

ISSUES IN CLINICAL TRIAL DESIGNS

JOHN M. KANE

The introduction of the randomized, double-blind, clinical trial was one of the major advances in the development of medical science. In the arena of psychotropic drug development this approach has proven to be of enormous value in advancing a field in which laboratory tests and strictly objective methods for diagnosis and outcome assessment are not currently available.

Designing trials in the treatment of schizophrenia highlights some of these challenges. Schizophrenia is a complex illness affecting to varying degrees a range of functions, including cognition, affect, behavior, mood, and motivation. The fact that this disorder affects so many different domains, varying from individual to individual, and to some extent within individuals over time, makes development of pharmacologic treatments even more challenging. Although there are core features of schizophrenia that involve perception (hallucinations), cognition (attention, working memory, etc.), motivation (avolition), inferential reasoning (delusions), and affect (blunted or inappropriate), there is no pathognomonic sign or symptom of the disease. This has important implications for the diagnostic process, which is also complicated by the fact that the evaluation of some core features (e.g., hallucinations and delusions) relies solely on subjective reporting, the accuracy of which is potentially influenced by the very symptoms themselves as well as by other social situational and personality variables.

In addition, the fact that such an array of domains and functions is disturbed in this illness creates a challenge for drug development. The tendency has been to conduct an array of assessments to evaluate drug effects in a number of domains concurrently, when in fact different domains may require different study designs, patient selection criteria, durations of treatment, etc. In the future, more attention will be given to those issues, and it is possible that multiple treatments will be studied rather with the goal of finding combinations able to improve outcome across a variety of domains. It is hoped that new treatments will be developed

with a focus on specific domains such as negative symptoms and cognitive dysfunction. Although a better understanding of basic mechanisms should facilitate further treatment advances, our current knowledge of pathophysiology remains limited. Advances in imaging techniques and pharmacogenomics are also important potential developments on the horizon that could have enormous impact on drug development and clinical evaluation.

Each area of psychotropic drug development has its own challenges in terms of rates of spontaneous remission, placebo response, patterns of relapse, domains of assessment, etc., but, in general, challenges of design and methodology involve issues that cut across the diagnostic domains.

DESIGN ISSUES

There are a number of critical issues in general design that need to be addressed in both the individual study as well as the particular program of drug development. A drug development program needs to be comprehensive as well as adaptive so that early results can inform subsequent evaluation. Although even when a drug is marketed there are still limitations in the amount of knowledge available to clinicians, several fundamental questions should have been at least partially addressed: (a) What benefits are likely to result from the drug? (b) What are its risks? (c) What dosage is indicated? (d) How does the new drug compare to alternative treatments? (e) Are there specific patients most likely to benefit from the drug?

There are a number of specific concerns that should be addressed when designing clinical trials of psychotropic drugs. Some of the most salient issues include dose finding; efficacy vs. placebo; efficacy vs. a standard reference compound; acute and long-term adverse effects; continuation and maintenance treatment efficacy; and relative efficacy or adverse effects in specific subgroups (e.g., early-phase illness, late-phase illness, refractory patients).

Dose-finding tolerability studies involving antipsychotic medications generally call for involvement of target patient populations earlier in the process than with other classes of drugs because it is difficult to ethically justify administering

J. M. Kane: Department of Psychiatry, Hillside Hospital, Glen Oaks, New York; Department of Psychiatry and Neuroscience, Albert Einstein College of Medicine, Bronx, New York.

these drugs to healthy volunteers for more than a week or two, and patterns of tolerance may be quite different in patients versus healthy volunteers.

It is not always possible to accurately predict clinical dosage requirements from preclinical studies; therefore, it is important to establish a full range of tolerable dosages in order to provide an appropriate range for efficacy studies. Drug development programs have been delayed and at times abandoned because of inadequate dose-finding efforts in the early stages of development (1). In addition, it is not unusual for dosage recommendations to change after a drug is marketed.

It is also important to have sufficient data on absorption, elimination, metabolism, and drug–drug interactions, to inform trial design.

Treatment trials generally fall into three broad categories: acute, continuation, and maintenance (or relapse prevention). Sometimes attempts are made to study two or even three phases in the same trial, but controversy surrounds the need to rerandomize patients before drawing conclusions about relative efficacy in maintenance-phase treatment. Patient characteristics may vary somewhat in terms of desirability within specific trials, but overall the following issues should be considered.

Patient Characteristics

It is important to be clear on whether or not patients are in a state of acute relapse or exacerbation as opposed to partial remission or a “stable plateau” of chronic symptomatology. At times investigators will withdraw patients from ongoing treatment in order to transition them to a clinical trial, resulting in some symptom exacerbation. The importance of these different approaches is that they may result in patients with very different degrees of drug responsiveness, different patterns of baseline symptomatology, and varying degrees of “stability” in baseline symptomatology.

The ideal sample of patients is probably those who have not already been partially treated so that the full degree and time course of response can be determined. However, given the way that subjects must be ascertained and recruited for trials, it is likely that some treatment will have already been administered. The fact that participants have been partially treated or are in a chronic symptomatic state does not necessarily preclude the detection of a subsequent, clinically significant drug effect, but it is likely that the nature and magnitude of the effect will be altered.

The subjectivity of many components of symptomatology in psychiatric disorders creates special challenges. Given the fact that many symptoms are subjective and cannot be confirmed or quantified using objective measures, the assessment of baseline status can be difficult. Clinicians are particularly familiar with patients suffering from psychoses who are more open and explicit about pretreatment psychopathology once they begin to improve. Some patients may

not appear eligible for a trial or be willing or able to give informed consent until they are partially treated.

A variety of subject characteristics should be considered in terms of inclusion and exclusion criteria. Specific decisions will be influenced by the nature and goals of the particular trial.

Age is often a basis for exclusion (either too young or too old). Age can certainly affect pharmacokinetics of particular drugs. The elderly are more likely to have comorbid medical conditions and be more sensitive to some adverse effects, and there are a variety of issues when young patients are included in trials. These and other factors have led to a paucity of subjects at the extreme age ranges in clinical trials. However, there has recently been increased recognition of the need for more early data on diverse age groups, and mechanisms are being implemented to encourage their inclusion in clinical trials.

Gender can be an important variable, and women are often underrepresented in clinical trials.

Ethnicity may have implications for drug metabolism and tolerance. In addition, as pharmacogenomic strategies are developed to extend clinical trial data, more accurate documentation of race will be critical.

Marital status can be a proxy for psychosocial adjustment and illness course, and may therefore be of prognostic significance.

Weight and body mass index have become an increasing concern from a public health standpoint and because of the considerable weight gain observed with some psychotropic drugs and in particular several new-generation antipsychotic medications (2).

Diagnostic subtype can be important in helping to characterize those patients most likely to benefit from specific treatments. Duration of illness and the duration of the current episode can be important in helping to define populations in terms of drug responsiveness as well as long-term course and outcome. A particular problem in many trials is categorizing patients' histories in terms of drug responsiveness. A current episode duration of more than 2 or 3 weeks could suggest that the patient is poorly or only partially responsive to the treatments that have already been administered, or, alternatively that some other factor is complicating treatment response (e.g., noncompliance, comorbid conditions, etc.). It can often be difficult to time the onset of illness or of a specific episode. As putative novel compounds are developed, it may become increasingly important to test these agents in patients who have not already been chronically exposed to other medications.

The specific type and severity of signs and symptoms required for entry into a trial will vary depending on the overall goals. Usually a minimal threshold of severity is established for core symptoms of interest. It is hoped that studies will also focus on patients selected on the basis of significant residual or secondary symptoms if they are associated with subjective distress and/or functional impairment.

If trials are designed to focus specifically on patients who were nonresponders or intolerant to other treatments, explicit criteria should be developed to identify such groups. There is debate as to whether or not a prospective trial is necessary to confirm treatment refractoriness, but this is certainly the most conservative approach because it also addresses to some extent the potential change in treatment milieu and attention resulting from participation in a research trial. In addition, there is enormous variability in the quality of retrospective assessment of treatment response.

Drug washout is a challenge in acutely ill patients. If some exacerbation in symptoms occurs, this complicates establishment of a baseline as well as adding to ethical concerns and management issues. On the other hand, absence of a washout means a true “baseline” is not achieved, assuming that there has been some degree of response and/or adverse effects from the prior treatment. The use of a concurrent placebo group in the treatment trial mitigates these concerns to some extent, but does not eliminate them entirely. The type, dosage, and half-life of prior treatments will influence how long a washout is necessary to prevent potential withdrawal effects from influencing baseline ratings. Whether or not a washout takes place (and how long it is) can have implications for assessing the effects of subsequent treatment. The effects of withdrawal are neither consistent nor predictable, which complicates establishment of an appropriate baseline.

Premorbid social adjustment is a variable that does have prognostic significance, particularly in schizophrenia. Poor adjustment is associated with poorer outcome, and may be an indicator of those patients in whom early neurodevelopmental abnormalities or prodromal symptoms were more severe.

Comorbid psychiatric disorders should be evaluated and documented. Though there are insufficient data to determine what influence comorbid conditions are likely to have on overall response to psychotropic medications, common comorbid conditions should be studied at some point to help assure generalizability and to inform clinical practice. In addition, some studies tentatively suggest that different medications may have more or less impact on measures of, for example, substance abuse, suggesting that this could also be an important outcome measure in appropriate populations.

In studying antipsychotic medications it is important to document the presence and severity of any preexisting movement disorders in order to have an adequate baseline assessment and to ensure that a preexisting condition (or withdrawal effect) is not attributed to subsequent treatment.

It is essential that patients be assessed for their capacity to give informed consent. It is beyond the scope of this chapter to discuss this in great detail, but patients should be able to describe and explain in their own words the research in which they are agreeing to participate, its goals, its experimental aspects, and its potential risks and benefits.

They must understand that they have the right to withdraw at any time and that they will not be penalized in any way if they choose to do so.

Trial Design

One of the most critical and difficult aspects of trial design is weighing and balancing what is ideal and what is feasible. An ideal trial for which patients cannot be recruited or in which they cannot be retained will not achieve its goals. In addition, though many questions ultimately need to be addressed, it is usually impossible to adequately address multiple questions in a single trial.

The duration of a trial will be influenced by whether or not a placebo group is included. The longer the duration, the more difficult to justify the retention of patients on placebo, and the higher the dropout rate, the less useful are the data.

The time course of response to psychotropic medication is generally variable. The modal time frame of response has to be factored into trial design in order to allow estimates of statistical power. In the acute treatment of schizophrenia, for example, most patients will experience at least half of the ultimate degree of improvement within the first 4 to 6 weeks (assuming that there was not an inordinately long titration phase). In many studies a significant drug effect is seen after only 1 to 2 weeks; however, different signs and symptoms are likely to have a different time course of response. For example, agitation is likely to respond more rapidly than delusions or thought disorder. In addition, there may be a subgroup of patients who are slower to respond, and for such patients longer trials may be needed. If a between-drug comparison of the full extent of response is ultimately important, then much longer trials are needed (e.g., 6 months or longer), and this begins to encompass the continuation phase of treatment. As more and more domains of outcome are of interest in clinical trials (such as primary negative symptoms or cognitive dysfunction in schizophrenia), it will be important to better characterize the time course of response for these variables in order to establish minimum and optimum durations of trials for these purposes. Estimates of expected degrees of improvement in various domains will be critical for statistical power calculations.

The Role Of Placebos

The decision as to whether or not to use a placebo in short-term, acute trials remains a topic of considerable controversy, and some dynamic tension continues to exist between “regulatory” requirements, investigators, institutional review boards, patients and families, and other interested parties. There are a number of important arguments that can be made against the routine use of a placebo in clinical trials. Rothman and Michels (3) argue that when an effective

treatment exists for a particular disease, the use of a placebo is inappropriate on both logical and ethical bases. However, the argument suggests that the use of a placebo is appropriate in cases when an effective treatment is not available. A problem remains in how to define effectiveness. The use of the term *effective* in this context is not necessarily identical to the current use of *effectiveness* as differentiated from *efficacy*.

In a complex disease such as schizophrenia, we continue to struggle with establishing the most meaningful definitions of efficacy and effectiveness. If we define response narrowly in terms of positive symptoms, then certainly some response to conventional agents is expected. In the case of severe deficit symptoms or in patients who have proven refractory to other drugs, the issue is less clear.

A particular problem arises when response to a proven effective treatment (or so-called gold standard) can vary enormously from trial to trial and in some cases be rather low, or when response to a placebo is generally high (4).

The argument is often made that in developing new drugs to treat a condition for which effective treatments are already available, the question should not be is the new drug superior to placebo but rather is the new drug superior to an already available agent. Unfortunately, given the nature of the diseases and the adverse effects associated with some psychotropic drugs, a new drug could be superior in one domain and inferior in another, while being a very valuable addition to the therapeutic armamentarium. The use of placebo controls can still be important to determine whether or not in some domains a drug is inferior, but still better than a placebo, or whether its inferiority in one domain is such that it would change the overall effectiveness equation.

To provide an example, suppose drug A were somewhat less effective than drug B in controlling acute symptoms, but some patients did quite well on drug A. At the same time, drug B was associated with serious side effects that might result in a substantial number of patients discontinuing the medication within a short period of time. Would we prefer to have drug A available to treat those patients who benefited from it, while then giving drug B to those who don't. Before approving drug A, we would want to be certain that it was superior to a placebo, though inferior to drug B in the particular domain of acute response.

There are a host of issues relating to the use of placebos that have been discussed in more detail elsewhere. As Lavori (5) has emphasized, the data sets available from current placebo-controlled trials are usually "heavily truncated, differentially by treatment groups, and certainly nonrandomly." He argues that most investigators "use ad hoc statistically unjustifiable maneuvers such as last observation carried forward (LOCF)" and that "the interpretation of positive results in the context of badly truncated data requires unverifiable assumptions, external to the observed data of the study."

Another important consideration in the use of a placebo is the potential harm resulting from a delay in instituting

active treatment. This is a difficult question to adequately address; however, there have been some attempts to examine the consequences, both short- and long-term, of receiving a placebo in the context of short-term trials. Overall, there do not appear to be demonstrable deleterious effects of participating in short-term trials (6,7). The issue of lengthy delays (i.e., 6 months or longer) in implementing treatment has been a topic of discussion in first-episode schizophrenia patients, with some authors suggesting that the longer duration of untreated psychosis is associated with poor outcome. In one patient cohort, this effect was reported in short-term outcome (8), but the effect was no longer evident in long-term follow-up (9). Short-term clinical trials usually involve durations of 4 to 8 weeks. Therefore, it is important to recognize potential differences in consequences between brief delays and relatively long delays in treatment. Lavori (5) argues that because assessments in placebo-treated patients are usually truncated because of high dropout rates, we do not know the full consequences of exposure to a placebo. The field would certainly benefit from more intent-to-treat analyses as well as long-term follow-up of patients who were involved in placebo-controlled trials.

Designs involving the treatment of patients who have failed on other treatments are another challenge. One could argue that placebo controls are more acceptable in this context because there is no effective treatment. However, it is usually the case that these patients have demonstrated some benefit from standard, albeit inadequate, treatment. Therefore, the appropriate comparison would be the new treatment versus standard treatment, with the only outcome of interest being the superiority of the former.

The decision as to whether or not to use placebo or active controls or both in a particular trial is not an easy one. There are complex issues that need to be considered, and it is hoped that further knowledge involving the determinants of heterogeneity in response will facilitate more rational and acceptable trial designs (10).

A related problem is the use of rescue medication. Balancing the desire to retain subjects and the desire to prevent harm and not withhold effective treatment is a critical issue. To what extent should other medications be available for those participants who would otherwise be dropped from a trial due to lack of efficacy and need for alternative treatment? Extensive use of rescue medication can make it difficult to accurately assess the drug effect (even though use of rescue medication can be a telling outcome in and of itself). The use of adjunctive medication to treat adverse effects that occur in the course of a trial can also be a concern (e.g., the use of antiparkinsonian medication). Here, too, rates of utilization can be an important outcome measure, yet at the same time the additional medication might have other undesirable effects (e.g., cognitive impairment).

A number of novel designs have not been widely used, and to some extent there is a disincentive to utilize them, particularly in a regulatory context.

Crossover designs have been suggested as one alternative, although some exposure to a placebo is still involved. A patient receives a potentially active compound and if response occurs, crossover to a placebo takes place. If response does not occur, the placebo phase is not required. The placebo phase in this context helps to determine whether or not the response to medication was a true drug effect or not. It is argued that this design has the advantage of each patient serving as his or her own control, allowing all patients to eventually receive active medication and increasing statistical power.

The applicability of this design varies depending on the nature of the disorder being studied, the time course of response, and the vulnerability to relapse or symptom exacerbation once active treatment is replaced by a placebo. For example, this design may be more informative in rapid cycling bipolar patients (11) than in the context of an acute treatment trial in other disorders. Also, this trial does not eliminate exposure to a placebo. From an ethical standpoint, how do we weigh the delay in providing active treatment against the withdrawal of effective treatment once a response occurs, with the outcome of interest being an exacerbation of symptoms?

Other alternative designs include adaptive allocation strategies. The intent of this approach is to reduce the number of subjects exposed to placebo, ineffective, or toxic treatments. This is achieved by altering the probability of a participant's receiving one treatment or another based on the probability established to that point in the trial of which treatment is associated with the best outcome. These designs are difficult to conduct, and they require knowledge of the results of completed subjects in order to allocate treatment for the next subject. In addition, the response criteria have to be clearly established a priori. The design becomes more complicated when three or more arms are included in a trial. Some studies have utilized such designs with success (12). The ultimate goal of reducing the number of subjects exposed to inferior treatments can be achieved; ultimately, however, the number of subjects required will depend on the effect size of interest. (For further discussion see ref. 13.)

Active Controls

Comparisons between experimental treatments and active controls require careful consideration in terms of specific drugs, dosage, adverse effects profiles, titration requirements, etc. If a dose of the comparator is too low, efficacy could be less than possible, and if the dosage is too high, then adverse effects may occur more often. This issue is often a particular concern in industry-sponsored studies, where marketing issues often influence the choice of comparator and even its dose. This highlights the potential value of studying a range of doses of both the comparator and the experimental drug. Though this is costly, the informa-

tion can be particularly valuable in informing clinical practice. Unfortunately, this is rarely done (14). To some extent, this results from unfounded assumptions that we have good data on dose-response relationships with drugs that have been in widespread use. Often that is not the case. In addition, dosage requirements will vary depending on the population. For example, in schizophrenia, first-episode patients in general respond to lower doses than multi-episode patients, and acute treatment usually requires higher doses than maintenance treatment.

Another design that is being increasingly utilized is the adjunctive or add-on strategy. This is particularly useful when subjects with partial or inadequate response are the focus of interest. Rather than switching participants from the unsatisfactory treatment to a new treatment, participants are randomized to an added placebo or added experimental treatment. In this approach, no drug withdrawal is necessary and the question of interest is whether or not the new treatment provides additional benefit.

The potential disadvantages of such a design include drug-drug interactions, particularly if a novel effect is anticipated from the adjunctive treatment. Will this be influenced by the original treatment (e.g., different receptor binding profiles)? This approach is particularly relevant when monotherapy is the exception rather than the rule. This type of design has been employed in the development of anticonvulsant medications (15).

Continuation Treatment

After improvement in acute symptomatology, there is a period of consolidation and stabilization often referred to as the continuation phase. It is assumed that discontinuation of medication during this period would be associated with a higher risk of relapse than subsequent discontinuation. It is difficult to specify when the transition from continuation treatment to maintenance (or prophylactic) treatment occurs, but at least 6 months is a reasonably conservative threshold. The question arises as to how to characterize those patients who have experienced clinically significant improvement, but continue to have more than mild symptoms. In such patients, the continuation phase could become indefinite rather than transitioning to maintenance treatment. This is a semantic distinction because the goal of maintenance treatment is to prevent a relapse or reexacerbation of psychotic signs and symptoms.

A continuation versus discontinuation design can be a sensitive test for drug effect. However, ethically, consent and protection issues are a major concern when any degree of worsening becomes an outcome measure. If such designs are considered, strategies such as sequential analyses or planned interim analysis would be important in terminating the study at the earliest appropriate time.

Maintenance Treatment

In any potentially recurring or chronic illness, the issue of long-term treatment is critical (15). Clearly, the more information on natural history and untreated course that is available from whatever source, the better in helping to define the goals and objectives of maintenance treatment. However, as is often the case, long-term outcome data in such a context are likely to be unavailable, and when comparisons are made with historical data there have often been changes in diagnostic criteria, ascertainment techniques, or other factors that would limit generalizability.

In considering the role of maintenance treatment, frequency, severity, and potential consequences of relapse are critical. Is maintenance treatment justified if a relapse is unlikely to occur for several years? This will be influenced not only by the consequences of a potential relapse, but also by the potential consequences of the prophylactic treatment itself.

In this context, the appropriateness both from a scientific and ethical standpoint of including a placebo control is an enormous concern. The fact that relapse rates on active medication and placebo can vary enormously from one study, one site, or one population to another is an important consideration. Some would argue that an active comparison involving an experimental medication could result in as many or more relapses than could occur in a placebo-controlled trial given the sample size needed to avoid a type II error. Concerns similar to those raised previously apply here as well in terms of multiple domains of outcome and benefit-to-risk ratio. If drug A had a significantly higher relapse rate than drug B but was much safer and more likely to be taken on an ongoing basis, would this drug be utilized if it were shown to be superior to a placebo? How much worse than standard treatment and how much better than placebo would a drug have to be in order to decide one way or the other? This is an unresolved issue in terms of regulatory, scientific, and ethical concerns.

Many of the issues raised previously in the discussion of acute treatment apply here as well. Patient characteristics, age, sex, ethnicity, age at onset of illness, duration of current or most recent episode, baseline psychopathology, comorbid conditions, etc. are all important issues. Even premorbid psychosocial adjustment has been shown to have some predictive power in relapse prevention studies in schizophrenia (8,17).

Issues such as reference comparator, dosage, route of administration, concomitant treatments (both pharmacologic and nonpharmacologic), a priori relapse or exacerbation criteria, duration, and strategies to enhance and measure compliance are all important in designing such studies.

The duration of such trials is critical in achieving overall goals. Results can be quite different during the first year of maintenance treatment as compared to the second, with relapse rates often being higher in the first year following

recovery from an acute episode as compared to the second year (18). At the same time, in some studies involving dosage reduction, relapse rates were higher in the second year than in the first (19).

This discussion also relates to the issue of time course of relapse in establishing appropriate durations for maintenance trials. In schizophrenia, for example, based on historical data most relapses do not occur for several months after complete drug discontinuation in stable outpatients. One context where time course of potential noncompliance and time course of relapse was such that trial designs were probably inadequate to find meaningful differences was in the comparison of oral and depot medications. A number of double-blind controlled trials were conducted in which patients were randomly assigned to depot or oral medications and therefore had to receive both injections and tablets, one of which was a placebo. The duration of all but one of these trials was 1 year. In general, they failed to find the significant differences that had been expected given high rates of non-compliance in schizophrenia and high rates of relapse following drug discontinuation. However, meta-analysis of these studies supports the value of long-acting injectable preparations (20).

It is likely that the less than expected effects were due to an inadequate duration. Given the fact that subjects agreeing to receive both injections and tablets in a double-blind design are on the more compliant end of the spectrum, one would not expect noncompliance to occur rapidly. In fact, it could take many weeks or months, particularly given the frequent assessments and the psychosocial support involved in being part of a research project. Because the relapse that ensues after complete discontinuation of medications is not likely to occur for several months, it would be unrealistic to expect to observe a difference between depot and oral medication in such a study if the duration was only 1 year (21). The only such study that lasted 2 years found no difference between treatments in the first year, but evidence of clear separation in the second (22). However, the sample size was inadequate to have sufficient statistical power to establish a significant difference, even in the second year.

The role of nonpharmacologic treatments and environmental factors in long-term studies is also important. There is clear evidence that application of nonsomatic interventions can have significant impact on relapse rates among individuals receiving pharmacotherapy. Although ideally nonpharmacologic treatment should be controlled, if it is not there should be documentation of availability and utilization so that potential confounds can be identified.

Another important issue in the design of maintenance trials is whether or not rerandomization following recovery from an acute episode is necessary to demonstrate efficacy in the maintenance phase. In some drug development programs, those patients who respond in the context of an acute trial will be followed and relapse rates reported in comparison to a reference drug. This design provides data

from only those patients who responded to each drug acutely. The argument is made that to demonstrate efficacy in relapse prevention, patients should be rerandomized or the study should be started after patients have been stabilized on any drug. This then allows conclusions to be drawn regarding prophylactic efficacy among patients in general, not just those who responded to an acute trial of a particular drug. (In addition, it is important to recognize high rates of attrition for other causes in acute treatment trials.) This is not to say that there is no value in collecting long-term continuation data on a particular medication, because these data are important in setting the stage for subsequent evaluation and comparisons.

As more domains of interest are examined in schizophrenia, it is necessary to consider the specific designs required to establish efficacy and particular outcome measures. In recent clinical trials, attempts have been made to collect data on an array of measures when at times important confounds can compromise interpretation. For example, in schizophrenia, primary negative symptoms are difficult to study in the context of an acute treatment trial that has selected patients on the basis of having clinically significant positive symptoms. Trials need to be conducted in patients selected on the basis of having residual negative symptoms not complicated by acute positive symptoms or significant extrapyramidal side effects. Remarkably few such studies have been done.

Similar concerns surround the issue of cognitive dysfunction. Newer antipsychotics show some promise in improving measures of cognitive function (23). However, studying these measures requires designs specific to their optimum assessment. In addition, the ultimate question in measuring cognitive performance will be what impact these changes have on functioning, either psychosocial or vocational, level of care, family burden, etc. To date, such studies have not been conducted, and it is premature to conclude that measurable differences on specific cognitive tests will translate into meaningful differences in functioning.

SELECTION OF PARTICIPANTS IN CLINICAL TRIALS

The issues discussed in the previous paragraph serve as examples of how patient selection becomes a critical focus in expanding our knowledge of specific drug effects.

Effectiveness Research

Increasing attention has been focused on the fact that traditional randomized clinical trials often include highly selected patients who may not be representative of the population at large. As new medications are used in routine clinical practice, there is often a considerable gap in the knowledge base needed to inform decision making. For example, many

patients with schizophrenia have comorbid conditions (e.g., substance abuse) that could influence dosing patterns, adverse effects, overall response rates, compliance, drug interactions, etc. The pharmaceutical industry does not necessarily have an incentive to conduct effectiveness research, as the narrowly defined clinical trial is the most useful and probably cost-effective approach to the drug approval process. In addition, including patients with comorbid psychiatric and medical conditions can potentially increase rates of apparent adverse effects where attribution can be difficult.

At the same time, mechanisms should be sought for conducting effectiveness trials, which are extremely important in informing clinical practice and public policy decisions.

Approaches to Subject Selection

Diagnosis and Phenomenologic Characterization

At present, diagnostic classification is an important element in patient selection. Although nosology shifts over time, it is important to incorporate into the selection criteria the use of an established diagnostic system with proven validity and established reliability. Ideally, research should involve a more systematic and formal diagnostic process than simply relying on a hospital chart diagnosis. Formal evaluation instruments are available for specific diagnostic systems. Although the use of the complete interviews may be overly time-consuming and not cost-effective for some types of research, at minimum a checklist indicating how patients met specific diagnostic criteria should be completed.

As discussed previously, diagnostic subtype has not been a consistent predictor of drug response; however, as classification systems improve and, it is hoped, subtypes become more meaningful, this element will have increasing importance in clinical trial design.

Because many psychotropic drugs are effective across a range of illnesses, a phenomenologic approach to characterizing pharmacologic effect could be reasonable. Although issues of reliability and generalizability would have to be carefully addressed, it is hoped that further research will lead to advances in this perspective.

Biological Classification

Although diseases such as schizophrenia have been characterized by a broad array of biologic abnormalities, there are as yet no well-validated biological classification systems that have proven to be useful in clinical trials or in drug development. This may be largely due to lack of systematic effects in this direction rather than an absence of potentially informative relationships. As further advances take place in diverse perspectives ranging from neuroimaging to pharmacogenomics, it is just a matter of time before biological classification becomes a critical ingredient in this context.

At present, many of the findings are based on group differences and are not necessarily appropriate as selection criteria for clinical trials. In addition, a variety of concerns including sensitivity and specificity will need to be addressed in further developing this perspective.

PHARMACOKINETIC ISSUES

The more knowledge available about pharmacokinetics and metabolism (including activity of metabolites) before large-scale clinical trials are designed, the better. Understanding potential relationships between blood levels and therapeutic response as well as adverse effects can be very helpful in optimizing treatment outcome. However, relevant data are often inadequate before critical decisions about dose and dosing schedules are made. If more attention were given to these issues earlier, clinicians would have to struggle less with establishing appropriate treatment strategies. Advances in brain imaging have set the stage for useful investigation during early stages of drug development; however, here, too, few systematic efforts have been made to take advantage of the potential of such studies to help establish optimum strategies for clinical trials.

Clinicians value the availability of different delivery methods for psychotropic medications, given the challenges of both acute and long-term treatment. Oral, liquid, intramuscular, and long-acting forms should be developed and tested in clinical trials as early as possible. Different clinical trial designs may be necessary with different preparations intended for different levels of acuity or phases of treatment. Here, too, the more information available about pharmacokinetic and pharmacodynamic issues, the better.

Given the heterogeneity of clinical response and the enormous variability in drug absorption and metabolism, randomly assigning patients to different plasma levels of interest can be a powerful tool in establishing dose-response relationships and optimum dosing guidelines. Though more difficult than the standard trial, such studies are feasible, but rarely done (24).

ASSESSMENT OF THERAPEUTIC EFFECTS AND CLINICAL CHANGE

There are many established instruments for the assessment of psychopathology in clinical trials. In some cases, these instruments have been utilized for many years. However, there continues to be a dearth of new scale development. This is partially due to the tedious nature of the development process and the reluctance of many sponsors of clinical trials to utilize a new instrument. As new drugs are developed with potentially different spectrums of activity, it would be useful to have new scales designed to be sensitive to specific therapeutic effects. This is particularly appropri-

ate since many of the original assessment scales were validated by proving sensitive to the effects of specific classes of psychotropic medications. (For detailed discussions of specific instruments for clinical assessment see refs. 25 and 26.)

As outcome measures of interest become more broad, an array of separate supplemental instruments are being employed to measure quality of life, social and vocational adjustment, cognitive functioning, and substance abuse. In designing assessment batteries, it is important to choose instruments with proven reliability and validity as well as instruments that are likely to be sensitive to the kind of treatment effect being measured. Meaningful clinical effects should be identified with specific measures of change in order to ensure that the sample size provides adequate statistical power.

As increasing numbers of assessments are employed, it is also important to recognize the burden created for patients and raters. Careful thought should go into selecting the most informative measures and planning a data analysis program with a priori primary and secondary hypotheses.

PROBLEMS IN ASSESSMENT

Because psychiatric disorders are often complex, multifaceted diseases and some key symptoms are purely subjective, the techniques used for assessment can be critical. Information regarding psychopathology is most frequently obtained from direct patient interview and observation, though information from other sources (e.g., family, nurses) is sometimes used. Patient report can be impeded by intentional concealment, lack of insight, paranoid ideation, and the overall acuity and severity of the illness. It is not uncommon for psychiatric patients to reveal more psychopathology as they begin to respond to treatment than they did prior to its initiation. The reliability and validity of different sources of information in assessing specific domains have not been adequately studied. In many trials assessors who are not familiar with the patient on an ongoing basis are asked to rate psychopathology. Although these ratings can be sensitive to treatment effects, it is likely that a person who has ongoing contact with the patient in a treatment context will provide a more accurate assessment. Here, too, research comparing different rater allocation strategies would be helpful in determining which is most valid and cost-effective. It is critical to have the same rater evaluating the patient throughout the trial whenever possible. Despite establishing high degrees of interrater reliability, this kind of continuity is important.

The timing of assessments and the time frame chosen for a given assessment should be determined by the goals in the study. In general, when rating psychopathology, the previous week is a reasonable time frame. Patients are less likely to accurately recall specific symptoms that are more

remote in time. The time frame used for a particular assessment does not need to coincide with the interval between assessments. In a long-term trial it is not necessary to rate patients weekly. But when they are assessed, the previous week can be the focus of the assessment.

ASSESSMENT OF ADVERSE EXPERIENCES

The two major goals of drug development—to enhance therapeutic efficacy and to improve tolerability—go hand in hand, particularly in the case of psychotropic medications, where many side effects of psychotropic drugs overlap clinical signs and symptoms of psychiatric illnesses. Given the frequent long-term nature of psychotropic drug treatment, adverse effects become critical in influencing compliance and determining the overall benefit-to-risk ratio.

In general, the methods for detecting adverse events have been given far less attention than the methods for evaluating efficacy. Controversy exists as to the most valid means of accurately estimating the incidence of adverse effects. Many clinical trials rely on patient self-report, with some specific queries or rating scales used to assess known adverse effects (e.g., extrapyramidal side effects or tardive dyskinesia) that are outcomes of interest. Given the subjective nature of many adverse events, there is a concern that detailed, specific queries across a broad range of possible symptoms will result in the elicitation of far more symptoms than an unstructured approach.

A methodologic comparison study (27) suggested that the general elicitation of adverse events is more practical and appropriate for routine clinical trials than a comprehensive and lengthy interview. At the same time the field needs to acknowledge the possibility of inordinate delays in recognizing the frequency of specific adverse events such as the sexual dysfunction associated with selective serotonin reuptake inhibitor (SSRI) antidepressants.

There is a strong argument for the use of data and safety monitoring boards when large and/or long-term studies are involved or high-risk treatments are being studied.

CONCLUSION

The clinical trial remains the mainstay of treatment development. It is always hoped that further advances will evolve more rapidly than they do, but there is reason for considerable optimism that over the next decade there will be important advances in predicting and understanding psychotropic drug response whether via functional neuroimaging, pharmacogenomics, or other potential developments. In addition, it is hoped that increasing emphasis on studying a broader array of functionally meaningful outcome measures in the context of better informed benefit-to-risk assessment and documentation of cost-effectiveness will lead to clinical

trial designs to better address the full range of public health issues.

REFERENCES

1. Klein DF. Improvement of phase III psychotropic drug trials by intensive phase II work. *Neuropsychopharmacology* 1991;4: 251–257.
2. Allison DB, Mentore JL, Heo M, et al. Antipsychotic-induced weight gain: a comprehensive research synthesis. *Am J Psychiatry* 1999;156(11):1686–1696.
3. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994;331:394–398.
4. Leber P. The placebo control in clinical trials (a view from the FDA). *Psychopharmacol Bull* 1986;22:30–32.
5. Lavori PW. Placebo control groups in randomized treatment trials: a statistician's perspective. *Biol Psychiatry* 2000;47: 717–723.
6. Curson DA, Hirsch SR, Platt SD, et al. Does short term placebo treatment of chronic schizophrenia produce long term harm? *Br Med J* 1986;293:726–728.
7. Wyatt RJ, Hinter ID, Bartko JJ. The long-term effects of placebo in patients with chronic schizophrenia. *Biol Psychiatry* 1999;46: 1092–1105.
8. Loebel AD, Lieberman JA, Alvir JMJ, et al. Duration of psychosis and outcome in first episode schizophrenia. *Am J Psychiatry* 1992; 149:1183–1188.
9. Robinson D, Woerner MG, Alvir JMJ, et al. Predictors of relapse following response to first episode schizophrenia or schizoaffective disorder. *Arch Gen Psychiatry* 1999;56:241–247.
10. Mattocks KM, Horwitz RI. Placebos, active control groups, and the unpredictability paradox. *Biol Psychiatry* 2000;47:693–698.
11. Post RM, L'Herrou T, Luckenbaugh DA, et al. Statistical approaches to trial durations in episodic affective illness. *Psychiatry Res* 1998;78:71–87.
12. Tamura RN, Farris DE, Andersen JS, et al. A case study of an adaptive clinical trial in the treatment of outpatients with depressive disorder. *J Am Stat Assoc* 1994;89:768–776.
13. Rosenberger WF, Hu F. Bootstrap methods for adaptive designs. *Stat Med* 1999;18:1757–1767.
14. Zimbroff DL, Kane JM, Tamminga CA, et al. Control, dose-response study of sertindole and haloperidol in the treatment of schizophrenia. *Am J Psychiatry* 1997;154:782–791.
15. Brodie MJ. Antiepileptic drugs, clinical trials, and the marketplace. *Lancet* 1996;347:777–779.
16. Greenhouse JP, Stangl D, Kupfer DJ, et al. Methodological issues in maintenance therapy clinical trials. *Arch Gen Psychiatry* 1991; 48:313–318.
17. Kane JM, Rifkin A, Quitkin F, et al. Fluphenazine versus placebo in patients with remitted, acute first episode schizophrenia. *Arch Gen Psychiatry* 1982;39:70–73.
18. Hogarty GE. Prevention of relapse in chronic schizophrenic patients. *J Clin Psychiatry* 1993;54(suppl 3):18–23.
19. Marder SR, Van Putten T, Mintz J, et al. Low and conventional dose maintenance therapy with fluphenazine decanoate. Two-year outcome. *Arch Gen Psychiatry* 1987;44:518–521.
20. Davis JM, Barter JT, Kane JM. Antipsychotic drugs. In: Kaplan HI, Sadock BJ, eds. *Comprehensive textbook of psychiatry*. Baltimore: Williams & Wilkins, 1989:1591–1626.
21. Kane JM, Borenstein M. Compliance in the long-term treatment of schizophrenia. *Psychopharmacol Bull* 1985;21:23–27.
22. Hogarty JE, Schooler NR, Ulrich R, et al. Fluphenazine and social therapy in the after care of schizophrenic patients: relapse analysis of a two-year controlled study of fluphenazine decanoate

- and fluphenazine hydrochloride. *Arch Gen Psychiatry* 1979;36: 1283– 1294.
23. Meltzer HY, McGurk S. The effects of clozapine, risperidone, and olanzapine on cognitive function in schizophrenia. *Schizophr Bull* 1999;25(2):233–255.
 24. Peck CC. The randomized concentration-controlled clinical trial: an information rich alternative to the randomized placebo-controlled trial. *Clin Pharmacol Ther* 1990;47:126.
 25. Barnes RE, Nelson HE, eds. *The assessment of psychosis: a practical handbook*. London: Chapman and Hall Medical, 1994.
 26. American Psychiatric Association. *Handbook of psychiatric measures*. Washington, DC: American Psychiatric Association, 2000.
 27. Rabkin JG, Markowitz JS, Ocepak-Welikson K, et al. General versus systematic inquiry about emergent clinical events with SAFTEE: implication for clinical research. *J Clin Psychiatry* 1992; 12:3–10.