The treatment of schizophrenia represents one of the most difficult areas of medicine for carrying out reliable and informative clinical trials of new medicinal products. The methodological issues that affect studies of neuroleptic agents are not unique. The evaluation of treatments for negative symptoms probably represents the most unusual methodological problem (not covered in this paper), but from a statistical perspective this problem has parallels elsewhere in medicine. The real reason why clinical trials in schizophrenia are so difficult is the fact that a number of methodological issues are present together and in a severe form.

This paper is concerned largely with trials that provide the confirmatory evidence of the efficacy of new medicinal agents, that is, those carried out during their phase 3 development or perhaps during the development of a new indication in phase 4. Hence, it is concerned only with controlled trials that provide the most reliable and informative evidence of efficacy for licensing decisions.

General guidance on the statistical issues that arise in confirmatory trials and that relate to regulatory decisions can be found in ICH E9 (ICH, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use).1 Guidance on the design, conduct, analysis, and interpretation of these trials in the field of schizophrenia can be found in the Committee for Proprietary Medicinal Products (CPMP) Note for guidance.2

Two issues require a broader introduction before discussing their impact on trials in schizophrenia. The first is the use of placebo. Recent debate of this topic was stimulated by changes to the Declaration of Helsinki. The second is the design of studies to evaluate maintenance treatment in the prevention of relapse and recurrence. With respect to both of these issues, specific implications for trials in schizophrenia are considered. Additional design topics addressed briefly are noninferiority designs, add-on designs, withdrawal designs, run-in periods on placebo, loss to follow-up, and short-term and long-term trials.

Keywords: neuroleptic; controlled trial; methodology; comparator; placebo; duration; Declaration of Helsinki; relapse; recurrence

Author affiliations: Statistical and Regulatory Consultant, Canterbury, UK; Visiting Professor of Medical Statistics, Department of Epidemiology and Public Health, University of Leicester, Leicester, UK

Address for correspondence: Prof J. A. Lewis, Hatch Green Cottage, Hatch Lane, Charnham Hatch, Canterbury, Kent, CT4 7LP, UK (e-mail: johnlewis@clinitrial.co.uk)
context of this paper, it is important to clarify the basis of the concerns surrounding its use and to explain the current resolution.

The second issue is the design of studies to evaluate the long-term maintenance treatment of an episodic, or partly episodic, condition. The use of medicinal products for this purpose arises in a number of psychiatric and other indications, and a terminology has been developed to communicate thoughts and conclusions. The words *relapse*, *recurrence*, and *rebound* all capture ideas that are important for the underlying scientific model and for the mechanism of action, but do they also capture ideas that are critical to the use and licensing of medicinal products and that must be individually assessed in clinical trials? Following a general discussion of these two methodological issues, this paper will then relate the discussion to the specific development of neuroleptics. A number of other methodological issues highlighted by the CPMP guideline will also be introduced at this point.

**The use of placebo**

In both Europe and the USA, the process of drafting clinical guidelines for the development of new medicinal products has often led to discussions concerning the acceptability of the use of placebo in controlled trials. There are those who take the view that it is unethical to expose patients to placebo treatment when approved medicinal products already exist for the condition in question. There are others who stress the vital nature of placebo-controlled clinical trials in establishing unequivocally the benefits of a new medication. At first sight, this appears to be a conflict between the optimal treatment of today’s patients and the optimal treatment of tomorrow’s patients. The ethics of this well-known conflict are a serious and difficult matter and one on which arbitration might reasonable be sought through the Declaration of Helsinki (hereafter referred to as the Declaration). Clinical trials sponsored by the pharmaceutical industry generally defer to the Declaration on ethical matters and a copy of it is attached to most protocols supported by the industry.

The wording of earlier versions of the Declaration did not provide much support for the use of placebo in controlled trials in the situations where doubt arose. However, there was uncertainty about its true interpretation, and there was also a widely held view that it was not intended to address the specific problems in question in pharmaceutical development. These doubts were sufficient to permit the use of placebo to continue relatively unhindered by these specific ethical concerns. There was hope that the revised version might clarify matters and provide comfort to those who felt that they might be in conflict with the wording, but not the spirit, of the Declaration. However, section 29 of the revised version contained the following text:

> “The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic and therapeutic methods. This does not exclude the use of placebo, or no treatment, where no proven prophylactic, diagnostic or therapeutic method exists.”

This clearly did not change matters. It was an immediate source of alarm to those responsible for the conduct and approval of clinical trials, whether based in research institutes, medical practice, the pharmaceutical industry, or regulatory bodies. The resulting arguments have been captured in a number of publications and official statements. Regulators in Europe and the USA all take the view that there are a number of circumstances where a placebo arm is acceptable and necessary in a controlled trial, even when alternative proven (and licensed) therapies exist. Firstly, it is argued that use of placebo is acceptable in the following situations:

- The period on placebo (and therefore not on the known effective agents) does not entail any additional risk of serious or irreversible harm to the patient.
- The patient (or their legal representative) is capable of giving, and gives, fully informed consent.
- The patient may request conventional treatment at any stage, or may be placed on such treatment by the treating physician.

The key item in this list is clearly the first one because the other items apply more generally to all clinical trial procedures. If the line of argument advanced in the first point is accepted, then it becomes important to have a clear understanding of what constitutes serious or irreversible harm in each specific medical situation. Secondly, it is argued that the use of placebo is scientifically necessary under those circumstances where active-controlled trials are unreliable and where their use would increase the proportion of erroneous clinical and regulatory decisions. This topic is thoroughly discussed in the ICH E10 regulatory guideline on the choice of control.12
In some areas of medicine, the sensitivity of a specific trial is an uncertain matter. For example, in the field of depression, there are plenty of examples of apparently good-quality, placebo-controlled trials of established and licensed agents that failed to detect a difference. Such trials are scientifically awkward and expensive, but clearly cannot and do not form the basis of regulatory approvals; they lead to delays and further research. However, in an area of medicine where the paradoxical failure of a placebo-controlled trial was a real possibility, a trial that used a licensed treatment as the sole comparator arm would also run the risk of failing to detect a real difference. That is, it might fail to pick up the inferiority of a test agent to a standard agent. The lack of difference from the standard agent could be equally paradoxical, but in this case it could lead to a positive licensing decision.

Thirdly, some alternative (and regularly used) design strategies using placebo avoid the apparent ethical dilemma, for example, the addition of a new medication or placebo on top of standard treatment (the “add-on” design). Following widespread critical comment, the World Medical Association (WMA) took the unusual step of issuing a statement on their website that modifies the position in the Declaration with respect to section 29. This states that:

“… a placebo-controlled trial may be ethically acceptable, even if proven therapy is available, under the following circumstances:

- Where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy or safety of a prophylactic, diagnostic or therapeutic method, or
- Where a prophylactic, diagnostic or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious or irreversible harm.”

The new statement reflects a position that is close to the regulatory viewpoint that has been expressed. The current website wording does not quite cover all eventualities in a watertight manner, but there is a reasonable chance that further negotiation could rectify this. Formal adoption of the new wording, or similar wording, within the Declaration requires ratification at the next meeting of the WMA in October 2002. Assuming that this is indeed the outcome, then a consensus on the continued ethical use of placebo appears to have been reached.

In various fields of medicine, there has been debate about the precise meaning of the terms relapse and recurrence when applied to the response of patients to drug treatment. Other terms such as rebound are often also brought into the same discussion. Relapse and recurrence both indicate a worsening of the patient’s symptoms. Relapse indicates an increase in the patient’s symptoms after successful treatment, but as part of the original episode of disease. It must therefore occur within a reasonably short time of treatment withdrawal. Recurrence, on the other hand, is a reemergence of the patient’s symptoms after a time without symptoms and is usually regarded as the onset of a new episode of disease. It is natural to wish to describe the effects of medicinal products on the possibility of relapse and recurrence because these are concepts in the treating physician’s mind that help to communicate the benefits and risks of treatment. Careful withdrawal of treatment may be needed to prevent relapse. Continuation of treatment may avoid relapse and prevent recurrence. However, in practice, it can be difficult to reliably distinguish between a relapse and a recurrence in an individual patient. It is even more difficult to carry out clinical trials that can distinguish between the effect of a treatment on relapse and its effect on recurrence.

Fortunately, for questions relating to the longer-term use of treatments, this distinction does not greatly matter. The key questions about length of treatment are usually straightforward questions such as:

- For how long should I continue treatment?
- If I have successfully treated a patient for 6 months, is further treatment clinically valuable?

There are designs of clinical trial that can answer these questions without necessarily distinguishing between effects on relapse and effects on recurrence. Although this may lead to problems concerning the drafting of indications, it does not affect decisions concerning how to use the treatment in the individual patient. A design that would shed some light on the first question above would be to randomize patients to, say, 2 months of active treatment followed by 4 months of placebo (regimen A), 4 months of active treatment followed by 2 months of placebo (regimen B), or 6 months of active treatment (regimen C). The outcome might be reappearance of positive symptoms in a suitably defined
manner. The superiority of the results on regimen B to those on regimen A accompanied by the similarity of the results on regimen B to those on regimen C would be an indication that B was the best treatment regimen.

An alternative approach to answering the first question would be to await a specific degree and duration of response, and then randomize to continuing or stopping treatment.

A design that would address the second question would be one that took patients successfully treated for 6 months and randomized them either to receive placebo or to continue active treatment for a further 3 months. Again the outcome could be taken as the reappearance of positive symptoms. The superiority of the active treatment arm would indicate the value of continuing treatment.

To detect an effect on recurrence per se would require the selection of patients who had been successfully treated with the agent under investigation. The treatment would then be stopped for a period of time in order to establish that the first episode was over and that the possibility of relapse had gone. Patients who relapsed (had positive symptoms) during this time would be withdrawn. The remaining patients would then be randomized to restarting active treatment or to placebo, and the reappearance of positive symptoms would be assessed and evaluated. Although a positive effect in such a study is likely to indicate a real effect on recurrence, it is hard to see that it would lead to the use of the treatment in a similar manner in clinical practice. Hence the practical importance of such a design must be doubtful, except as an exploratory research tool. Phase 3 trials should reflect the intended manner of use.

**Specific issues for clinical trials in schizophrenia**

**Placebo-controlled, parallel-group comparisons**

It should be clear from the discussion above that the scientific need to use placebo as a comparator depends upon whether trials against the currently licensed and standard agents would reliably detect differences between treatments if they existed. In practice, it also depends upon confidence that standard treatments will exhibit approximately the same size of effect in a new trial as they did when they were originally tested against placebo. This latter condition arises because it is necessary to be able to judge what proportion of the benefit of the comparator might be eroded by its replacement by the treatment under test. Any “noninferiority” trial (trial to show that the test treatment is no worse) against an active comparator involves prespecifying a “noninferiority margin” to define the degree of difference that is clinically important and that it is necessary to exclude.

In schizophrenia, the main problem relating to the use of active controls arises from lack of confidence that the size of the treatment effect of a comparator agent could be reliably predicted in a new trial setting. This is because there have been changes and developments in many aspects of the disease and its treatment that are critical to the size of the observed benefits in clinical trials. There have been changes to diagnostic criteria, efficacy criteria, the range of available treatments, and the way in which they are used. However, in addition to the problem of defining a noninferiority margin, the different profile of effects of newer agents on negative symptoms and extrapyramidal side effects may confound direct comparisons of their effects on positive symptoms. Furthermore, the large number of dropouts and noncompliers in schizophrenia trials makes it harder to ensure the robustness of the conclusions of a trial claiming to show the absence of a difference between two agents, than it would be for a trial claiming to show a difference from placebo.

Given that the historical basis for predicting the effect of potential active comparators is shaky, it appears that there is a scientific need to fall back on placebo-controlled trials for straightforward head-to-head comparisons. But are they acceptable from an ethical perspective? Is there any danger of serious or irreversible harm? This is a question for practicing clinicians to answer, but it would seem likely that, at the very least, patients for placebo-controlled trials would have to be selected with considerable care.

Assuming that a placebo-controlled trial is possible, there are strong reasons for including, in addition, a third arm on a standard agent at an adequate dose. The third arm provides an internal calibration of the efficacy results, which is especially useful when the primary outcome is a subjective rating scale. It also provides valuable information about relative effects on positive and negative symptoms.

**Alternative designs**

If there are ethical problems about the use of placebo, then alternative designs must be considered. The add-on
design is often a useful possibility, taking patients who are not satisfactorily controlled on standard therapy and randomizing them to a test treatment in addition or to placebo in addition. If this were successful, then a later trial could take patients satisfactorily treated with the combination of treatments (test plus standard) and randomize them to continuing on the combination or the test treatment alone. Patients randomized to test treatment alone would have suitable rescue therapy available. Data collected in this way would form a sound basis for subsequent extension to broader first-line use of the test agent through active-controlled trials.

Another possibility is to carry out a straightforward placebo-controlled comparison, but to withdraw patients from placebo or test treatment as soon as they exhibit unacceptable symptoms. The time to withdrawal is then a suitable outcome measure. Patients on placebo are only exposed to risk for as long as they are acceptably treated by placebo.

**Run-in periods on placebo**

Initial washout periods are commonplace in trials of antischizophrenic agents. This is for sound methodological reasons related to properly characterizing the patients and to ensuring that the effects of withdrawing any previous treatment do not contaminate the observed effects of the test treatment. It is also common to treat all patients with placebo during the run-in.

The case against using placebo during the run-in has been argued strongly by Senn. He points out that this stratagem involves the treating physician deceiving the patient, whereas in more conventional uses of placebo both are in the same state of ignorance. As far as the subsequent comparison of randomized treatment arms is concerned, it would be just as acceptable to have a run-in without treatment: it does no harm to the main objective of the study. The onus to prove their case lies squarely on those who believe that placebo treatment is necessary during a run-in.

**Losses of patients from clinical trials**

The incidence of dropout from clinical trials in schizophrenia is high. This is one of the factors that make these trials particularly difficult to interpret because the biases introduced by dropouts are difficult to assess. All possible steps should be taken to minimize the number of dropouts and to shed light on the potential bias they induce. The reasons for dropout should be carefully documented. After stopping their trial medication, dropouts should still be followed up as fully as possible as planned in the protocol. Key measurements should also be made at the time of stopping treatment.

The primary analysis of a placebo-controlled comparison should include all randomized patients regardless of dropout. A “per protocol” analysis should support the primary analysis. There should be a full exploration of the sensitivity of the main results of the trial to the influence of the dropouts, taking into account the reasons for dropout and the corresponding potential biases that they might cause.

**Short-term trials**

The efficacy of a neuroleptic agent can generally be established in a short-term trial lasting about 6 weeks, studying acute exacerbations of the disease. A dose-ranging study might include three or more doses, placebo (ethically justified, as described earlier), and a standard treatment arm, making five treatment arms in all, to establish the optimal dose and the lower end of the dose range. A phase 3 confirmatory study would use the dosing regimen intended for licensing and would also ideally include placebo and active control.

**Long-term studies**

The difficulties inherent in schizophrenia trials make it imperative that licensing decisions are made on the basis of controlled trial data. It is not sufficient to monitor a group of patients exposed to long-term therapy and record their progress. The data from such a study would probably be supportive, especially for safety purposes, but would not establish a regulatory claim. The duration of controlled data adequate to establish use as maintenance therapy is of the order of 1 year. It is difficult to envisage ethical use of placebo over this timescale and so active control seems inevitable. If a straightforward randomized comparison over a period of 1 year is undertaken, then it will be necessary to defend the sensitivity of the trial, that is, its ability to detect clinically important differences from the active control, if they exist. This will probably have to take into account a high level of dropout and noncompliance, and that could clearly pose problems.
Because of these problems, it may be more profitable to make use of the designs described earlier in the section *Long-term studies of efficacy: relapse and recurrence*. This might be done sequentially, first establishing that 3 or 6 months’ treatment was better than treatment that stopped after the acute exacerbation, and then going on to 1 year. Patients whose acute episode was successfully treated by the test treatment could be randomized to placebo (stopping treatment) or test treatment. Those who survived successfully on test until 6 months, say, could then be randomized again to placebo or to test treatment. In this way, the value of continuing treatment at each selected time point would be established. The problem of dropouts would be reduced because only those who reached each time point would be rerandomized.

In trials of this nature, a natural primary outcome measure would be the time to the reappearance of positive symptoms, suitably defined. A “time to event” analysis of this outcome would be appropriate. In this analysis, no distinction need be made between relapse and recurrence in the primary analysis, although secondary analyses might consider this distinction. Other measurements of symptoms and adverse effects could also be used to support the primary outcome.

A positive conclusion of a trial using this type of design implies that continued treatment up to and beyond the point of randomization is worthwhile. Hence the later that randomization is deferred, the longer the treatment period that can be supported by the trial. However, the later that randomization is deferred, the more patients will leave the trial before randomization, and so the more must be entered at the start. In addition, after randomization the trial must continue for a reasonably long period of time in order to collect sufficient “events.” There are likely to be limits on the numbers of patients that can be recruited initially and on the overall length of the trial that will place practical restrictions on this design.

---

**Diseños y duración de estudios y elección de fármacos de comparación incluyendo el empleo de placebo**

Este artículo discute algunos aspectos metodológicos que son relevantes para el diseño de ensayos controlados de nuevos productos medicinales para ser utilizados en el tratamiento de la esquizofrenia. En forma más general y en mayor extensión se abordan dos aspectos. El primero es el empleo de placebo. Cambios en la Declaración de Helsinki estimularon un reciente debate acerca de este tópico. El segundo aspecto es el diseño de estudios para evaluar el tratamiento de mantención en la prevención de recaídas y recurrencias. Se han considerado las implicancias específicas que tienen ambos aspectos para los ensayos en esquizofrenia. Otros temas de diseño, mencionados brevemente, son diseños de “no inferioridad,” diseños de tratamientos combinados, diseños de evaluación de abandono de tratamiento, estudios con fare de preinclución con placebo, con pérdidas del seguimiento y estudios a corto y a largo plazo.

**Schémas d’études, durée des études et choix des comparateurs y compris d’un placebo**

Cet article passe en revue quelques points de méthodologie concernant les schémas d’études contrôlées portant sur les nouveaux médicaments utilisés dans le traitement de la schizophrénie. Deux points sont traités de façon plus générale et plus longuement. Le premier concerne l’utilisation d’un placebo ainsi que le récent débat sur ce sujet suscité par les modifications apportées à la Déclaration d’Helsinki. Le second se rapporte aux schémas d’études destinées à évaluer le traitement d’entretien pour prévenir les rechutes ou les récurrences. Les aspects spécifiques de ces deux points vis-à-vis de la schizophrénie sont détaillés, avant d’aborder plus rapidement d’autres aspects méthodologiques tels que les schémas de « non-infériorité », les schémas de traitement associé, les schémas avec sorties d’étude, les études avec phase de préincluision sous placebo, avec perdus de vue et les essais à court et à long terme.
REFERENCES

1. Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonization E9 Expert Working Group. Stat Med. 1999;18:1905-1942.
2. Committee for Proprietary Medicinal Products (CPMP). Note for Guidance on the Clinical Investigation of Medicinal Products in the Treatment of Schizophrenia (CPMP/EWP/SS9/95). London, UK: European Agency for the Evaluation of Medicinal Products; 1998.
3. The World Medical Association, Inc. Declaration of Helsinki (revised). Available at www.wma.net. Ferney Voltaire, France: World Medical Association, Inc; 2000. Accessed 26 November 2002.
4. Rothman KJ, Michels KB. The continued unethical use of placebo controls. N Engl J Med. 1994;331:394-398.
5. Committee for Proprietary Medicinal Products. EMEA/CPMP Position statement on the use of placebo in clinical trials with regard to the revised Declaration of Helsinki. Available at www.emea.eu.int. London, UK: European Medicines Evaluation Agency (EMEA); 2001. Accessed 26 November 2002.
6. Lewis JA, Jonsson B, Kreutz G, Sampaio C, van Zwieten-Boot B. Placebo-controlled trials and the Declaration of Helsinki. Lancet. 2002;359:1337-1340.
7. Temple R, Ellenberg SS. Placebo-controlled trials and active-controlled trials in the evaluation of new treatments. Part 1. Ethical and scientific issues. Ann Intern Med. 2000;133:455-462.
8. Temple R, Ellenberg SS. Placebo-controlled trials and active-controlled trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. Ann Intern Med. 2000;133:464-470.
9. Emanuel EJ, Miller FG. The ethics of placebo-controlled trials—a middle ground. N Engl J Med. 2001;345:915-919.
10. Shapiro HT, Meslin EM. Ethical issues in the design and conduct of clinical trials in developing countries. N Engl J Med. 2001;345:139-141.
11. Moreno JD. Placebos on trial. Ann Neurol. 2001;49:558-560.
12. Choice of Control Group in Clinical Trials. ICH topic E10, ICH Harmonised Tripartite Guideline. London, UK: European Agency for the Evaluation of Medicinal Products; 2000.
13. Storosum JG, Elferink AJA, van Zwieten BJ, van den Brink W, Gersons BPR, Broekmans AW. Short-term efficacy of tricyclic anti-depressants revisited: a meta-analytic study. Eur J Neuropsychopharmacol. 2001;11:173-180.
14. Senn S. Statistical Issues in Drug Development. Chichester, UK: John Wiley & Sons; 1997:71-72.
15. Senn S. Are placebo run-ins justified? BMJ. 1997;314:1191.