State of the art

Randomized controlled trials in schizophrenia: opportunities, limitations, and trial design alternatives

Christoph U. Correll, MD; Taishiro Kishimoto, MD; John M. Kane, MD

Randomized controlled trials (RCTs) have become a cornerstone of evidence-based medicine, and therefore have an important impact on clinical decision-making and clinical practice. The clinical trial can be a much more precise and accurate assessment of therapeutic potential than the anecdotal report or uncontrolled case series. However, clinical trials have important limitations in terms of feasibility and generalizability

State-of-the art clinical trial design and methodology are enormously important for the advancement of the field. In contrast, the critical relevance of trial conduct and implementation have only more recently been the focus of discussion and research. Although randomized controlled trials are generally considered the gold standard for the assessment of pharmacologic and nonpharmacologic interventions in medicine, trials are vulnerable to complications and influences that can seriously compromise their success. Like interventions, trial design and conduct are also contextual. They need to be individualized and adapted to a number of relevant variables, such as setting, population, illness phase, interventions, patient and rater expectations and biases, and the overall aims of the investigation. While this means that there is no unified approach possible, certain general principles and guidelines require careful consideration. Knowledge of basic solutions and alternatives, and the recognition of the complex challenges that need to be addressed proactively can help to minimize unwanted outcomes, including trial failure and uninformative or falsely negative outcomes. Moreover, novel design alternatives need to be explored that target sample enrichment according to the study question and enhancement of precision in the measurement of relevant outcomes. We propose two novel design strategies that take advantage of the recently validated early antipsychotic response paradigm (that has also been observed with antidepressants and mood stabilizers). In the "early responder randomized discontinuation design" all patients are assigned to the active drug, and only those who had at least a minimal response at 2 weeks are enrolled in a double-blind, placebo-controlled discontinuation trial, enriching the placebo controlled trial portion with true drug responders. In the mirror image "early nonresponder randomized dose increase or augmentation design," early nonresponders at 2 weeks are assigned to staying on the medication or going either to a higher dose or an augmentation agent. It is hoped that through increased attention to the issues raised in this article and further refinement of trial methodology and conduct, the field will make much needed additional progress in the prevention and treatment of schizophrenia.

Author affiliations: The Zucker Hillside Hospital, Psychiatry Research, North Shore - Long Island Jewish Health System, Glen Oaks, New York, USA (Christoph U. Correll, Taishiro Kishimoto, John M. Kane); Albert Einstein College of Medicine, Bronx, New York, USA; The Feinstein Institute for Medical Research, Manhasset, New York, USA; Hofstra North Shore LIJ School of Medicine, Hempstead, New York, USA (Christoph U. Correll, John M. Kane)

Address for correspondence: Christoph U. Correll, MD, Division of Psychiatry Research, The Zucker Hillside Hospital, 75-59 263rd Street, Glen Oaks, NY 11004, USA
(e-mail: ccorrell@lij.edu)

and can also fail or prove to be erroneous in their conclusions. The process of patient selection in clinical trials further highlights the strengths and weaknesses of the current nosology, and the prevalence of comorbid conditions and other factors can also influence treatment response. Moreover, the clinical trial serves to highlight the ethical and scientific tension between striving for the common good and the treatment of the individual person. When and to what extent the use of placebos is appropriate when proven effective treatments are available is an important and complex issue about which reasonable people may disagree. In order for RCTs to serve the common good in an optimal fashion, clinicians, health care policy makers and other individuals with a stake in influencing and evaluating clinical care must be informed consumers of clinical trial data. Similarly, for clinical trials to be informative, those involved must carefully consider the opportunities and challenges of trial design, methodology, conduct, implementation, and interpretation.

In designing and conducting clinical trials, there is a constant tension between the "perfect" and the "feasible," the desirable and doable, and between striving for scientific excellence and clinical impact. This dichotomy needs to be resolved proactively, consciously realizing and weighing the cost and benefit of each decision aimed at resolving this unavoidable conflict. It is in that spirit that we provide the following overview of controlled clinical trials in schizophrenia. We will first discuss the changing clinical and scientific context in which RCTs are taking place, followed by a discussion of specific trial components and their importance.

Historical developments

The somewhat serendipitous observation that chlorpromazine had a pronounced "calming" activity that extended to benefits for psychotic signs and symptoms was one of the great advances in 20th-century medicine. This effect was observed without the benefit of an RCT. Chlorpromazine was subsequently approved by the Food and Drug Administration in 1954, and by 1964, approximately 50 million people around the world had been treated with this medication.

In 1949, the World Health Organization published the sixth revision of the *International Statistical Classification of Diseases (ICD),* which for the first time included a section on mental disorders.[1] The first official *Diagnostic and Statistical Manual of Mental Disorders (DSM)* was published in 1952 by the American Psychiatric Association.[2] Diagnostic criteria were not really specified for discrete disorders until the third edition of *DSM (III),*[3] which attempted to improve the validity and reliability of psychiatric diagnosis. This, in turn, had enormous implications for clinical practice, clinical research and drug development.

In 1969, Klein and Davis published a seminal work entitled *Diagnosis and Drug Treatment of Psychiatric Disorders.*[4] In the introduction, they wrote,

> "We may be fortunate to be entering a period in which rational comparative study will become standard for therapeutic decision. Although clinical hunches and results of clinical experience are important factors in the determination of proper treatment, the findings of research studies, particularly those which are done with controlled double-blind technique, provide the behavioral scientific data for informed decision."

By 1969, Klein and Davis identified 126 controlled studies comparing antipsychotic drugs and placebo in which the medications were found to be more effective and 26 comparisons in which they were not.[4] They also examined the role of dose adequacy and found that most of those studies that found chlorpromazine to be ineffective used very small doses, and all 23 studies that employed doses over 500 mg/day were positive. Similarly, in all studies, which were judged to be methodologically rigorous, the phenothiazine derivatives (and reserpine) were shown to be more effective than the control conditions. These data led to an enormous shift in clinical practice, with antipsychotic drugs becoming the critical component in the treatment of schizophrenia.

Treatment-refractory illness and clozapine

With the development and initial evaluation of clozapine in Europe, early observations suggested a novel compound had been developed with a qualitatively different clinical profile that included the relative absence of drug-induced extrapyramidal effects and potentially superior efficacy in patients who had failed other anti-dopaminergic agents. However, a series of cases of agranulocytosis[5] led to a delay in the further development of clozapine in the US. Based on a large RCT with prospective validation of treatment refractoriness demonstrating clozapine's superiority over chlorpromazine in refractory schizophrenia,[6] the FDA approved

clozapine with the narrow indication for treatment resistant patients in 1990. Since then, clozapine's singular role in treatment-refractory patients with schizophrenia has been confirmed[7] and its role in the management of suicidality has also been established.[8] Nevertheless, recent meta-analyses[9] did not uniformly confirm clozapine's superiority over other antipsychotics in schizophrenia. Again, several design issues need to be considered when evaluating this inconsistency, including inappropriately low doses of clozapine[9], as well as the lack of selection for truly resistant patients.

Attention to first-episode schizophrenia

Beginning in the mid 1980s increased attention to first-episode patients seemed warranted to evaluate treatment outcomes that are unconfounded by the effects of prior treatment, multiple relapses, and chronic illness.[11-13] Studies revealed cognitive and psychosocial deficits that were present at illness onset,[14] a long duration of untreated psychosis prior to first mental health contact,[15] increased sensitivity to medication side effects,[16] but also a better treatment response compared with more chronically ill patients.[17] Exploring biological heterogeneity and treatment response at this phase has become an important focus. In addition, as part of the move toward the early treatment of schizophrenia, and the response to new FDA incentives, the efficacy of antipsychotics has also demonstrated in adolescents with schizophrenia.[18] In adolescents, appropriate selection criteria and trial design considerations are also critical.

Comparative efficacy and effectiveness of first-generation and second-generation antipsychotics

With the introduction of second-generation antipsychotics, there were observations of lower extrapyramidal side-effect burden and tardive dyskinesia risk and expectations of superior efficacy for positive, negative, and cognitive symptoms.[19] Initial efficacy studies seemed to confirm the superiority of second-generation antipsychotics, but the comparator consisted predominantly of haloperidol, used at moderate to high doses and often without anticholinergic cotreatment, which made early treatment discontinuation and secondary negative symptoms more likely in haloperidol treated patients. Since then, a series of acute phase and longer-term studies

have been completed,[20-24] including large efficacy-effectiveness hybrid trials that compared first- and second-generation antipsychotics. These data have led to varying conclusions. Interpretations include that there is no difference between first- and second-generation antipsychotics, that second-generation antipsychotics are superior to first-generation antipsychotics, that some second–generation antipsychotics are superior to either all or some first-generation antipsychotics, in general, or in certain efficacy and/or side effect domains, or in patient subgroups that are not yet easily identified prior to choosing a specific agent. Since such a number of divergent interpretations have been offered, this indicates that blanket statements do not do justice to the complex data base. Moreover, in comparative trials, design issues are highly relevant to interpretation of the data,[25] including sample size, choice of dose of the study drug and active comparator, prior treatment, blinding, duration of treatment, patient and rater expectations and biases, choice of outcomes, handling of dropouts and interpretation of the data, all of which we will discuss in detail below.

Shifting adverse event focus to physical health

Following the predominant use of second-generation antipsychotics, there has been a shift in side effect concerns from Parkinsonism and tardive dyskinesia, to physical health risks and outcomes that are associated with decreased longevity.[26-30] Even more so than the study of tardive dyskinesia, the study of adverse effect risks that are either rare or distal outcomes that occur after many years of illness and antipsychotic exposure pose formidable challenges. This applies to study of sudden cardiac death as a potential consequence or QTc prolongation or other arrhythmogenic properties of antipsychotics,[31] as well as to diabetes and cardiovascular and cerebrovascular morbidity and mortality. Since these outcomes occur generally after many years or represent premature onset of disorders that also occur in the general population, RCTs might not be the best way to assess the comparative safety of antipsychotics.[32] In fact, RCTs have largely focused on the assessment of risk factors (such as weight gain, lipid and glucose abnormalities) for cardiovascular and cerebrovascular illness, rather than on the development of such illnesses themselves. An exception is the assessment of death and cere-

brovascular events associated with antipsychotics in the elderly, which, by definition, is an enriched, high-risk cohort. Even here, however, the increased risk with antipsychotics was only uncovered after pooling all data from placebo controlled RCTs in meta-analyses.[33] These examples highlight the fact that for some outcomes, such as rare and temporally distal events, meta-analyses,[34] pharmacoepidemiologic studies,[31,35] and cohort studies[36] or registries are more useful than RCTs. This is true, despite the considerable limitations of meta-analyses and pharmacoepidemiologic studies and registries. Thus, these methods are merely hypothesis-generating, requiring follow-up with mechanistic studies and RCTs in enriched samples that were informed by meta-analytic, pharmacoepidemiologic, and registry data.

Raising the bar for outcomes

In addition to a broadened focus on physical health, outcomes other than symptomatic improvement have become standard in the field including more systematic and operationalized approaches to measuring response, remission and recovery,[37-40] subjective well-being and quality of life,[41,42] cognition[43-45] and psychosocial/vocational performance.[46-48] These have become end points of increasing importance in routine practice and in clinical trials. However, focusing on such outcomes also requires specific considerations, including treatment modalities, trial duration, assessments, end points, etc.

Targeting individualized treatment

In addition to raising the bar for outcomes, the ways to measure and predict them have also become topics of increasing interest. Efforts at increasing the predictability of outcomes for individual patients (ie, personalized medicine) have included clinically driven nosological and phenomenological approaches, but, so far, these have not really succeeded.[49] Current approaches that do not yet have consistent clinical applicability include the use of genetics, neuroimaging, neurocognition, and blood- or tissue-based biomarkers and sets of biomarkers, also called biosignatures.[50] Similarly, developments are underway to define biomarkers as surrogate end points in drug development.[51] To achieve personalized psychiatric treatments, specific design considerations are needed. These include ways of decreasing the heterogeneity of the study population and the parsing of clini-

cal and biological variables that are relevant for specific mechanisms and treatment effects. In addition, relevant treatment mediators and moderators can serve as selection criteria and randomization or stratification variables. However, obviously, these approaches will depend on the identification of markers that predict treatment outcome to specific interventions and that are not "just" markers of general illness severity and responsiveness (although even such general markers could, for example, help to facilitate early identification of patients who should have a trial of clozapine).[52,53] One such potentially useful "biomarker" or endophenotype that we will discuss subsequently is the presence or absence of early minimal clinical response, which might be useful to enrich samples for specific studies.

Commercialization and globalization of clinical trials

Another historical element in any discussion of RCTs is the "commercialization" and "globalization" of research. In the early days of clinical trials in psychopharmacology, there were a relatively small number of largely academic sites which participated in the design and conduct of such investigations. The development of a specialized clinical trials network was greatly facilitated by the establishment of the "Early Clinical Drug Evaluation Units" by the Psychopharmacology Research Branch at NIMH. Over the last few decades, there has been an enormous shift in the locus of clinical trials from academia to more commercial sites, from the US and Europe to many other countries, and a much greater involvement of a variety of vendors and middlemen in the management and conduct of such trials. The reasons for and consequences of this shift are complex and varied, and a detailed discussion is beyond the scope of this paper. This phenomenon will be discussed subsequently in relation to patient recruitment as well as study implementation and management.

Designing RCTs in schizophrenia

The essential first step in designing any trial, however, is to determine 'what is the question?' All too often investigators attempt to address more than one question in the same clinical trial. Although there is often an opportunity to collect meaningful data on several primary and secondary outcome measures simultaneously,

in some cases (eg, the efficacy of a putative therapeutic agent for cognition on negative symptoms or agitation), a specific and distinct type of sample and trial design is needed. Once the primary question is established, patient selection, randomization strategy, treatment selection and controls/comparator(s), trial duration, assessment measures, power analysis, and statistical plan will be the focus of attention. The degree to which appropriate decisions are made regarding these issues will be critical in the success of the trial. We will return to these issues in the subsequent sections of this paper.

Types of trials

Like any other scientific method, RCTs have specific strengths and weaknesses *(Table I)*. These need to be considered and adapted to the specific aim of the investigation. One important decision is the degree to which real world characteristics of populations, treatments, and procedures are systematically restricted and standardized. There are a number of broad categories in which RCTs can be placed. Efficacy studies involve clearly defined and often narrow populations of patients who can be studied with some frequency and intensity with a variety of measures, which would not likely be used in routine clinical practice. Primary outcomes of interest ordinarily include symptom reduction on a validated and reasonably comprehensive scale that is rarely used in routine treatment. While this procedure increases the chance of finding specific efficacy or tolerability signals, the sample and settings in which this signal is detected might become so restricted that as few as 10% to 20% of individuals with a given diagnosis are enrolled,[54,55] affecting the generalizability of the findings.

Effectiveness trials, on the other hand, involve a broader patient population intended to facilitate greater (although still imperfect) generalizability of the trial results to "real-world" patients and less intensive and frequent assessments with more objective and easily determined outcome measures, such as all-cause discontinuation, hospitalization, or death. Effectiveness studies are more likely to include issues such as adherence, tolerability, and quality of life as well as evaluation of the overall impact of a specific treatment. Large "simple" trials or pragmatic/practical trials ideally should include much larger numbers of patients than would be feasible to include in an efficacy study; however, large, multicenter, recruitment efforts can certainly result in

sizeable numbers of subjects participating in efficacy studies as well. Other types of trials include crossover and adaptive or sequential designs, which are described in more detail below.

Patient characteristics and selection

Sponsors and funding agencies usually conceptualize acute treatment as that involving recently relapsed or exacerbated patients (often newly admitted to a hospital). However, some investigators include patients who have been chronically symptomatic, but with a symptom severity level great enough to qualify for the study. This alternative approach can lead to the inclusion of patients with varying degrees of potential drug responsiveness. In addition, even if patients are recently relapsed and admitted to hospital, if they have been treated for 2 or more weeks prior to going into the study, their partial treatment (or established poor response) might limit the determination of the full therapeutic potential of the treatments to which they are assigned. That patients have already been partially treated does not necessarily mean that a significant drug effect will not be evident, but the magnitude and time course of that effect and, ultimately, the statistical power can be affected.

Age is another important consideration not only because of the relative prevalence of comorbid medical conditions with increasing age, but also because of potential differences in patterns of response depending upon the phase of illness. Consideration should be given to both age and duration of illness in selecting patient populations for clinical trials. Response patterns, dosage requirements, and vulnerability to side effects are, for example, quite different in first-episode vs more chronically ill patients.[16,56]

The duration of the current episode is a factor which has been given insufficient attention in most clinical trials. Recent work[57-61] has suggested that a substantial proportion of the response to antipsychotic agents (at least for "positive" symptoms) is likely to occur in the first 1 or 2 weeks of treatment. Therefore, as suggested previously, if a patient has been treated for more than a week or 2 prior to entering the trial it is likely that considerable drug response has already taken place. In some cases, patients need to be at least partially treated before they will be available to participate in a clinical trial, but documentation of length and type of prior treatment in the current episode would help in interpreting results, par-

Selected trial characteristics	Strengths	Weaknesses
Restricted inclusion criteria	High internal validity	Limited external validity/generalizability
	Specific signal detection capacity in carefully selected target population	Difficulty assessing optimal dosing in unrestricted, more heterogeneous or seriously ill sample
	Usable for regulatory and registration purposes	Decreased knowledge about response and side effect patterns in patients with comorbid psychiatric and/or medical conditions
		Slow enrolment
Randomization	Controlling for measured and, especially, unmeasured group differences	Selection bias towards a less generalizable sample (less severely ill, more chronically ill patients; patients with prior stabilization or treatment phase)
	Challenges associated with placebo controls, maintaining blind in the face of specific adverse effects	Selection bias towards a less generalizable sample (less severely ill, more chronically ill patients; patients with prior stabilization or treatment phase)
Small-to-medium sized samples	More homogeneous and carefully characterized samples	Reduced generalizability
		Need for multiple research-oriented sites
		Low signal detection for rare outcomes
Specialized settings	Greater potential for careful selection and diagnostic/assessment	Reduced generalizability
	More control over study procedures	Fewer potential sites
	Greater comfort using placebo controls	Greater likelihood of professional patients
	Study conduct by well-trained personnel with allocated research time	Lower enrolment rates
		Potentially less access to patients of interest (eg, acute exacerbations, drug-naïve)
Frequent visits	Controlled treatment and assessment	Reduced generalizability
	Better assurance of patient safety	Reduced enrolment
	More systematic quantitative assessments for therapeutic and adverse effects	Increased burden and dropouts
	Increased opportunity to facilitate/monitor adherence	Frequent quantitative assessments not done in clinical practice
	Greater ability to use informative laboratory tests	
	Better assessment of use of ancillary services or use of other medications	Potential influence on specific, investigated effect by increased contact
Comprehensive assessments	Specific assessment of measurable outcomes (including safety and tolerability) using validated and reliable scales administered by well-trained personnel	Primary/secondary outcomes rarely assessed in clinical practice.
	Usable for regulatory and registration purposes	Patient/caregiver rated outcomes, quality of life and functional capacity rarely assessed in clinical practice
		Use of quantitative measures unlikely in clinical practice and clinicians not trained in their use
		Need for careful training and ongoing supervision of raters
		Increased burden and dropouts
Cost	Usable for regulatory and registration purposes	Increased per-patient costs
	Usable for potential marketing	

Table I. Randomized controlled trials: strengths and weaknesses.

ticularly in failed trials. This also could help to eliminate patients who have been chronically symptomatic and have not had a recent exacerbation.

In this context, it is also important to consider the presence and potential duration of a washout period (which may or may not be placebo-controlled, double- or single-blind). In some cases, investigators might use this not only to eliminate the adverse effects of other drugs, but also to see the severity of untreated symptoms. It is important from an ethical and scientific standpoint that it is clear what the expectations from a washout period really are. In addition, as we will discuss subsequently, we do not have good data on the time course of potential symptom worsening among those patients who are partially treated and who then have their medication discontinued. There are a number of practical constraints on drug washout, which include increased length of trial duration, hospital stay, staff requirements, etc. Without an adequate washout, however, it can be difficult to establish a true baseline for both psychopathology and adverse effects. In addition, there might be withdrawal effects from the prior agent,[62] which could complicate interpretation of data on a newly instituted medication. The use of a concurrent placebo group can help to mitigate some of these concerns, and this is one important argument for the use of placebos in such trials.

Documentation of response to previous treatment in prior episodes can be difficult to obtain on a retrospective basis; however, this can also be important in eliminating potential poor or partial responders. Many current trials stipulate this exclusion criterion, but little is done to validate these assumptions and, given the pressure to enter subjects, there is likely some slippage in this domain.

There is increasing focus on studying patients who have poor or partially responsive symptoms in specific domains of psychopathology, such as negative, positive, or cognitive dysfunction. The selection of such patients is also a challenge, as the retrospective documentation of the stability and persistence of symptoms despite adequate trials (and with some evidence of treatment adherence) can also be difficult and requires special effort in both the design and execution of appropriate RCTs. In addition, cross-sectionally, negative symptoms can be difficult to differentiate from drug-induced Parkinsonism, depression, demoralization or guardedness. Negative symptoms may also be a secondary (possibly adaptive) response to positive symptoms. Therefore, a longitudi-

nal, prospective approach might be necessary to identify eligible patients for such studies.

The presence of comorbid substance abuse and/or other psychiatric conditions is often an exclusion criterion in efficacy trials, but not in practical trials. Suicidality has usually been an exclusion criterion for clinical trials, but it has been shown to be a feasible target for RCTs it its own right with appropriate safeguards.[63] The presence of comorbid medical conditions or abnormal laboratory tests has also become an increasing concern given the potentially high rates of metabolic syndrome, smoking, etc. among people with schizophrenia.

Randomization strategies

Most RCTs use even randomization strategies, which assign equal numbers of patients to each treatment arm. This makes sense since the smallest arm will determine the statistical power involved in the comparison between treatments. However, there are some situations where equal randomization might not be appropriate. Equipoise randomization[64] has been used in some trials and can provide patient and physician input into the potential options involved in the treatment assignment without doing away with randomization altogether. Adaptive design is another strategy which entails the use of data already collected to influence subsequent randomization of treatment groups.[65] This can be particularly useful in studies involving dose-finding; however, it can also lead to premature closure of a study arm or an entire study based on relatively small amounts of data. Sequential design allows pairing of individuals who are assigned to alternate treatments with the results of each subsequent pair contributing to an ongoing analysis of the odds of reaching a significant effect or not.[65] This can provide statistical significance with a small number of patients if the treatment effect is substantial, or it can provide an initial confirmation of the null hypothesis. Such a design has rarely been used in RCTs in schizophrenia,[66] but should be considered.

Stratified randomization is an important tool to ensure that the treatment arms are balanced on a small number of potentially important mediating or moderating variables. The number of such variables will be determined by the overall sample size. Crossover studies are also conducted in some situations, but given potential medication carryover effects and uncertainty as to how vulnerable patients are to returning to their baseline state

make placebo-controlled crossover trials less informative in psychiatry than they might be in some medical conditions. So-called "switch" studies are often done in psychiatry, but unless there is a control group they cannot be considered RCTs. Switch studies are obviously much more informative if patients are randomized to be switched or to stay on the original treatment, in a double-blind fashion.

Blinding

While blinding/masking is an important feature to minimize expectation biases, blinding patients and providers to the treatment does not match clinical practice. Moreover, this feature can reduce enrolment and retention and blinding can be undone by known or expected adverse effects, pill characteristics (overcoating, taste) or administrative/procedural issues, such as different treatment providers for different interventions or availability of laboratory test results that are differentially affected by treatments. A compromise that is sometimes used is randomized, open treatment with utilization of masked raters, but care needs to be taken to maintain the masking.

Treatment selection

The choice of treatment and control(s) will, of course, be heavily influenced by the basic question that the RCT is intended to address. The choice of active control and the target dose(s) of both the investigational medicine and the control agent are important. Estimates of therapeutic equivalence and comparative adverse effect profiles are affected by these choices. If a dose is too low, efficacy may be suboptimal, but if a dose is too high it might inflate the incidence of adverse effects. Titration schedules can also be important for some drugs as well as bioavailability issues related to food ingestion, or metabolic issues related to smoking, body weight, concomitant medications, etc. The side effect profiles of the experimental drug and comparator can also lead to functional unblinding and should be considered from that standpoint as well, or methods can be used to reduce the likelihood of such effects by using an ineffective low dose of the experimental drug as a pseudoplacebo, or separating the ratings of efficacy from those of tolerability, or using centralized raters who do not follow the same patient through a trial.

An important and potentially difficult issue is the extent to which and what kind of "rescue" medication should be made available to those individuals who might otherwise drop out of the trial due to lack of efficacy and need for further treatment. This possibility can complicate the assessment of the therapeutic agent. However, in some settings it is difficult to conduct a controlled trial without such a provision. As will be discussed subsequently, the possibility of treating all patients initially with active agents, identifying those with a clear early response and then enrolling only the latter subjects in a double-blind, placebo controlled discontinuation study could be a powerful strategy to detect a true drug effect while exposing a minimal number of patients to placebo.

Comedications

The permission, timing, and dosing of comedications also requires consideration. Comedications are useful to limit adverse effect burden and dropouts, but can obscure true treatment effects. Moreover, differential washout of comedications in treatment groups prior to randomization can create confounds, whereas overly limited use of comedications might limit the feasibility of the trial and not match clinical reality.

Placebo controls

Recent discussion regarding placebo controlled clinical trials in schizophrenia has largely focused on ethical issues. The World Medical Association´s Declaration of Helsinki[67] stipulates that "The use of placebo is acceptable in studies where no current proven intervention exists; or where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine efficacy or safety of an intervention and the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm. Extreme care must be taken to avoid abuse of this option." Many investigators and, very importantly, regulatory agencies, such as the Food and Drug Administration in the US and the European Medicines Agency, have taken the position that a valid evaluation of a treatment for schizophrenia (in terms of both efficacy and safety) is not possible without a placebo-controlled design, unless the goal is to demonstrate superiority of the experimental agent over existing treatments. As a result, every antipsychotic that has been approved

for the treatment of schizophrenia in either the US or Europe in the past 20 years has been assessed for acute efficacy in placebo-controlled clinical trials.

However, such designs have been challenged.[68-70] In addition, ethical committees in many settings are implementing stricter standards, making it increasingly difficult to conduct placebo controlled clinical trials in schizophrenia. Furthermore, high dropout rates have been reported in clinical trials utilizing placebo controls,[71] and there has also been a decrease in the drug effect observed in clinical trials comparing both experimental and approved antipsychotics with placebo.[72-74] There are a number of potential factors which contribute to these findings ranging from protocol design to patient selection and assessment procedures. Moreover, unexpectedly high placebo response is also seen in patients enrolled in augmentation studies who were supposed to have stable, unresponsive residual symptoms.[75] Taken together, all of these factors underscore the importance of carefully considering the benefits and risks of placebo controlled trials, evaluating alternative strategies to achieve needed goals in drug development and ensuring that when placebos are involved that trials are implemented and conducted in such a way as to not inflate or exaggerate the placebo response. It is also important to distinguish between different types of trials, since acute treatment and maintenance treatment trials, or studies of treatment resistant patients, etc. might provide varying challenges in this context.

Trial duration

Both feasibility and scientific considerations influence the length of a trial. Though the full therapeutic benefit of antipsychotics might not be seen for weeks or months, the greatest proportion of response occurs within the first few weeks,[57,58] although this pattern is somewhat less clear for first-episode patients.[76,77] Improvement in positive symptoms can even be seen in a matter of hours or days.[78] The potential use of placebo controls in short-term, acute treatment trials argues for as short a duration as possible, in that those patients who are assigned to placebo are more likely to experience further exacerbation or lack of response and, therefore, terminate prematurely. Those patients who respond to placebo will be more likely to remain in the study, thereby complicating statistical analysis, diminishing the drug placebo difference and reducing generalizability. In many stud-

ies, a significant drug effect can be seen after 1 or 2 weeks. However, if drugs with novel mechanisms are the focus of investigation then assumptions about time course of response might be less reliable. There is also a subgroup of patients who are slower to respond and if the ultimate goal is to compare the full therapeutic potential of alternative treatments, then a longer trial might be desirable. Issues related to trial duration in maintenance of effect/relapse prevention studies will be discussed subsequently.

Outcome and assessment measures

The selection of assessment measures and instruments will be largely driven by the choice(s) of the primary and secondary outcome measures as well as by feasibility and rater/patient burden. Often, too many scales are included in a clinical trial and some of the data are never analyzed or published. Attention to scale validity and reliability is also important, and in regulatory trials there is a particular emphasis on instruments, which have been demonstrated on a broad scale to have the desired characteristics. If a new scale is introduced, it is often recommended to have an existing and widely utilized scale for the same domain included as a reference point. As there is increasing emphasis on patient reported outcomes, however, there is also some concern as to the validity of such measures for those individuals who are lacking in insight or unable to reliably evaluate their own subjective and/or functional state. In the case of schizophrenia, informant information can also be extremely valuable. Patient-reported outcomes in some cases can be impeded by willful concealment or distrust of the interviewer or interview situation. Clearly, as broader outcome assessments are called for, measures of negative symptoms, cognitive function, social and vocational performance/quality of life, subjective well-being, family burden, etc should be considered.

In contrast to some dimensions of psychopathology, the longer time frame needed to assess the amelioration of negative symptoms and cognitive dysfunction or improvements in overall social and vocational adjustment might require trials of much longer duration. This is especially true if issues of persistence of effect are to be clarified. Finally, it has been recognized that adverse effects are not as carefully and comprehensively measured as efficacy measures.[79] This should also be reme-

diated by adding a brief interview based or self-administered adverse effect check list to spontaneous reporting.

Quality of ratings and fidelity of assessments

In order to have a sufficient signal-to-noise detection ratio for diagnostic, symptomatic and side effect assessments, both the utilized tools and the raters performing the interviews and ratings need to be highly reliable. In fact, the intraclass correlation coefficient (ICC) between ratings has a profound impact on sample size requirements. For example, an ICC of 0.9 requires 111 patients compared with 200 patients if the ICC is 0.5 in order to achieve the same statistical power.[80] The way that raters are trained and the manner in which reliability is established varies. In fact, true interrater reliability is rarely established in multicenter clinical trials. Specifically, having prospective interviewers only rate videotaped assessments performed by an expert does not establish the kind of reliability that is necessary. Even high ICCs with the expert rater do not in any way establish the ability of the rater to elicit the same symptoms when conducting an independent interview that he/she was able to rate when being fed the patient responses in an idealized training tape. Moreover, the method of rating even taped interviews is not usually standardized, so that it is not clear to what degree ratings occur completely independent in the classroom. In addition, a sufficient number of such assessments to establish statistical correlations is rarely done. Furthermore, even if reliability was established for both the interview and the rating, rater drift needs to be countered by reassessing the reliability of the ratings periodically throughout the trial, as well as training new raters when there is staff turnover. Other methods of increasing precision of ratings include comparing similar outcome dimensions across different assessment scales (ie, convergent validity) or checking rater-assessed outcomes against patient reported outcomes or against the evaluation of quality control by remote expert raters (ie, external consistency). In case of obvious inconsistencies, raters can then be approached and simply be given feedback or they can be retrained. However, even though expert raters can be used to check or adjudicate site based ratings, they have to rely on the interviews that may be less than optimal in obtaining a full clinical picture. Research has shown that many assessments were deficient when site based interviews were audiotaped

and randomly assessed by expert raters.[81] Another method, particularly for multisite studies that has shown considerable promise to increase the reliability of ratings and reduce placebo response,[82] includes the use of remote centralized expert raters who perform the assessments via live, two-way video. This method can be expensive and poses some logistical challenges, but is in keeping with the desire to centralize and standardize assessments whenever possible, as has increasingly been done with cardiology, pathology, radiology, and laboratory tests in multicenter trials.

Relapse prevention

Relapse prevention in schizophrenia remains a major public problem. However, the number of studies focusing on relapse prevention/maintenance treatment is substantially smaller compared with acute phase trials.[83,84] A particular challenge in relapse prevention studies is the definition of relapse, for which no universal criteria exist.[83] Given ethical concerns about placebo-controlled trials in relapse prevention, it has become customary to utilize relapse criteria which do not require a full-blown psychotic exacerbation, but rather rely on minimally clinically significant early signs of relapse. Subsequently, relapse rates might be higher than in studies conducted previously, and there are a number of potential false positives.

The use of placebo controls in relapse prevention studies is another source of controversy, and opinions of regulatory authorities also differ on this topic. Some would argue that the demonstration of non-inferiority in comparison to a proven efficacious compound should be sufficient. However, both dropout and response rates vary whether an active or placebo control is used,[85] and relapse rates vary enormously across trials. For example, a recent trial comparing depot and oral medications reported rehospitalization rates of 39% and 45%, respectively, in a 2-year study.[86] By contrast, other trials reported rehospitalization rates as low as 1.3% and 5.8% with depot and oral medications, respectively, at 1 year,[87] and 9.3% and 15.2%, respectively, at 2 years.[88] Therefore, it is difficult to be certain if one is dealing with an ineffective medication or with a patient population that is highly vulnerable to relapse regardless of medication status.

Another important issue that needs to be considered in the design of maintenance and relapse prevention studies is the timing of the randomization. In most tri-

als, patients are randomized in the acute treatment phase and then continued into an extension maintenance study. However, if patients are not rerandomized after stabilization, the concern is that by including randomly assigned, acutely exacerbated patients, only those patients at risk for relapse who had responded to and tolerated the specific acute treatment participate in the maintenance portion of the trial. This could lead to a selection bias for patients who experienced less side effects or experienced more improvement on the allocated medication. This concern is particularly relevant when there are unequal proportions of patients in each originally randomized group that enter the maintenance and relapse prevention phase of the study.

The degree to which patients entering the trial are stable and whether this is established retrospectively or prospectively are other important considerations. As for relapse, stability criteria and the required duration of stability or remission are insufficiently standardized. Another important issue is the duration of the trial. Since some long studies suggest different patterns of relapse during the first and second years,[89,86,90] a duration of 2 years or longer is ideal. But, of course, the longer the duration, the higher the dropout rate might be. The dropout rate varies from study to study, but some surpass 50%. This issue is clearly more important in active-active comparisons, as drug placebo differences can usually be seen in shorter time frames.

Finally, the determination of optimum dosing for maintenance/relapse prevention is a particular challenge, given the unpredictable time course of relapse. Even when patients are completely withdrawn from antipsychotics when in a state of remission or stability, the resulting relapse might not occur for weeks or months. Therefore, if a flexible dose is used, it is difficult to determine whether or not a relapse is due to an ineffective medication or if it is due to reducing the dose below an efficacy threshold for that patient. Therefore, fixed-dose studies are valuable in the maintenance phase to evaluate the dose response relationship, which might be quite different from that observed in acute efficacy trials where the goal is reducing acute and severe psychopathology. Importantly, given high potential rates of non adherence, the use of long-acting injectable medications can be very valuable in this context to ensure that nonadherence does not confound the interpretation of dose-response relationships.[91-92]

Statistical issues

Several issues of clinical trial design influence sample size estimates and the power to detect a clinically meaningful treatment effect, while maintaining a nominal level of type I error. For example, multiple outcomes can inflate type I error, and unreliable assessment processes and imprecise measurements can introduces biases and reduce statistical power.[93] In addition, missing data pose considerable challenges. It is increasingly recognized that last-observation-carried-forward (LOCF) analytic methods are problematic and that mixed models repeated measures (MMRM) analyses for continuous outcomes and Generalized Estimation Equation (GEE) models are a superior way of handling missing data. It took a while for regulatory agencies to agree to this, but nowadays MMRM analyses are also an acceptable analysis method for registration trials. However, there are really no good solutions for dealing with missing data that are almost never missing truly at random. Even methods like MMRM and GEE that adjust the analyses based on results from patients who continued in the trial have their limitations, as their validity is based on the assumption of ignorable attrition,[94] highlighting the importance of minimizing dropouts and missing data as much as possible. In fact, dropout rates have become an increasing problem like placebo response rates.[95] Thus, studies need to be designed in ways to minimize dropout rates, for example by not creating incentives for leaving the study early. Incentives for patients may include a rollover in an open long-term extension phase study where treatment is free, while incentives for investigators might include recruiting patients to a subsequent randomized study phase. In addition, comedications to minimize the severity of adverse effects and time limited rescue strategies for inefficacy that do not compromise the primary outcome should also be considered. Another proposed possibility is the measurement of the "intent to attend" the next study visit and to use this as a covariate to decrease the attrition bias.[94] In addition, identifying patients who are not likely to continue in the study or who find participation burdensome prior to their dropping out can help research personnel to proactively address barriers to trial completion. Finally, allowing for in-person, two-way video or telephone assessments in the patient's home should also be considered to reduce the amount of missing data.

State of the art

Trial implementation and conduct

One of the most neglected areas of clinical trials is trial management and oversight. As signal detection has become increasingly difficult and sample sizes have increased, the conduct and quality control of large-scale RCTs has become increasingly complex and difficult.[65] Subsequently, many companies have outsourced this important aspect of trial implementation and performance. As a result, there is the danger of a loss of control and diffusion or narrowing of responsibility, in that clinical research organizations are mostly in charge of assuring that increasingly tight time lines are kept and quota are met. The enormous time pressures can lead to a problematic disconnect between the desired quantity and the desired quality of enrolled patients and assessments. Moreover, increasing regulatory requirements can also lead to an overburdening of sites and investigators who are not part of professional clinical trials sites and who might drop out of multisite RCTs, thereby narrowing the settings in which patients are studied. Furthermore, the focus on assuring adherence to formal requirements, which has appropriately attracted scrutiny and attention, should not distract from assessing the quality of the trial conduct that is not equivalent to following checklists and increasingly complex documentation.

To overcome some of these problems, recent concern has focused on finding ways to encourage trial management organizations to broaden their responsibility beyond considerations of documentation and fulfilling quota. Rather, methods need to be considered that provide incentives to these organizations to assure adherence to appropriate standards of patient selection and high quality assessments, follow-up, protocol adherence, and retention. However, a high placebo response or lack of separation of the investigational drug or standard comparator from placebo cannot automatically or solely be used as a quality indicator. Moreover, quality adherence should also be measurable and achievable during the conduct of the trial and not only after its conclusion. Therefore, the field needs to develop standards against which sites and clinical trial management organizations can be assessed and which can provide clear guidance to all parties involved in the conduct of the trial.

Novel trial designs

In addition to addressing the issues and challenges previously discussed, there is also the need to explore novel trial designs which can help to facilitate signal detection and overall trial efficiency. There is a powerful clinical tool that uses the patients' own response pattern to predict outcomes. This intraindividual test of early response/nonresponse as a predictor of subsequent response[96,97] or the predictive value of dysphoric response[98] had been studied briefly in the 1980s. As much as 15 to 20 years later, these findings have been revisited and expanded upon, stimulated by analyses showing that, at least at a group level, the majority of antipsychotic response occurs within the first few weeks[57,58] and, even days[99] after antipsychotic initiation. Building on these findings, a series of post-hoc analyses[59,60,100-102] plus a recent prospective study[61] showed that nonresponse at study end point can be predicted with high sensitivity, specificity and predictive power by presence of less than a minimal response, equivalent to less than 20% reduction in the Positive and Negative Syndrome Scale[103] or Brief Psychiatric Rating Scale[104] total score at 2 weeks after antipsychotic initiation. However, having identified this general response pattern, questions remain as to whether such trajectories are similar in the more likely heterogeneous first-episode schizophrenia samples and in treatment-refractory patients.[76,77] In addition, it needs to be determined whether or not a limited set of specific symptom items that could be used in clinical practice are equally valid and reliable[105] and what one can learn from symptom trajectories at an individual patient level.[106-109]
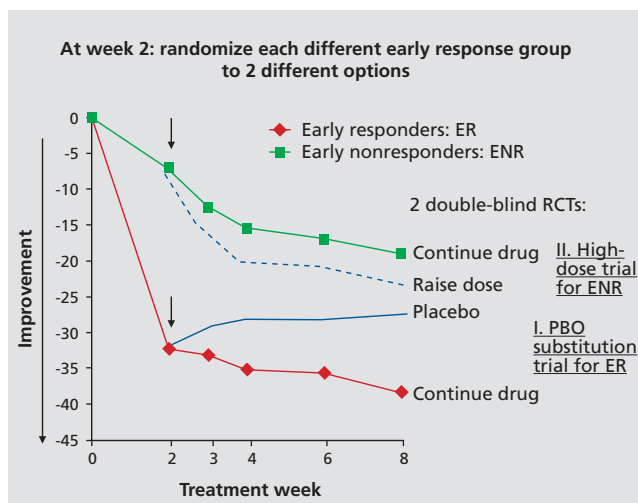

Figure 1. Novel drug development design utilizing early response/nonresponse for sample enrichment. ↓, randomization time point

This strategy would be very valuable in helping to determine what alternative treatments are likely to be more successful after early nonresponse has been identified.[61] A novel design to help enhance signal-to-noise ratio in an acute trial could take advantage of the response patterns that have been identified *(Figure 1)*. In the "early responder randomized discontinuation design" all patients are assigned to active drug, and then only those who had at least a minimal response at 2 weeks are enrolled in a double-blind, placebo-controlled discontinuation trial. This design could potentially enrich the placebo controlled portion of the trial with true drug responders and thereby expose fewer patients to placebo. A recent report by Marques et al[110] suggests that those patients with a robust early response are less likely to include placebo patients than other trajectories of response. Appropriate data should be collected to determine what proportion of early responders would show an exacerbation following placebo substitution and within what time-frame. The ethical implications of such a design should also be considered. Is an exacerbation on placebo substitution for a subgroup of patients more acceptable than a 4- to 6-week exposure to placebo for a larger number of patients? At the same time, early nonresponders at 2 weeks can also be used for rerandomization to test strategies focusing on enhancing response. In the mirror image "early nonresponder randomized dose increase or augmentation design," early nonresponders at 2 weeks are assigned to staying on the medication or going either to a higher dose or an augmentation agent. The dose increase or augmentation option will likely mostly be studied separately. However, including both options in a three-arm design would also be possible. This might be espcially attractive when studying the addition of a second antipsychotic as the augmentation strategy. Having the dose increase arm in this design would allow distinguishing the effect of non-dopaminergic receptor synergies vs mere increased antipaminergic "dose" increase.

In effect, this proposed design, "the early responder randomized discontinuation design" is an alternative to a previously proposed study design, "the sequential parallel comparison design."[111] In contrast to the design that we are proposing which has a 2-week active drug run-in phase, the sequential parallel comparison study consists of two phases of randomized treatment of equal duration. The first phase involves an unbalanced randomization between placebo and active treatment with over-

sampling of placebo randomization. The second phase involves re-randomization of placebo nonresponders to active treatment or placebo. As patients in the second phase "failed" placebo before, they are less likely to respond to placebo, which diminishes the placebo response and has the potential of enhancing power. However, at the same time, drug response rates are also likely to be reduced. The complication with this design is the proposed data analytic technique that does not only use patients from the second phase, ie, in an enhanced sample of placebo nonresponders. Rather, outcomes from both phases are combined in a complicated pooling ratio.[111] However, were only patients from phase two to be used, this would necessitate a very high number of patients to undergo the first phase.

Conceivable alternatives to this design include two phases of unequal duration with rerandomization of early placebo nonresponders (in schizophrenia: <20% reduction in PANSS total score) after only 2 weeks to either placebo or active treatment (ie, early placebo nonresponder sequential parallel comparison design). Alternatively, a triple-blind, 2-week placebo lead-in phase could precede randomization to drug or placebo in patients with <20% reduction in PANSS total score, rather than randomizing patients without a "full" response (however defined in a given study and disease) at 4 weeks or longer, as proposed in the sequential parallel comparison design.

As can be seen from these examples, the enrichment strategy of focusing on early responders or nonresponders to either drug or placebo could potentially be leveraged to enhance the signal-to-noise ratio, depending on the question that is being asked.

Summary and conclusions

Clinical trials, like everything else man-made, are imperfect. Their specific content and success are context dependent. A number of factors that were outlined in this article need to be considered, controlled, monitored, and improved upon. In addition to a number of standard features, the design of RCTs needs to be tailored to the research question, population, illness phase, setting, active treatment, control condition and outcome under investigation. Patient selection, blinding, ratings, study/site management and adherence are important aspects. Innovative designs should be considered in order to deal with some of the inevitable

State of the art

compromises involved in designing and conducting RCTs. For some research questions, alternative study types might need to be considered, such as cohort, pharmacoepidemiologic database or registry studies. Importantly, measurable quality standards for RCTs need to be developed. Applying these standards along with novel ways to incentivize all of the parties involved in order to achieve increased adherence to quality measures need to be explored. To achieve this, the different stakeholders should share experiences and actual data to come up with appropriate solutions. We need to learn from the past as much as possible and we need to appreciate that failed and uninformative trials, increasing placebo response rates and increased sample size requirements in the context of decreasing effect sizes are a critical and destructive, but shared problem that needs viable solutions. Without this shared responsibility for the design and conduct of high quality trials, the development of new compounds and the broadening of indications for patients in strong need of effective and safe treatment alternatives will become increasingly difficult. In addition, more and more companies will be discouraged from pursuing these therapeutic targets for drug development. Finally, the utility of novel trial designs that decrease placebo response and enrich samples should be tested and their appropriateness for regulatory approval pathways needs to be explored. ❑

REFERENCES

1. World Health Organization. *International Statistical Classification of Diseases, Injuries, and Causes of Death. Sixth Revision of the International Lists of Diseases and Causes of Death.* Geneva, Switzerland: WHO; 1949.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* Washington, DC: American Psychiatric Association; 1952.
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 3rd ed. Washington, DC: American Psychiatric Association; 1980
4. Klein DF, Davis JM. *Diagnosis and Drug Treatment of Psychiatric Disorders.* Baltimore, MD: The Williams & Wilkins Company; 1969.
5. Griffith RW, Saameli K. Clozapine and agranulocytosis. *Lancet.* 1975;2:657.
6. Kane, J. M, Honigfeld, G, Singer, J, et al. Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch Gen Psychiatry.* 1988;45:789-796.
7. Essali A, Al-Haj Haasan N, Li C, Rathbone J. Clozapine versus typical neuroleptic medication for schizophrenia. *Cochrane Database Syst Rev.* 2009;1:CD000059.
8. Meltzer HY, Alphs L, Green AI, et al; International Suicide Prevention Trial Study Group. Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch Gen Psychiatry.* 2003;60:82-91. Erratum in: *Arch Gen Psychiatry.* 2003;60:735.
9. Leucht S, Komossa K, Rummel-Kluge C, et al. A meta-analysis of head-to-head comparisons of second-generation antipsychotics in the treatment of schizophrenia. *Am J Psychiatry.* 2009;166:152-163.
10. Asenjo Lobos C, Komossa K, Rummel-Kluge C, et al. Clozapine versus other atypical antipsychotics for schizophrenia. *Cochrane Database Syst Rev.* 2010;11:CD006633.

11. Kane JM, Quitkin F, Rifkin A, Ramos-Lorenzi JR, Nayak DV. Fluphenazine versus placebo in patients with remitted acute first episode schizophrenia. *Arch Gen Psychiatry.* 1982;39:70-73.
12. Robinson DG, Woerner MG, Alvir JM, et al. Predictors of treatment response from a first episode of schizophrenia or schizoaffective disorder. *Am J Psychiatry.* 1999;156:544-549.
13. Melle I, Larsen TK, Haahr U, Friis S, et al. Prevention of negative symptom psychopathologies in first-episode schizophrenia: two-year effects of reducing the duration of untreated psychosis. *Arch Gen Psychiatry.* 2008;65:634-640.
14. Bilder RM, Goldman RS, Robinson D et al. Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *Am J Psychiatry.* 2000;157:549-559.
15. Loebel AD, Lieberman JA, Alvir JM, Mayerhoff DI, Geisler SH, Szymanski SR. Duration of psychosis and outcome in first-episode schizophrenia. *Am J Psychiatry.* 1992;149:1183-1188.
16. Robinson DG, Woerner MG, Delman HM, Kane JM. Pharmacological treatments for first-episode schizophrenia. *Schizophr Bull.* 2005;31:705-722.
17. Robinson DG, Woerner MG, Alvir JM, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Arch Gen Psychiatry.* 1999;56:241-247.
18. Correll CU, Kratochvil CJ, March J. Developments in pediatric psychopharmacology: focus on stimulants, antidepressants and antipsychotics. *J Clin Psychiatry.* In press.
19. Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. *J Clin Psychiatry.* 2010;71:1115-1124.
20. Rosenheck R, Perlick D, Bingham S, et al; Department of Veterans Affairs Cooperative Study Group on the Cost-Effectiveness of Olanzapine. Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. *JAMA.* 2003;26;290:2693-2702.

Ensayos controlados randomizados en esquizofrenia: oportunidades, limitaciones y alternativas de diseño de ensayos

El estado del arte en el diseño y la metodología de los ensayos clínicos es de gran importancia para el avance en este campo. A la inversa, la importancia crítica del manejo e implementación del ensayo sólo ha sido recientemente el foco de discusión e investigación. Aunque los ensayos controlados randomizados son considerados generalmente el gold standard para la evaluación de las intervenciones farmacológicas y no farmacológicas en medicina, los ensayos son sensibles a las complicaciones e influencias que pueden comprometer seriamente su éxito. Al igual que las intervenciones, el diseño y manejo de los ensayos también dependen del contexto. Ellos necesitan estar individualizados y adaptados a un número de variables relevantes, tales como el lugar, la población, la fase de la enfermedad, las intervenciones, las expectativas y sesgos del paciente y del evaluador, y los objetivos generales de la investigación. Dado que no existe un enfoque uniforme factible, se deben considerar cuidadosamente ciertos principios y guías generales. El conocimiento de soluciones y alternativas básicas, y el reconocimiento de los complejos desafíos que necesitan emprenderse proactivamente pueden ayudar a minimizar los resultados no deseados, incluyendo la falla del ensayo y los resultados mal informados o falsos negativos. Sin embargo, las alternativas de un nuevo diseño necesitan ser exploradas con el aporte de una muestra objetivo de acuerdo a la pregunta del estudio y con un aumento de la precisión en la medición de resultados relevantes. Se proponen dos estrategias de un nuevo diseño que aprovechan el paradigma de respuesta de antipsicóticos recientemente validado (que también se ha aplicado a antidepresivos y estabilizadores del ánimo). En el "diseño de discontinuación randomizado del respondedor precoz" todos los pacientes son asignados al fármaco activo, y sólo los que tienen al menos una mínima respuesta a las dos semanas se incorporan a un ensayo de discontinuación doble-ciego, placebo-controlado, con lo que aumenta la porción del ensayo controlada por placebo con los verdaderos respondedores al fármaco. En la imagen especular del "diseño de aumentación o aumento randomizado de dosis de los no respondedores precoces", los no respondedores precoces a las dos semanas son asignados a mantener la medicación o bien a una dosis mayor o a un agente potenciador. Es de esperar que mediante una mayor atención a los temas planteados en este artículo y a un mayor refinamiento en la metodología y el manejo de los ensayos, esta área aporte progresos adicionales en la prevención y tratamiento de la esquizofrenia.

21. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med.* 2005;353:1209-1223.
22. Jones PB, Barnes TR, Davies L, et al. Randomized controlled trial of the effect on Quality of Life of second vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Arch Gen Psychiatry.* 2006;63:1079-1087.
23. Leucht S, Komossa K, Rummel-Kluge C, et al. A meta-analysis of head-to-head comparisons of second-generation antipsychotics in the treatment of schizophrenia. *Am J Psychiatry.* 2009;166:152-163.
24. McEvoy JP, Lieberman JA, Stroup TS, et al; CATIE Investigators. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypicalantipsychotic treatment. *Am J Psychiatry.* 2006;163:600-610.
25. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry.* 2006;163:185-194.
26. Newcomer JW. Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review. *CNS Drugs.* 2005;19(suppl 1):1-93.
27. Fleischhacker WW, Cetkovich-Bakmas M, De Hert M, et al. Comorbid somatic illnesses in patients with severe mental disorders: clinical, policy, and research challenges. *J Clin Psychiatry.* 2008;69:514-519.

28. Leucht S, Burkard T, Henderson J, Maj M, Sartorius N. Physical illness and schizophrenia: a review of the literature. *Acta Psychiatr Scand.* 2007;116:317-333.
29. Colton CW, Manderscheid RW. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Prev Chronic Dis.* 2006;3:A42.
30. De Hert M, Correll CU, Bobes J, et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry.* 2011;10:1052-1077.
31. Ray WA, Chung CP, Murray KT, Hall K, Stein CM. Atypical antipsychotic drugs and the risk of sudden cardiac death. *N Engl J Med.* 2009;360:225-35. Erratum in: *N Engl J Med.* 2009;361:1814.
32. Correll CU, Nielsen. J Antipsychotic-associated all-cause and cardiac mortality: what should we worry about and how should the risk be assessed? *Acta Psychiatr Scand.* 2010;122:341-344.
33. Schneider LS, Dagerman KS, Insel P. Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. *JAMA.* 2005;294:1934-1943.
34. Smith M, Hopkins D, Peveler RC, Holt RI, Woodward M, Ismail K. First. vs. second-generation antipsychotics and risk for diabetes in schizophrenia: systematic review and meta-analysis. *Br J Psychiatry.* 2008;192:406-411.
35. Nielsen J, Skadhede S, Correll CU. Antipsychotics associated with the development of type 2 diabetes in antipsychotic-naïve schizophrenia patients. *Neuropsychopharmacology.* 2010;35:1997-2004.

Études contrôlées randomisées dans la schizophrénie : opportunités, limites et schémas d'étude alternatifs

La méthodologie et le schéma des études cliniques les plus récentes sont extrêmement importants pour les progrès dans ce domaine. À l'inverse, la pertinence de la mise en œuvre et du déroulement de l'étude n'a été que plus récemment le centre des discussions et des recherches. Bien que les études contrôlées randomisées soient généralement considérées comme la méthode de référence pour l'évaluation des traitements pharmacologiques et non pharmacologiques en médecine, elles sont soumises à des complications et à des biais qui peuvent gravement compromettre leur succès. Comme toute intervention, la conduite et le schéma des études, dépendent du contexte. Ils doivent être individualisés et adaptés à un certain nombre de variables pertinentes, comme l'environnement, la population, la phase de la maladie, les traitements, les biais et les attentes des patients et des évaluateurs ainsi que les objectifs globaux de la recherche. Même si une approche unifiée est impossible, certains principes et recommandations généraux demandent un examen attentif. La connaissance des solutions de base et de leurs alternatives, et celle des défis complexes nécessitant d'être pris en charge préventivement, peuvent aider à minimiser des résultats non désirés, comme l'échec de l'étude et l'obtention de résultats sans intérêts ou faussement négatifs. De plus, il faut étudier de nouvelles alternatives de schémas d'études afin d'enrichir les échantillons en fonction des questions de l'étude et améliorer la précision de mesure de résultats pertinents. Nous proposons deux nouvelles stratégies de schémas d'études qui profitent du modèle de réponse antipsychotique précoce récemment validé (qui a aussi été observé avec des antidépresseurs et des régulateurs de l'humeur). Dans le schéma de « répondeurs précoces randomisés pour l'arrêt de traitement » tous les patients reçoivent un médicament actif, et seuls ceux qui avaient au moins une réponse minimale à 2 semaines sont inclus pour une étude d'interruption en double aveugle, contrôlée contre placebo, ce qui permet d'enrichir la partie de l'étude contrôlée contre placebo avec de vrais répondeurs au traitement. À l'inverse,« les non-répondeurs précoces randomisés pour escalade de dose ou coprescription», doivent à 2 semaines continuer leur traitement et soit augmenter leur dose soit bénéficier d'une prescription complémentaire. Nous espérons que grâce à l'intérêt accru pour les problèmes soulevés dans cet article et à la précision plus rigoureuse dans le déroulement et la méthodologie des études, des progrès supplémentaires bien nécessaires verront le jour dans le domaine, pour la prévention et le traitement de la schizophrénie.

36. Correll CU, Manu P, Olshanskiy V et al. Cardiometabolic risk of second-generation antipsychotic medications during first-time use in children and adolescents. *JAMA*. 2009;302:1765-1773.

37. Andreasen NC, Carpenter WT Jr, Kane JM, Lasser RA, Marder SR, Weinberger DR. Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry*. 2005;162:441-449.

38. Kane JM, Leucht S. Unanswered questions in schizophrenia clinical trials. *Schizophr Bull*. 2008;34:302-309.

39. Leucht S, Heres S, Hamann J, Kane JM. Methodological issues in current antipsychotic drug trials. *Schizophr Bull*. 2008;34:275-285.

40. Liberman RP, Kopelowicz A, Ventura J, Gutkind D. Operational criteria and factors related to recovery from schizophrenia. *Int Rev Psychiatry*. 2002;14:256-272.

41. Naber D, Karow A, Lambert M. Subjective well-being under the neuroleptic treatment and its relevance for compliance. *Acta Psychiatr Scand Suppl*. 2005;427:29-34.

42. Lambert M, Schimmelmann BG, Naber D, et al. Prediction of remission as a combination of symptomatic and functional remission and adequate subjective well-being in 2960 patients with schizophrenia. *J Clin Psychiatry*. 2006;67:1690-1697.

43. Green MF, Penn DL, Bentall R, Carpenter WT, et al. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophr Bull*. 2008;34:1211-1220.

44. Harvey PD, Green MF, Bowie C, Loebel A. The dimensions of clinical and cognitive change in schizophrenia: evidence for independence of improvements. *Psychopharmacology (Berl)*. 2006;187:356-363.

45. Keefe RS, Vinogradov S, Medalia A, et al. Report From the Working Group Conference on Multisite Trial Design for Cognitive Remediation in Schizophrenia. *Schizophr Bull*. 2010. [Epub ahead of print].

46. Lehman AF, Goldberg R, Dixon LB, et al. Improving employment outcomes for persons with severe mental illnesses. *Arch Gen Psychiatry*. 2002;59:165-172.

47. McGurk SR, Schiano D, Mueser KT, Wolfe R. Implementation of the thinking skills for work program in a psychosocial clubhouse. *Psychiatr Rehabil J*. 2010;33:190-199.

48. Evans JD, Bond GR, Meyer PS, et al. Cognitive and clinical predictors of success in vocational rehabilitation in schizophrenia. *Schizophr Res*. 2004;70:331-342.

49. Correll CU, Cañas F, Larmo I, et al. Individualizing antipsychotic treatment selection in schizophrenia: characteristics of empirically derived patient subgroups. *Eur Psychiatry*. 2011;26 (suppl 1):3-16.

50. Falkai P, Mike O, Inez MG, Paul H, Andras BG, Sophia F; MEOS Consortium. A roadmap to disentangle the molecular etiology of schizophrenia. *Eur Psychiatry*. 2008;23:224-232.

51. Lathia CD, Amakye D, Dai W, et al. The value, qualification, and regulatory use of surrogate end points in drug development. *Clin Pharmacol Ther*. 2009;86:32-43.

52. Agid O, Foussias G, Singh S, Remington G. Where to position clozapine: re-examining the evidence. *Can J Psychiatry*. 2010;55:677-684.

53. Agid O, Arenovich T, Sajeev G, et al. An algorithm-based approach to first-episode schizophrenia: response rates over 3 prospective antipsychotic trials with a retrospective data analysis. *J Clin Psychiatry*. 2011. [Epub ahead of print] .

54. Robinson D, Woerner MG, Pollack S, Lerner G. Subject selection biases in clinical trials: data from a multicenter schizophrenia treatment study. *J Clin Psychopharmacol*. 1996;16:170-176.

55. Hofer A, Hummer M, Huber R, Kurz M, Walch T, Fleischhacker WW. Selection bias in clinical trials with antipsychotics. *J Clin Psychopharmacol*. 2000;20:699-702.

56. Salimi K, Jarskog LF, Lieberman JA. Antipsychotic drugs for first-episode schizophrenia: a comparative review. *CNS Drugs*. 2009;23:837-855.

57. Agid O, Kapur S, Arenovich T, Zipursky RB. Delayed-onset hypothesis of antipsychotic action: a hypothesis tested and rejected. *Arch Gen Psychiatry*. 2003;60:1228-1235.

58. Leucht S, Busch R, Hamann J, Kissling W, Kane JM. Early-onset hypothesis of antipsychotic drug action: a hypothesis tested, confirmed and extended. *Biol Psychiatry*. 2005;57:1543-1549.

59. Leucht S, Busch R, Kissling W, Kane JM. Early prediction of antipsychotic non-response. *J Clin Psychiatry*. 2007;68:352-360.

60. Kinon BJ, Chen L, Ascher-Svanum H, et al. Predicting response to atypical antipsychotics based on early response in the treatment of schizophrenia. *Schizophr Res*. 2008;102:230-240.

61. Kinon BJ, C, Ascher-Svanum H, Stauffer VL, et al. Early response to antipsychotic drug therapy as a clinical marker of subsequent response in the treatment of schizophrenia. *Neuropsychopharmacol*. 2010;35:581-590.

62. Correll CU. From receptor pharmacology to improved outcomes: individualizing the selection, dosing, and switching of antipsychotics. *Eur Psychiatry*. 2010;25 Suppl 2:S12-S21.

63. Meltzer HY. Suicidality in schizophrenia: a review of the evidence for risk factors and treatment options. *Curr Psychiatry Rep*. 2002;4:279-283.

64. Lavori PW, Rush AJ, Wisniewski SR, et al Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol Psychiatry*. 2001;50:792-801.

65. Marks DM, J T, Pae CU. Innovations in clinical research design and conduct in psychiatry: shifting to pragmatic approaches. *Psychiatry Investig*. 2009;6:1-6.

66. Kane JM, Rifkin A, Quitkin F, et al. Low dose fluphenazine decanoate in maintenance treatment of schizophrenia. *Psychiatry Res*. 1979; 1:341-348.

67. World Medical Association (WMA). Declaration of Helsinki-Ethical principles for medical research involving human subjects. Adopted by the 18Th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the 59th WMA General Assembly, Seoul, October 2008.

68. Fleischhacker WW. Burns T. European Group For Research In Schizophrenia. Feasibility of placebo-controlled clinical trials of antipsychotic compounds in Europe. *Psychopharmacology*. 2002;162:82-84.

69. Hummer M, Holzmeister R, Kemmler G, et al. Attitudes of patients with schizophrenia toward placebo-controlled clinical trials. *J Clin Psychiatry*. 2003;64:277-281.

70. Roberts LW. The ethical basis of psychiatric research: conceptual issues and empirical findings. *Comp Psychiatry*. 39:99-110.

71. Kemmler G, Hummer M, Widschwendter C, Fleischhacker WW. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Arch Gen Psychiatry*. 2005;62:1305-1312.

72. Kemp AS, Schooler NR, Kalali AH, et al. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? *Schizophr Bull*. 2010;36:504-509.

73. Loebel A, Cucchiaro J, Siu C, Daniel D, et al. Signal detection in clinical trials: A post-study survey of schizophrenia trial sites. Paper presented at: ISCTM 2010 Autumn Conference. 13-14 October. Baltimore, MD.

74. Chen Y-F, Wang S-J, Khin NA, Hung MJ, et al. Trial design issues and treatment effect modeling in multi-regional schizophrenia trials. *Pharmaceutical Statistics*. 2010;9, 217-229.

75. Kane JM, Correll CU, Goff DC, et al. A multicenter, randomized, double-blind, placebo-controlled, 16-week study of adjunctive aripiprazole for schizophrenia or schizoaffective disorder inadequately treated with quetiapine or risperidone monotherapy. *J Clin Psychiatry*. 2009;70:1348-1357.

76. Emsley R, Rabinowitz J, Medori R. Time course for antipsychotic treatment response in first-episode schizophrenia. *Am J Psychiatry*. 2006;163:743-745.

77. Gallego JA, Robinson DG, Sevy SM, et al. Time to treatment response in first episode schizophrenia: must acute treatment trials last several months? *J Clin Psychopharmacol*. In press.

78. Agid O, Kapur S, Warrington L, Loebel A, Siu C. Early onset of antipsychotic response in the treatment of acutely agitated patients with psychotic disorders. *Schizophr Res*. 2008;102:241-248.

79. Papanikolaou PN, Churchill R, Wahlbeck K, Ioannidis JP. Safety reporting in randomized trials of mental health interventions. *Am J Psychiatry*. 2004;161:1692-1697.

80. Kobak KA, Brown B, Sharp I, Levy-Mack H, Wells K, Ockun F, Williams JB. Sources of unreliability in depression ratings. *J Clin Psychopharmacol*. 200929:82-5. Erratum in: *J Clin Psychopharmacol*. 2009;29:221.

81. Kobak KA. Feiger AD. Lipsitz JD. Interview quality and signal detection in clinical trials. *Am J Psychiatry*. 2005;162:628.

82. Kobak KA, Leuchter A, DeBrota D, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J Clin Psychopharmacology*. 2010;30:193-197.

83. Leucht S, Barnes TR, Kissling W, Engel RR, Correll C. Kane JM. Relapse prevention in schizophrenia with new-generation antipsychotics: a systematic review and exploratory meta-analysis of randomized, controlled trials. *Am J Psychiatry*. 2003;160:1209-1222.

84. Leucht C, Heres S, Kane JM, Kissling W, Davis JM, Leucht S. Oral versus depot antipsychotic drugs for schizophrenia--a critical systematic review and meta-analysis of randomised long-term trials. *Schizophr Res*. 2011;127:83-92.

85. Woods SW, Gueorguieva RV, Baker CB, Makuch RW. Control group bias in randomized atypical antipsychotic medication trials for schizophrenia. *Arch Gen Psychiatry*. 2005;62:961-970.

86. Rosenheck RA, Krystal JH, Lew R, et al for the CSP555 Research Group. Long-acting risperidone and oral antipsychotics in unstable schizophrenia. *N Engl J Med*. 2011;364:842-51.

87. Li, R, Zhang, M, Shi, S,. The effect of haloperidol decanoate on negative and positive symptoms of schizophrenia. *Chin J Nerv Ment Diseases*. 1996;22:9–12.

88. Gaebel W, Schreiner A, Bergmans P, et al. Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable vs quetiapine: results of a long-term, open-label, randomized clinical trial. *Neuropsychopharmacology*. 2010;35:2367-77.

89. Csernansky JG, Mahmoud R, Brenner R. A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *N Engl J Med*. 2002; 346:16-22.

90. Schooler N, Rabinowitz J, Davidson M, et al. Risperidone and haloperidol in first episode psychosis: a long-term randomized trial. *Am J Psychiatry*. 2005;162:947-953.

91. Kane JM, Rifkin A, Woerner M, et al. Low-dose neuroleptic treatment of outpatient schizophrenics: I. Preliminary results for relapse rates. *Arch Gen Psychiatry*. 1983;40:893-896.

92. Kane JM, Davis JM, Schooler N, et al. A multidose study of haloperidol decanoate in the maintenance treatment of schizophrenia. *Am J Psychiatry*. 2002;159:554-560.

93. Leon AC. Implications of clinical trial design on sample size requirements. *Schizophr Bull*. 2008;34:664-669.

94. Leon AC, Demirtas H, Hedeker D. Bias reduction with an adjustment for participants' intent to dropout of a randomized controlled clinical trial. *Clin Trials*. 2007;4:540-547.

95. Wahlbeck K, Tuunainen A, Ahokas A, Leucht S. Dropout rates in randomized antipsychotic drug trials. *Psychopharmacology (Berl)*. 2001;155:230-233.

96. Nedopil N, Pflieger R, Ruther E. The prediction of acute response, remission and general outcome of neuroleptic treatment in acute schizophrenic patients. *Pharmacopsychiatria*. 1983;16:201–205.

97. Bartko G, Herczeg I, Bekesy M. Predicting outcome of neuroleptic treatment on the basis of subjective response and early clinical improvement. *J Clin Psychiatry*. 1987;48:363–365.

98. Awad AG, Voruganti LN. Neuroleptic dysphoria: revisiting the concept 50 years later. *Acta Psychiatr Scand*. Suppl. 2005;427:6-13.

99. Kapur S, Arenovich T, Agid O, Zipursky R, Lindborg S, Jones B. Evidence for onset of antipsychotic effects within the first 24 hours of treatment. *Am J Psychiatry*. 2005;162:939-946.

100. Correll CU, Malhotra AK, Kaushik S, McMeniman M, Kane JM. Early prediction of antipsychotic response in schizophrenia. *Am J Psychiatry*. 2003160:2063-5. Erratum in: *Am J Psychiatry*. 2005;162:1774.

101. Chang YC, Lane HY, Yang KH, Huang CL. Optimizing early prediction for antipsychotic response in schizophrenia. *J Clin Psychopharmacol*. 2006;26:554-559.

102. Ascher-Svanum H, Nyhuis AW, Faries DE, Kinon BJ, Baker RW, Shekhar A. Clinical, functional, and economic ramifications of early non-response to antipsychotics in the naturalistic treatment of schizophrenia. *Schizophr Bull*. 2008;34:1163-1171.

103. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261-276.

104. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep.* 1962;10:799–812.

105. Ruberg SJ, Chen LL, Stauffer V, et al. Identification of early changes in specific symptoms that predict longer-term response to atypical antipsychotics in the treatment of patients with schizophrenia. *BMC Psychiatry*. 2011;11:23.

106. Stauffer VL. Case M. Kinon BJ, et al. Early response to antipsychotic therapy as a clinical marker of subsequent response in the treatment of patients with first-episode psychosis. *Psychiatry Res*. 2011;187:42-48.

107. Levine SZ, Rabinowitz J. Trajectories and antecedents of treatment response over time in early-episode psychosis. *Schizophr Bull*. 201036:624-632.

108. Levine SZ. Leucht S. Elaboration on the early-onset hypothesis of antipsychotic drug action: treatment response trajectories. *Biol Psychiatry*. 2010;68:86-92.

109. Levine SZ, Rabinowitz J, Case M, Ascher-Svanum H. Treatment response trajectories and their antecedents in recent-onset psychosis: a 2-year prospective study. *J Clin Psychopharmacol*. 2010;30:446-449.

110. Marques TR, Arenovich T, Agid O, et al. The different trajectories of antipsychotic response: antipsychotics versus placebo. *Psychol Med*. 2010. [Epub ahead of print].

111. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the place-bo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom*. 2003;72:115-127.