surrounding these small trials is the level of proof provided by the results of these studies, which reflects the relevance of using non-standard methods.

Conducting a small trial through necessity reveals various levels of difficulties, some of which are not entirely specific to small trials, but are increased because of the reduced patient numbers. These include the following:

1. The risk of not reaching a conclusion, even though there is a real difference between the therapeutic regimens being compared, because of the great variability in random fluctuations. Only a quantitatively important effect is likely to be given prominence.

2. The caution necessary when extrapolating the observed results to the whole population likely to be receiving the treatment, since the conditions required to achieve these results should not deviate too much from the circumstances of usual practice.

3. The near impossibility of considering covariables when analysing the response to treatment, thereby making it difficult to specifically identify responder patients.

4. The occasional absence of an explicit biological mecha-

<sup>†</sup> For a list of participants, please see the end of the article.

nism allowing reliance on an objective criteria or biological marker to support the results, since the reduced number of patients limits the level of proof provided by the clinical evaluation.

5. The particular circumstance of conducting such trials specifically because of the lack of an alternative treatment, the special relationship between patients or patient associations and therapists, the absence of any prior data on the variability in the proposed criteria and the subsequent risks – such are the many sources of bias in objectively evaluating the treatment.

The sum of these uncertainties and constraints increases the particular needs for precision and transparency in these trials, which have to be the same as, if not greater than, those in standard trials, when ensuring that the results are not simply due to chance.

In the present article, the proposal is to first touch on approaches derived from classical methodology, but with particular relevance to the context of small clinical trials and, later on, to address the more unusual approaches.<sup>[1]</sup>

## 1. Classical Methods of Evaluation

Regardless of the methodology used, a small trial cannot contravene general principles, nor can it be undertaken unless the three following conditions exist: a relevant and significant medical issue; a procedure that adheres to a rigorous methodology that can answer the question raised; and fulfilment of the rules and ethical principles.

The reference model is the controlled, randomised double-blind trial using two parallel groups. This approach should be systematically favoured, and should also give preference to an international multicentre trial if a local or national recruitment plan cannot enrol the patient numbers usually required. The use of protocols and online notebooks may facilitate these multicentre trials, as advocated by, for example, an organisation such as GEREQ.<sup>[2]</sup> The reduced number of subjects requires a reduced variability in the response, which may be provided by the classical method of variance reduction.

### 1.1 Crossover Trials

In crossover trials, the subject is his own control and randomly receives two or several sequences of treatment. When comparing two treatments, A and B, each subject receives either sequence A then B, or the reverse, B then A, with the sequence given by balanced random allocation. The advantages of this are as follows: (i) for the clinician, a comparison of effects in the same patient, therefore eliminating intersubject variability, and (ii) for the methodologist, the increase in power is especially important since the intrasubject variability (variability of criteria measured in the same patient) is low compared with the inter-

subject variability (variability of criteria measured between subjects), and since the difference in the expected effect between the two treatment groups is low compared with the intersubject variability. Nevertheless, this experimental design is problematical in that it concerns stable chronic diseases; it assumes that each subject enters the second phase of treatment in a state comparable to that when starting the first phase, thereby eliminating *a priori* any patients who have recovered, died or have been lost to follow-up during the study; the main endpoint of the intervention should not be modified by learning processes (for example, coronary artery disease and the treadmill) or habituation (diet regimen) phenomena. Furthermore, the effect of the treatment received during the first phase should be exhausted when the subject starts the second phase, thereby imposing a fairly lengthy period known as the ‘washout’ in the case of any possible remaining effect. The total duration of the treatment for each patient is therefore twice as long as in a parallel-group study, culminating in an increased number of patients lost to follow-up, even though the final analysis concerns only the subjects with a measured criterion for each of the two treatment phases. This therefore sets out criteria that change rapidly during treatment in order to limit the total duration of the study. Finally, the usual analysis requires the elimination of an interaction effect between the phase and the treatment before the treatment effects can be compared.<sup>[3,4]</sup>

### 1.2 Factorial Design

The factorial design is normally used to simultaneously evaluate the actual effects of two or several treatments and their synergistic or antagonistic effects. Thus, a  $2 \times 2$  factorial design concerning two products, A and B, comprises four arms: *verum* A + *verum* B; *verum* A + placebo B; placebo A + *verum* B; placebo A + placebo B. The expected benefit in terms of power depends on the degree of interaction between the two treatments. Such an approach is often used to evaluate different dose strategies.

### 1.3 Trial with Repeated Criteria Measurements Over Time

Rather than using a unique measure in a large number of subjects, measuring a given criterion several times in a smaller number of patients can be proposed. The power of such a study, and therefore the number of subjects required, depends on the degree of correlation between the individual measurements over time, the number of these measurements and the inter- and intrasubject variabilities.

The concern over ‘non-loss of chance’ has led to suggesting a few specific experimental designs, although these meet the rules of classical methodology.

#### 1.4 The 'Add-On' Design

All patients benefit from the reference treatment. During or towards the end of a trial using this design, and depending on circumstances, patients are randomised between the investigational add-on treatments versus placebo. Such a design, which presupposes that the reference treatment and the investigational treatment have different mechanisms of action, is often the only possible one in the case of a lethal disease, where a reference treatment has already proved its efficacy.<sup>[5]</sup>

#### 1.5 The Randomised Withdrawal Design

All patients initially receive the investigational treatment, but only those responding to the treatment continue in the trial. They are then randomly chosen to either continue with the treatment or receive the placebo. Patients from both groups are followed-up regularly on the basis of clinical or biological criteria until evidence of a relapse appears. They are then put back on the 'effective' treatment. This approach helps to reduce the period during which the patients receive the placebo, while exposing only the subjects who responded to the treatment to the placebo, thereby optimising the possibility of observing the effect of the treatment in the subjects who responded to the treatment.<sup>[6]</sup> However, in the case of slowly evolving and unpredictable illnesses with spontaneous remissions (multiple sclerosis [MS], for example), this trial design may needlessly expose patients to the harmful adverse effects of the treatments, without the possibility of proving the efficacy of the treatment in achieving remission.

This method is accepted in the US within the context of obtaining marketing authorisations.

#### 1.6 The 'Early Escape' Design

Patients are 'withdrawn' from a study with the early escape design if they meet a criterion of failure, defined *a priori*. The evaluation concerns the rate of failure, thereby minimising the duration of exposure to an ineffective treatment.

One of the significant difficulties of these trials concerns the definition of binary endpoints of failure/relapse, and the loss of power in the case of a significant number of patients lost to follow-up.

## 2. More Uncommon Methods

### 2.1 Sequential Trials

The technique of sequential trials was developed with the aim of reducing the time before making a decision, while also reducing the average number of subjects included in the study.

We can distinguish between sequential trials, in which the subjects enter the study 'in pairs' with a decision-making evaluation focusing on comparing the response of each pair, and group sequential trials that are more suited to clinical practice.<sup>[7,8]</sup> These group sequential designs are widely performed. They introduce interim analyses, either according to the group of patients included or to the number of events observed; the number and repetition of the analyses, more or less regularly spaced out over time, and planned in advance, guarantee pre-set type I and II risks of error. Two types of procedures for repeated analyses have been developed, namely the 'boundary' approach and the 'spending function' approach. The triangular test<sup>[9]</sup> comes under the first approach. In practice, the follow-up consists of recording the point (x, y) that corresponds to the two statistical values calculated from the data observed at each interim analysis on a graph with two boundary areas that delimit the continuation area. If the point is outside the continuation area, the trial is stopped with rejection of the null hypothesis or the alternative hypothesis, according to the respective position of the point in relation to the boundary curves. If the point is within the continuation areas, inclusion of the patients continues. In the second approach, conventional tests are repeated during the interim analyses, with the thresholds adjusted according to the number of tests. Here, we refer to the O'Brien and Fleming<sup>[10]</sup> procedure and the general context of the ' $\alpha$ -spending function' of Lan and DeMets.<sup>[11]</sup>

### 2.2 Flexible Trials

Flexible trials are extensions of sequential trials, in which the experimental design and/or the therapeutic plan are modified on the basis of study findings, or even data collected outside of the study during the interim analyses.<sup>[12,13]</sup> The major difference lies in the possibility of revising the protocol in steps planned in advance and replacing the original experimental design with another one, on condition that the type I risk of error dependent on the available observations is not modified during the controlled stages of the trial. The adaptation rules do not need to be set *a priori*. Changes to the experimental design may be based on the data collected during the trial (with or without unblinding) or outside the trial. Nevertheless, this approach does not necessarily lead to a reduced number of subjects, even though this may be the case when data from an internal pilot study are used. In this case, data from the pilot study may help to set up the rest of the study and also be considered for measuring the final effect of the treatment.

### 2.3 Adaptive Randomisation Methods (By Extension, Adaptive Trials)

On the one hand, adaptive randomisation methods answer the ethical concern of allocating the most effective treatment to a majority of the patients included, even though the ambivalence principle ('equipose') at the start of the study that justifies its 'raison d'être' does not favour the allocation of one of the treatments. On the other hand, these methods answer the concern of maximising the power of the trial in the case of unequal variances between the two treatment arms.<sup>[14]</sup>

Several adaptive allocation methods have been suggested, of which the most well known is the 'play the winner'<sup>[15,16]</sup> method, which is based on the ballot box model; in this model, the box is initially filled with the same number each of two differently coloured balls corresponding respectively to the two treatments being compared. Depending on the answer of each patient included, the allocation ratio of the treatments for patients continuing treatment is modified by adding to the ballot box, in the case of a positive response to the treatment, a ball of the colour corresponding to the treatment received and, in the case of a negative response to the treatment, the differently coloured ball. More complex plans may consider the predictive factors with allocations according to strata.

If such a method appears appealing as a concept, it remains relatively uncommon in practice. The major logistical difficulty is undoubtedly the capacity to quickly evaluate an endpoint that needs to be binary, as each included patient's follow-up is performed. It requires very precise planning.<sup>[17]</sup>

### 2.4 The 'N of 1' Method

The clinician is often faced with a patient for whom treatments are necessary, yet there is no trial that has proven their efficacy in this patient type. In this situation, the clinician may be tempted to successively administer several treatments recognised as being effective for other patient types in order to evaluate their effect on this particular patient and to choose the most effective treatment. In this 'N of 1' strategy, inspired by the field of behavioural sciences, the patient is his own control and the result is directly adapted to meet his needs. The increase in the number of 'micro-trials' over a population of heterogeneous patients may help to advance a few more general conclusions that are secondary to any treatment.

Guyatt et al.<sup>[18]</sup> have suggested a controlled, randomised strategy in which each patient receives pairs of treatment over consecutive periods and according to allocation methods that are similar to those of the crossover trial. Each sequence of pairs of treatment is replicated several times until both patient and doctor

are convinced of the best treatment in terms of benefits (tolerability and efficacy).

In theory, this approach is interesting when previous treatments have shown that the treatment was completely ineffective in certain patients or when a doubt persists on the possible efficacy in a given patient, or even when a patient wishes to try a treatment deemed more or less subjectively dangerous or ineffective by the practitioner. The approach is suitable for a chronic condition requiring a long-term treatment with the possible occurrence of associated adverse events, which appear soon after initiation of treatment and disappear just as quickly when the treatment is discontinued. The number of sequences of pairs of treatments administered is not pre-set, but the decision will be all the stronger since this number of pairs is high.

The experience of such an approach remains very limited.

### 2.5 The Pragmatic Trial

At this point, it is useful to mention the methodology of pragmatic trials, even though it was developed over 30 years ago by Schwartz and Lellouch,<sup>[19,20]</sup> and follows the same principle of helping practitioners to make therapeutic decisions adapted to the real prescribing situation, mostly after a proven intrinsic efficacy has been demonstrated. In this experimental approach, the treatment or strategy ultimately chosen is the one that provides the best results evaluated according to 'pragmatic' criteria, which evidently summarise the advantages and inconveniences of the therapeutic strategies being compared (success/failure or scale of quality of life). In theory, the analysis is reduced to its simplest expression, since the treatment chosen is the one that provides the best result without any calculation. The difficulties essentially concern measuring the difference between the treatments, which may be disregarded, and establishing an evaluation criterion that must reflect the overall 'medical service provided'.

### 2.6 Synthesis Methods *a Posteriori*

Synthesis methods *a posteriori* are those focusing on the analysis of existing data and do not initiate any new studies.

### 2.7 Meta-Analyses of Small Trials

As long as studies of sufficient quality are well conducted, the techniques of quantitative pooling may help to achieve enough power to demonstrate a therapeutic effect. The major difficulty lies in identifying protocols and unpublished data essential to ensure the homogeneity of the evaluation criteria and the patients included. The perspective of a meta-analysis conducted *a posteriori* should not prevent *a priori* consideration of

the possibility of conducting a multicentre trial, taking into account the characteristics of each centre.

### 2.8 The Decision-Making Analysis

The interest in the purely analytical approach of a decision-making analysis lies in the possibility of carrying out sensitivity analyses using various hypotheses relying on data derived from the literature and opinions of experts in the field. It aims to evaluate the possible changes in decisions in relation to the hypotheses regarding the models of effect and the probabilities of carrying out these decisions. The quality of a decision-making analysis largely depends on the initial planning to structure the question asked, and the clarity of the stages behind its execution.

## 3. Recommendations

After a review of the existing methodologies the following recommendations can be made, based on the principle that the quality criteria should not differ because there are a small number of patients in a trial:

- *Prefer international, multicentre protocols*  
Whenever possible, it is recommended that the number of patients guaranteeing sufficient power to prove the expected effects should be included – by preferring multicentre studies, by providing access to information on current or planned clinical trials, and by involving patient associations when these exist. International trials, more difficult in terms of logistics, should still be encouraged and may be facilitated by online protocol platforms and the transmission of e-CRFs (electronic case report forms) [GEREQ<sup>[2]</sup>].
- *Ensure impeccable methodological quality.* The same methodological quality as for ‘standard’ trials is set out as follows:
  - A written protocol with clearly defined objectives.
  - Methodology and assessment criteria adapted to the objectives.
    - A follow-up conducted in accordance with good clinical practice.
    - Validated data as well as a planned analysis within the protocol. If analyses that differ from those initially anticipated are necessary, they should be clearly justified.
- *Avoid an increase in the number of limited experiments without protocols from the outset*

The same methodology should be put in place for all trials whether institutional or industry-based, in particular when collecting data in relation to the planned objective. This point is particularly important, since it imposes the need to differentiate between the ‘unauthorised’ experiment performed by clinicians faced with the need to care for their patient and a scientific pro-

ocol, which, though based on a low number of patients, will have been planned, written and submitted to the competent authorities for validation.

- *Always include a prospective component*  
A prospective component is necessary to guarantee a sufficient level of proof. New data or information derived from previously recorded data in which the studied events occurred prior to the decision used to set up the study are often marred by biases that are fairly easy to control, particularly in small numbers.
- *Always identify initial hypotheses on the basis of a rationale that is as developed and detailed as possible*  
The hypotheses should be clearly defined from the beginning, even if this means using flexible methods that enable repeated adaptations throughout studies. The prerequisites available at the start of small trials are often limited and the proposed hypotheses are often not widely documented. Under these conditions, it would be justifiable to adapt the follow-up or analysis procedures in relation to the results obtained during the study. Nevertheless, the initial hypotheses should be identifiable and all changes during the study should be clearly detailed. Similarly, the arbitrary value of  $p < 0.05$  may need to be discussed again.
- *Set up a Data & Safety Monitoring Committee (DSMC)*  
Here more than elsewhere, the DSMC, a trial ‘safeguard’, should consist of members who are not involved in the recruitment and treatment of patients. Its mission should be to monitor how a study progresses, and to suggest decisions that would increase study performance as much as possible, by avoiding investigators from being both judge and judged. It should represent protection for patients and reassurance for clinicians.
- *Present individual data comprehensively and clearly*  
Data, including individual data, should be presented comprehensively and clearly. Because of the lack of quantitatively significant data, it is necessary for all information collected during these trials to be presented to the entire community with an interest in it. The graphic representation of individual data helps to develop a better understanding of the results obtained.
- *Systematically recommend a follow-up survey for treated patients, when marketing authorisations are granted as a result of small trials, especially when they are based on surrogate criteria*

When marketing authorisation are obtained from small trials, subsequent studies (observational studies) are recommended to confirm the observed results by means of clinical criteria, all the more so since the marketing authorisation will have been given according to surrogate criteria. Most of the time, surrogate criteria have not been validated, since they do not fit the requirements that allow the clinical effect to be assessed. If surrogate criteria are specifically justified in the context of small trials, it

is even more important for a follow-up of all treated patients to be collected to confirm the benefit to patients on the basis of clinically relevant criteria measured in the long term. This also enables the long-term efficacy and safety of the treatment to be confirmed, since the marketing authorisation has been granted on the basis of trials of limited duration.

- *Never set the rules of the game in stone*

The rules of the game should never be set in stone, since the information collected may change decisions. For example, for a rare disease without an available treatment, highlighting an unexpected beneficial effect from a treatment should authorise modification of the design of the analysis, given the difficulty of setting up a new trial with this new criterion.

- *Involve patient associations in all stages of the studies*

Patient associations should be involved, especially in the case of rare diseases and at all stages of the study. They help with recruitment by informing patients that the study is being carried out and also by facilitating the gathering of patient informed consent for the investigator. They may also have an advisory role during the drafting of the protocol regarding the feasibility of expected tests in relation to any possible patient handicaps, and in helping to distribute results when these are positive so that other patients may have access to the treatment.

- *Never forget the immediate interest of the present and future patient*

The interest of the present and future patient is, in essence, the fundamental element in these studies.

## Participants

C. Auriche (Afssaps, Saint-Denis), J. Bernard (Alliances Maladies Rares, Marly-le-Roi), D. Boichut (DRC, Paris), M. Bouhassira (Lilly, Suresnes), P. Boutouyrie (HEGP, Paris), P. Chaumet-Riffaud (AFM, Ivry), C. Chiron (Hôpital Saint-Vincent de Paul, Paris), Y. Costa (DRC, Paris), F. de Crémières (Wyeth-Lederlé, Puteaux), S. Etienne (CEREP, Rueil-Malmaison), P. Fender (CNAMTS, Paris), R. Giorgi (Hôpital de la Timone, Marseille), F. Girard (Sanofi-Synthelabo, Le Plessis-Robinson), D. Girault (Laboratoires Takeda, Puteaux), C. Grosskopf (Roche, Neuilly-sur-Seine), F. Gueyffier (Hôpital Cardiologique, Lyon), N. Hoog Labouret (Afssaps, Saint-Denis), Y. Juillet (LEEM, Paris), P. Lechat (Hôpital Pitié-Salpêtrière, Paris), S. Leroy (LEEM, Paris), L. Parmentier (Serono, Boulogne-Billancourt), J. Ropers (Afssaps, Saint-Denis), B. Salanave (CNAMTS, Paris), T. Simon (Faculté de Médecine

Saint-Antoine, Paris), E. Van Ganse (Hôpital Edouard Herriot, Lyon), M. Vray (Institut Pasteur, Paris).

## References

1. Evans CH, Ildstad ST. Small clinical trials: issues and challenges. Washington, DC: National Academy Press, 2003
2. GEREQ. Gestion électronique des données cliniques [online]. Available from URL: www.gereq.ca [Accessed 2004 July 1]
3. Matthews JNS. Small clinical trials: are they all bad? *Stat Med* 1995; 14: 115-26
4. Senn SJ. Cross-over trials in clinical research. Chichester: John Wiley & Sons, 1993
5. Chiron C, Marchand MC, Tran A, et al. Stiripentol in severe myoclonic epilepsy in infancy: a randomised placebo-controlled syndrome dedicated trial. STICLO Study Group. *Lancet* 2000; 356: 1638-42
6. Temple RJ. A Regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, editors. *Clinical measurements in drug evaluation*. Chichester: John Wiley & Sons Ltd, 1995
7. Armitage P. *Sequential medical trials*. 2nd ed. Oxford: Blackwell Scientific Publications, 1975
8. Whitehead J, Jones DR. The analysis of sequential trials. *Biometrika* 1979; 66: 105-13
9. Whitehead J. *The design and analysis of sequential Trials*. 2nd ed. Chichester: Ellis Horwood, 1992
10. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; 35: 549-56
11. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70: 659-63
12. Posch M, Bauer P, Brannath W. Issues in designing flexible trials. *Stat Med* 2003; 22: 953-69
13. Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Stat Med* 2003; 22: 971-93
14. Rosenberger WF, Lachin JM. *Randomization in clinical trials*. New York: John Wiley & Sons, 2002
15. Zelen M. Play the winner and the controlled clinical trial. *J Am Stat Assoc* 1969; 64: 131-46
16. Wei LJ, Durham SD. The randomized play-the-winner rule in medical trials. *J Am Stat Assoc* 1978; 73: 840-3
17. Rosenberger WF. Randomized play-the-winner clinical trials: review and recommendations. *Control Clin Trials* 1999; 20: 328-42
18. Guyatt GH, Sackett D, Adachi J, et al. A clinicians' guide for conducting clinical trials in individual patients. *CMAJ* 1988; 139: 497-503
19. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in clinical trials. *J Chronic Dis* 1967; 20: 637-48
20. Vray M. Pragmatic and explanatory trials: ask and answer different questions. *Appl Clin Trials* 1999; 8: 42-50

Correspondence and offprints: *Muriel Vray*, Institut Pasteur, Unité d'Epidémiologie Pathologies Emergentes, 28 rue du Dr Roux, 75015 Paris, France.  
E-mail : vray@pasteur.fr