Conducting Clinical Trials for Novel Treatments in Psychiatry: A Global Perspective

www.ppdi.com | November 2012
EXECUTIVE SUMMARY

In psychiatry, new pharmacologic treatments must show a meaningful clinical advantage over existing standard-of-care treatments to gain regulatory approval and achieve commercial success. Designing and executing clinical trials to measure these advantages requires careful consideration of several factors, including unmet medical need, use of placebos, drug-drug interactions, implementation of rating scales and relevant social, cultural and regulatory differences between geographic regions.
INTRODUCTION

The last decade of the 20th century and the first decade of the 21st was a period of near-frenzied expansion of clinical trials in psychiatry as the second-generation antipsychotics, selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs) demonstrated their efficacy and safety relative to both placebo and other drugs in their therapeutic classes. These compounds, which are now largely generic, have established a standard of care for the treatment of schizophrenia, depression and all phases of bipolar disorder. Any new therapy must show a meaningful clinical advantage over existing treatments to achieve both regulatory approval and commercial success.

Designing and executing clinical trials to measure these advantages requires careful consideration of several factors. This paper addresses five key areas:

- **Unmet medical need:** The current standards of care for schizophrenia, depression and bipolar disease have well-appreciated inadequacies that result in several unmet medical needs
- **Drug-drug interactions:** Add-on trials of experimental medications are based on the assumption that the pharmacokinetics and pharmacodynamics of the background regimen are unaffected by the new drug
- **Use of placebos:** While placebos (either alone or in combination with background therapy) are theoretically the ideal mechanism for measuring the therapeutic and adverse effects of an experimental drug, their use is complicated by practical hurdles (including the placebo effect) and, in some cases, ethical concerns
- **Implementation of rating scales:** Clinical trials in psychiatry cannot be successful without a sensitive, specific, validated rating instrument (with verified and culturally relevant translations, if applicable) administered by trained, certified personnel
- **Regional differences:** Many clinical trials now have sites in multiple countries, and the social, cultural and regulatory differences between nations cannot be overlooked when designing a clinical trial

UNMET MEDICAL NEEDS IN PSYCHIATRY

The greatest unmet need in psychiatry may be in schizophrenia, where cognitive disturbance and negative symptoms (e.g., diminished emotional response, lack of motivation, social withdrawal) frequently persist in patients whose positive symptoms (e.g., hallucinations, delusions, disorganization of thought) are well-controlled by medication. Many schizophrenic patients are affected by both cognitive disturbance and negative symptoms and these symptoms often co-exist in the same patient.

While approved assessment scales exist for both conditions, many sponsors focus their clinical trial efforts on negative symptoms because a statistically significant improvement in these symptoms is considered clinically relevant and sufficient for an indication. In contrast, regulatory agencies have mandated that drugs seeking an indication for cognitive disturbance must demonstrate efficacy using an approved cognitive assessment scale as well as a clinically relevant “real life” functional endpoint. Unfortunately, a suitable functional scale has been difficult to establish and few trials utilize this approach.

Appropriate assessment of both cognitive disturbance and negative symptoms requires maintenance of the ongoing antipsychotic medication(s) to ensure the clinical stability of positive symptoms over the course of a trial. Coupled with a trend toward longer study durations that can document preservation of any observed improvement, most trials in this area utilize an add-on design in which patients are randomized to the experimental drug or placebo in addition to their underlying medication(s). The continuation of antipsychotic medication and the absence
of any approved treatment for cognitive disturbance and/ or negative symptoms eliminate any ethical objection to the use of placebo in this context.

Another unmet medical need in schizophrenia is pharmacotherapy for patients whose positive symptoms are inadequately controlled by clozapine, with or without other antipsychotic and anti-mania drugs. This indication can be studied using two clinical trial designs: (1) patients are randomized to receive either the experimental drug or placebo in addition to their current treatment regimen (the add-on approach); or (2) patients are randomized to receive the experimental drug as monotherapy or continue on their current regimen.

The main unmet medical need in major depressive disorder (MDD) is effective therapy for patients who have failed two or more drugs of different classes given at adequate dose (maximum tolerated) for adequate duration (at least six weeks) and in which at least one of the failures was prospectively observed. Most novel agents in this area do not increase intra-synaptic dopamine, norepinephrine or serotonin and can be evaluated using an add-on study design based on the patient’s last (unsatisfactory) regimen. As with refractory schizophrenia, success is defined as statistical superiority of the experimental drug compared with current therapy. Following demonstration of success as adjunctive therapy, the new agent also may be evaluated as monotherapy.

Other indications in psychiatry have significant unmet needs for effective pharmacologic intervention. Medications for autism, fragile X syndrome, adult attention deficit disorder (ADD) and post-traumatic stress disorder (PTSD) have entered clinical studies in recent years, as have treatments for nicotine, ethanol, stimulant and opiate abuse. In most instances, these compounds can be tested in placebo-controlled monotherapy trials, although the courageous may attempt to show the experimental drug to be superior to the (generally inadequate) standard of care.

**DRUG-DRUG INTERACTIONS**

Add-on trials of experimental medications are based on the assumption that the pharmacokinetics and pharmacodynamics of the background regimen are unaffected by the new drug. If the plasma concentration of the underlying medications and/or their active metabolites is raised or lowered by the experimental treatment, any observed changes in efficacy or adverse events cannot confidently be attributed directly to the add-on treatment. Thus, add-on treatments and their metabolites should not inhibit or induce the major hepatic enzyme systems or transporters of underlying drugs and their active metabolites.

The metabolic inhibition or induction of an underlying medication may be of limited consequence if there is a large margin of safety in the drug’s plasma concentration. Alternatively, if the variability of the inhibition/induction is limited, it may be clinically feasible to increase/reduce the dose of the underlying treatment by a fixed amount. In such cases, it may be possible to employ a blinded study design using either standard or modified doses. For medications with narrow therapeutic indices and/or significant variability in pharmacokinetic interaction with the experimental drug, plasma monitoring may be valuable. However, this approach is only practical in unblinded studies.

Add-on therapies in depression and schizophrenia should not inhibit or induce CYP 1A2, 2D6, 2C19 and 3A4, as the commonly used second-generation antipsychotics, SSRIs and SNRIs are metabolized by one or more of these enzymes. Ziprasidone is also metabolized by aldehyde oxidase. The effects of the experimental treatment and these antipsychotics and antidepressants on protein transporters and any possible effects of the underlying treatments on the metabolism of the experimental drug also need to be considered.
REduCing tHe PlaCeto reSPonSe

Over the past 20 years, the placebo response has increased dramatically in placebo-controlled trials of schizophrenia and major depression.\(^1\)\(^3\) This change has resulted in reduced effect sizes and failed studies for both new chemical entities and drugs with proven efficacy. Potential explanations have included regression to the mean, poor compliance, improper dosing and a higher chance of being on active treatment, but these factors were also in play when tricyclic, SSRI and SNRI antidepressants and second-generation antipsychotics were successfully tested and launched.

One proposed method for reducing the placebo response in trials of major depression was the exclusion of patients who improved during a placebo run-in, but this approach has not found empirical support.\(^6\) Another possible source of decreased drug/placebo difference is the more effective study blinding resulting from the relatively benign side effect profile of SSRs and SNRIs. With fewer adverse events, there is a reduced expectation of treatment effects by both investigator and patient. This theory is supported by Moncrieff, et al., who found no difference between active treatment and pharmacologically active, side effect-inducing placebo.\(^7\) The argument is further bolstered by the demonstration that naloxone can block the placebo effect in analgesia trials,\(^8\) but the unblinding conjecture does not lead to improved clinical trial efficiency. Shorter trials (four to six weeks) may be warranted in acute depression to reduce the placebo effect resulting from natural remission of acute depression and to ease the ethical burden of the placebo arm.

Increased baseline depression severity has been shown to predict the efficacy of active treatment over placebo due to decreased placebo response\(^5\) and greater effect of active treatment.\(^3\) These data suggest requiring higher baseline scores for entry into depression trials. However, this approach reduces the available patient population and the generalizability of study findings, prolongs recruitment and increases the possibility of rater inflation (which defeats the purpose of the higher baseline threshold for entry). Baseline rater inflation can be significantly reduced by implementing concomitant patient self-rating, either online or through an interactive voice response system (IVRS), or through centralized rating of the baseline visit.

In trials of schizophrenia, a longer period of acute illness has been associated with decreased effect size.\(^3\) Another potential factor is the inclusion of “professional patients,” an issue that can be addressed by excluding patients who have participated in more than a certain number of similar trials. A third possibility is regional differences in baseline symptom severity and increased variability deriving from scale translation and rater training in multiple languages, but this factor is difficult to prove. Finally, increasing the minimum baseline severity score for inclusion introduces the problem of rater inflation, which cannot be reliably addressed in schizophrenia trials using patient self-rating. However, centralized baseline rating can be used.

One direct and compelling explanation for the failure of an active drug to separate from a placebo is incomplete compliance in the active treatment arm. For lack of a better alternative, pill counts were, until recently, the inadequate industry standard for measuring adherence. New digital technology offers possible solutions through smart bottles, hand-held devices and, most recently, chips encapsulated in the medication. Proteus Biomedical developed a biodegradable chip that sends signals to a device worn by the patient that transmits time of ingestion and other data through a smart phone to a central database.

The proportion of patients who respond to placebo in major depression trials varies from 10 percent to 50 percent.\(^9\) This frightening statistic highlights the need to minimize rater variance in studies of depression and schizophrenia. The next section provides guidance on the
Throughout the iterative process, the instrument is evaluated for its:

- **Reliability**: Yields consistent, reproducible estimates of true treatment effect
- **Content validity**: The instrument measures the concept of interest. Qualitative research is used to demonstrate that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population and use. Testing other measurement properties will not replace or rectify problems with content validity
- **Construct validity**: Relationships among items, domains and concepts conform to a priori hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups
- **Ability to detect change**: Instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept

Clearly, developing and validating a new patient- or physician-reported instrument (or revising an existing scale) is not a trivial undertaking. However, as new central nervous system therapeutic targets emerge, it may become necessary to reconsider the adequacy of tools currently used to measure therapeutic effect and clinical relevance. Sponsors should always discuss construction or use of a new instrument with regulatory authorities before launching a Phase III study.

**Deployment of Rating Scales**

Sponsors choosing to use an established rating scale that is not in the public domain must obtain permission from its developers and often must pay a licensing fee. If the scale is needed in languages that the developer cannot provide, the sponsor also must request permission to translate the instrument.
The translation process is performed by specialized vendors who ensure that the scale measures the same constructs regardless of language or culture. In most cases, the original scale is translated by two independent local language experts, merged into a single translated version and then back-translated into the original language and compared with the source document. Cultural adaptation may also be necessary. For example, “shortness of breath” is nonsensical when translated directly into Mandarin Chinese, because in that language breathing is not described as “short” or “long.” Visual cues should not be overlooked, as objects that are readily identifiable by Americans (e.g., a football, the Statue of Liberty) may be completely unknown to patients in other countries. Translated instruments are also reviewed by a medical expert to confirm clinical relevance. Vendors maintain detailed records and issue a certificate of linguistic validation documenting the process.

Best Practices for Training Raters to Properly and Consistently Administer Rating Scales
Verifying the sensitivity and specificity of the rating instrument and the quality and cultural relevance of its translation are necessary steps in initiating a clinical trial, but training the raters to properly and consistently administer the scale is critical for success. The validity and reliability of outcome measures in psychiatry clinical trials are improved when using a pool of similarly qualified raters trained via a centralized program. This section reviews some best practices for the selection, training and calibration of raters in a global clinical trial.

Rater Selection
Potential raters for a clinical study should be assessed against pre-specified criteria for education, training and experience with both the indication and the relevant clinical scales (if applicable). For example, the sponsor of an MDD trial may mandate that raters have a Ph.D. in psychology and more than three years of experience administering the Montgomery-Asberg Depression Rating Scale (MADRS) in clinical settings or a master’s degree and more than five years of clinical experience with MADRS.

While some leeway in qualifications may be acceptable (e.g., a Ph.D. with 10 years of experience administering the Hamilton Depression Rating Scale but only one year with MADRS), better continuity in the rater population is likely to yield more reliable trial results.

Trials that mandate raters be blinded to baseline information, outcomes from other rating scales or adverse events will require multiple raters at each site to administer all scales and clinical care for a given patient. Even in trials without blinding restrictions, it is advisable to select at least two raters for each site to ensure availability despite vacations, sick leave and other staffing issues.

Rater Training
Rater training programs are customized based on the disease, protocol-specified scale and expected level of rater experience. Ideally, all raters will receive the same training using the same materials. Recording the original live training session or developing virtual training modules used by all raters are cost-effective ways to ensure consistency even if raters are added while the trial is underway. Validated translations of the materials and training should be obtained and utilized, if needed.

Applied skills training workshops (either live or via web conference) are an option for augmenting the didactic training. During these workshops, raters conduct mock interviews with actors and receive expert feedback on their performance.

Rater Certification, Quality Assurance and Recalibration
Once training is completed, a rater is certified on the primary and important secondary outcomes measures in the trial. The most common certification process asks the rater to score the relevant rating scale(s) using a standardized video of a patient or actor, then compares the rater’s results with scores obtained by an expert (or panel of experts). Raters whose outcomes are within an acceptable range are certified to rate that scale for that trial, while raters whose outcomes are near the acceptable range may
receive additional training and a second certification test. Raters not demonstrating sufficient skills with the scale(s) cannot participate in the trial. It is imperative that raters be trained and certified before the first scale is administered to the first subject at a given clinical trial site.

Once a rater has begun assessing patients, a centralized quality assurance program is critical to ensuring that the instruments are administered accurately and consistently. Quality assurance, which should be focused on a rater’s first few patients, can be performed by reviewing data for internal inconsistencies, evaluating audio or video from recorded assessments, examining rater worksheets and, where applicable, comparing rater assessments and patient self-assessments.

Trials of longer duration may include recalibration of raters at specific time points (for example, every 12 months) to enhance inter-rater reliability and prevent rater drift. Recalibration involves the raters scoring another video presentation or a written case interview. Quality assurance assessments may also be repeated at similar intervals.

Documentation of a rater’s qualifications, training, certification, quality assurance and recalibration (if applicable) is stored centrally and at the study site. These materials are used to prepare reports on inter-rater reliability at the conclusion of the study.

Other Options

Inter-rater reliability also can be addressed with the use of rigorously trained central raters at baseline (to reduce rater inflation) and/or throughout the trial (which may reduce the variability in scores for the primary efficacy variables). For trials utilizing cognitive batteries to measure dementia or cognitive disturbance in schizophrenia, fully computer-based assessments are available. Both of these approaches are under active investigation and may be appropriate in specific circumstances.

UNDERSTANDING REGIONAL DIFFERENCES THAT MAY AFFECT TRIAL DESIGN

The need for patients to participate in ever-larger studies in psychiatry brought many new investigators into clinical research, globalized clinical trials and resulted in the establishment of a clinical trial infrastructure in China, Eastern Europe, India, Latin America and other emerging economies. However, even a well-designed trial that addresses an unmet medical need, incorporates appropriate placebo controls and uses well-trained raters with validated instruments may fail to gain local approval or fully enroll if the social, cultural and regulatory norms of prospective clinical sites are not considered.

This section discusses several factors that may influence trial design and conduct in four major global regions: the United States, Europe, Latin America and Asia Pacific. When planning clinical trials, sponsors and clinical research organizations (CROs) should take into account the guidance of local/regional key opinion leaders, national consultants and responsible governing bodies (e.g., National Institute for Health and Clinical Excellence in the United Kingdom).
United States

- **Access to care and medication**: Broad but variable. Many psychiatric medicines are generic, and newer medications are generally covered by insurance (although restrictions are increasing)

- **Primary patient motivation/incentive**: Access to newer, better or more expensive treatment

- **Patient availability**: Highly variable. There are a few large concentrations of patients in specialized or central facilities; the remaining patients are spread across a wide range of public, private, inpatient and outpatient facilities

- **Use of placebos**: Not restricted by the U.S. Food and Drug Administration (FDA). Ethics committees occasionally limit trials in which use of placebo places vulnerable patients in imminent danger. The threshold for imminent danger is generally higher in the U.S. than in other regions

- **Other factors**:
  - Advertising for patients is common
  - Patients are highly mobile, so long-term trials can suffer from excessive loss to follow-up. This issue is less of a concern in schizophrenia studies

Europe

- **Access to care and medication**: Generally good but highly variable, especially between Western and Eastern European countries

  - Referral systems are common
  - Access to new, more expensive medications is sometimes limited in Eastern Europe, although the broadening availability of generics is mitigating this issue

- **Primary patient motivation/incentive**: Access to unapproved medications, although outpatient incentives, such as transportation or meals, have been shown to significantly boost recruitment

- **Patient availability**: Variable
  - Most countries have well-developed networks of public and private psychiatry clinics, but investigator experience is a limiting factor
  - Investigators must have their own patient databases and strong relationships with independent psychiatrists as well as medical personnel from district dispensaries and regional hospitals
  - Many trials are conducted in centralized hospitals that provide care to severely psychotic or depressed inpatients
  - In Russia, facilities must be accredited by the Ministry of Health before participating in clinical trials. At present, there are 63 accredited psychiatry institutions throughout the country, making Russia an attractive location for clinical trials in psychiatry

- **Use of placebos**: Controversial, especially for trials involving outpatients, acute episodes or a significant risk of suicide

  - Placebo-controlled trials can be approved if the protocol clearly describes all relevant factors, such as duration of placebo treatment, rescue medications and conditions of patient withdrawal

- **Other factors**:
  - Advertising for patients is common
  - Patients are highly mobile, so long-term trials can suffer from excessive loss to follow-up. This issue is less of a concern in schizophrenia studies

  - Acceptance of mandatory hospitalization during the trial, as well as the per bed cost, varies by country
Conducting Clinical Trials for Novel Treatments in Psychiatry: A Global Perspective

Latin America
- Access to care and medication: Heterogeneous across the region. This variability encourages matching the protocol to a specific country/countries based on local health care systems, regulations and ethical standards
- Primary patient motivation/incentive: Access to better care and medications
- Patient availability: Extensive network of psychiatry practices based in private institutions, research institutions, academic sites and public hospitals. In some countries, such as Mexico, clinical trial patients with severe disorders are typically hospitalized
- Use of placebos: Varies by country
  - In Argentina and Brazil, it is very difficult to obtain approval for trials in psychiatry that alter the standard of care. Use of placebo groups in monotherapy trials is rare
  - Trials involving placebos or wash-out periods in ambulatory patients are challenging throughout the region, as local health insurance does not cover medical care, including hospitalization, resulting from participation in a clinical trial
  - Add-on studies (even with a placebo arm) are generally acceptable
- Other factors:
  - Investigators in Latin America are very experienced and motivated to participate in psychiatry clinical trials

Asia Pacific: China
- Access to care and medication: Reasonable for patients with depression or schizophrenia living in large cities, although newer medications may require patients to pay out-of-pocket
- Primary patient motivation/incentive: Access to newer, better or more expensive treatment
- Patient availability: High in large urban facilities
  - All centers conducting Phase I to III trials for marketing approval must be recognized by the Chinese State Food and Drug Administration (SFDA)
  - There are approximately 38 recognized centers for psychiatry trials, all of which are specialized psychiatry hospitals
- Use of placebos: Can be difficult depending on the indication. Even if the study is approved, recruitment may be disappointing
- Other factors:
  - Some conditions, such as ADD, are not well-known among patients and general health care providers
CONCLUSION

Since any new therapy for psychiatric indications would require significant clinical advantage over existing treatments to achieve both regulatory approval and commercial success, the design and execution of clinical trials to measure these advantages will require careful consideration. The points discussed above should prove helpful in planning for and conducting clinical programs using the following approaches:

• Placebo-controlled adjunctive therapy for most programs in schizophrenia and depression where generally safe and effective treatments are generically available
• Monotherapy trials for other new indications.

Asia Pacific: India

• Access to care and medication: Highly variable depending on social class and geographic location
• Primary patient motivation/incentive: Access to better care and medications
• Patient availability: Limited because of the relatively small number of qualified psychiatry investigators per capita
  ◦ The few high-quality, experienced sites are often saturated with studies
  ◦ When required by the study protocol, it can be difficult for sites to find independent external raters
• Use of placebos: Challenging when there are ethical issues regarding the risk of worsening of illness
• Other factors:
  ◦ Psychiatry studies conducted in developing countries attract considerable scrutiny, particularly regarding the validity of informed consent
  ◦ India has numerous official languages, requiring multiple translations of patient materials. The issue is compounded if an independent rater is required for each language

Asia Pacific: Other

• For most other Asia-Pacific countries (e.g., Australia, New Zealand, Singapore, Taiwan, Thailand, Malaysia and South Korea), the treatment setting for psychiatry studies is similar to the U.S. and Western Europe
REFERENCES


2 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.


ADDITIONAL REFERENCES


