Exploratory Analyses of Efficacy Data From Schizophrenia Trials in Support of New Drug Applications Submitted to the US Food and Drug Administration

Ni A. Khin, MD; Yeh-Fong Chen, PhD; Yang Yang, PhD; Peiling Yang, PhD; and Thomas P. Laughren, MD

ABSTRACT

Objective: There has been concern about a high rate of placebo response and a decline in treatment effect over time in schizophrenia trials as well as the implications of increasing conduct of such trials outside North America. This report explores differences in efficacy data over an 18-year period from randomized placebo-controlled trials submitted in support of new drug applications (NDAs) for the treatment of schizophrenia and differences in results between trials conducted in North America and elsewhere.

Data Sources: Clinical trial data that were submitted to the US Food and Drug Administration (FDA) as part of NDAs for the indication of schizophrenia between 1991 and 2009.

Study Selection: Efficacy data were compiled from 32 clinical trials with 11,567 evaluable patients with schizophrenia. Data from completed, randomized, multicenter, double-blind, placebo-controlled, 4- to 8-week clinical trials in adult patients diagnosed with schizophrenia according to *DSM-III* or *DSM-IV* criteria were included.

Data Extraction: Baseline demographic and disease characteristics, including mean Positive and Negative Syndrome Scale (PANSS) total scores, were summarized and compared between North American and multiregional trials. Mean change from baseline to endpoint in PANSS total scores was utilized as the primary outcome of interest. We explored differences in treatment effect and success rate of these trials based on when and where the studies were conducted, sample size, trial duration, and baseline patient characteristics.

Results: Twenty-one of the 32 trials were conducted solely in North America, and 11 were carried out in multiple regions. Of those 11 multiregional trials, 2 were conducted exclusively in foreign countries. Although the observed responses (change from baseline) in placebo and drug-treated groups in multiregional trials tended to be larger than in North American trials, the treatment effects (drug-placebo difference) were –9 and –8 PANSS units for North American and multiregional trials, respectively. When time of trial conduct was taken into account, an increasing placebo response and a diminishing treatment effect over time were observed in North American trials from –10.8 PANSS units for the first period (1991–1998) to –6.0 PANSS units for the later period (1999–2008). The overall trial success rate over the almost 2 decades was 78%, declining slightly in trials conducted after 1999, the time period during which multiregional trials were first conducted (74% for 1999–2008 vs 85% for 1991–1998), despite increasing sample sizes in the later period. The mean baseline PANSS total score was in the range of 87–100 for most of these trials. Trials in patients with higher mean baseline PANSS total scores tended to show larger treatment effects than those in patients with lower scores. The mean body weight and body mass index (BMI) were higher in patients in North American trials and North America–predominant multiregional trials compared to those in foreign-predominant multiregional trials (mean body weights of 85 kg and 81 kg vs 72 kg, and BMIs of 29 and 27 vs 25, respectively). Treatment effects decreased as body weights increased, especially in North American trials. In foreign-predominant multiregional trials, there were higher proportions of women than in North American trials and North America–predominant multiregional trials (40% vs 22% and 27%, respectively) and a relatively larger proportion of Asians (21% vs 1% and 8%, respectively).

Conclusions: A high and increasing placebo response and a declining treatment effect are of great concern in schizophrenia trials conducted in North America. In this era of global clinical trials, close attention is needed to the design and conduct of these trials.

J Clin Psychiatry 2012;73(6):856–864

Submitted: *November 15, 2011; accepted February 15, 2012.*
Online ahead of print: *May 15, 2012 (doi:10.4088/JCP.11r07539).*
Corresponding author: *Ni A. Khin, MD, Division of Psychiatry Products, HFD-130, Food and Drug Administration, 10903 New Hampshire Ave, Bldg WO22, Rm. 4110, Silver Spring, MD 20993-0002 (ni.khin@fda.hhs.gov).*

In global drug development, central nervous system disorders rank second among the diseases studied, including many programs for drugs to treat depression and schizophrenia.[1] Schizophrenia is a chronic, disabling mental illness that affects approximately 2.4 million adults in the United States, with a lifetime prevalence of about 1.5% and with a relatively similar rate across countries.[2–4] In 2001, it was reported that, of the randomized placebo-controlled multicenter trials conducted for the treatment of schizophrenia, 25% of these trials failed to show an effect.[5] It has also been reported that the treatment effect in randomized controlled trials of atypical antipsychotic drugs has diminished over time.[6] This apparent reduction in antipsychotic treatment effect has been suggested to result from an increase in the placebo response in clinical trials conducted in recent years.[7,8]

It has been observed that an increasing number of clinical trials are being conducted in countries located in emerging regions such as Asia, Eastern Europe, and Latin America.[9] Data submitted to the US Food and Drug Administration (FDA) in support of new drug applications (NDAs) for an antipsychotic claim show that schizophrenia trials are increasingly being conducted as multiregional trials.[8] FDA accepts data generated from foreign sites for NDAs as long as they are from adequate and well-controlled trials that are conducted in compliance with standards of Good Clinical Practice.[10–13] Nevertheless, there have been ongoing concerns about the applicability of foreign data, particularly from regions outside North America or Western Europe, to US practice. These concerns are based largely on possible regional differences in patient populations, disease characteristics, and medical practice that could result in differences in placebo and treatment responses.[14,15]

This article provides the results of exploratory analyses that examine differences in response in drug- and placebo-treated groups, treatment effect size (drug-placebo difference), and success rate of placebo-controlled schizophrenia trials submitted in support of NDAs from 1991–2009. Differences were

Clinical Points

Of great concern in schizophrenia trials conducted in North America are

- A high and increasing placebo response and
- A diminishing treatment effect.

examined with regard to when and where the studies were conducted, sample size per treatment arm, trial duration, and baseline patient characteristics.

METHOD

Data Collection

Twelve antipsychotic drug programs in support of an indication for schizophrenia were identified among original NDAs submitted to the FDA between 1991 and 2009. Trials from all NDAs for oral formulations, regardless of approval status, were considered for inclusion in the database. They were all randomized, multicenter, double-blind, placebo-controlled clinical studies of 4 to 8 weeks' duration with 40 or more patients in at least 1 treatment arm. Patients enrolled in these trials were adults (age ≥ 18 years) diagnosed with schizophrenia according to the *Diagnostic and Statistical Manual of Mental Disorders (DSM)-III* or *DSM-IV* criteria. Trials limited to known drug responders, such as those in maintenance studies using a randomized withdrawal design, were excluded. Data used in this exploration were derived from trials that required informed consent.

The 2 mostly widely accepted and used rating scales in schizophrenia trials are the Positive and Negative Syndrome Scale (PANSS)[16,17] and the Brief Psychiatric Rating Scale.[18] Most studies used change from baseline to endpoint in PANSS total score as the prespecified primary efficacy variable. For our database, the mean change from baseline to endpoint in PANSS total scores was utilized as the primary outcome of interest. We included data from all trials that used the PANSS as either the primary or a secondary efficacy measure. Missing data were imputed using the last-observation-carried-forward (LOCF) approach.

To achieve a more comprehensive overview and to reduce selection bias, we considered it critical to include as many trials (including failed and negative trials) as possible in our data collection. Our initial search identified 34 schizophrenia trials with a total of 12,697 patients. Because the question of greatest interest is what trial design and other factors might affect the trial outcome for drugs and doses known to work, we further limited our database by only collecting trial data for drugs widely viewed as effective antipsychotics in the United States and around the world, and, for these drugs, at doses considered to be within an effective range. Many of the trials (n = 25) included more than 1 dose group, and some (n = 23) included an active control. After removing the doses and drugs we viewed as ineffective, we were left with data from 32 schizophrenia trials with 11,567 evaluable patients, defined as patients with a baseline and at least 1 postbaseline efficacy assessment. The date of trial conduct initiation was also noted; if it was not available, the NDA submission date was used.

Data Analysis

This exploratory analysis included all submitted short-term, randomized, placebo-controlled trials that were determined to be of adequate size and to have appropriate patient populations and entry criteria. Our analyses for this article included trials submitted in support of NDAs from 1991–2009, and analyses of these studies were based on the aggregated trial-level data drawn primarily from sponsors' study reports and the FDA's reviews. Statistical modeling of treatment effect using patient-level data pooled from trials submitted between 1993 and 2005 has already been performed; however, prior to 1997, some patient-level data were not available.[8]

Baseline demographic characteristics (age, gender, race, body weight/body mass index [BMI]), dropout rate, and baseline mean PANSS total scores were summarized and compared between North American (United States and Canada) and multiregional trials. Comparisons based on duration of trials, which ranged from 4 to 8 weeks, were also carried out.

We assessed mean treatment effect and trial success rate, where mean changes from baseline in PANSS total score at final visit for placebo and drug-treated groups were calculated based on LOCF data with analysis of covariance for each trial. The findings reported in this article are based on estimates from the aggregated trial-level data, using an LOCF imputation method. We understand that estimates from LOCF data are likely to be biased when the mechanism of missing data is not completely at random, particularly in the presence of high dropout rates.[19] However, this was the only approach for which results were available for all trials; the LOCF analysis was the protocol-specified primary analysis in most of these trials and at least a sensitivity analysis otherwise. For almost all of the trials, PANSS total score at baseline was included as a covariate. Mean treatment effect was calculated as the drug-placebo difference, ie, mean change in PANSS total score for the antipsychotic drug group minus mean change for the placebo group. Each trial was rated as a success or failure on the basis of whether it succeeded in showing statistical superiority for at least 1 investigational drug group over the placebo group on change from baseline to endpoint in the PANSS total score after adjusting for multiplicity. The multiplicity adjustment was based on the preplanned analysis procedure that in every instance adequately controlled the overall study-wide type I error rate.

For the purpose of assessing regional variations in treatment effect, these trials were categorized into "North American" trials, with all patients enrolled solely from North America, and "multiregional" trials—clinical trials that were not solely North American trials. To explore the impact of the proportion of North American patients, we further grouped multiregional trials into "North America–predominant multiregional trials" (trials with more than 50%, but < 100%,

Table 1. Demographic and Baseline Characteristics in Schizophrenia Trials[a]

Characteristic	North American Trials (n = 21)	Multiregional Trials[b] North America–Predominant (n = 7)	Multiregional Trials[b] Foreign-Predominant (n = 4)	Multiregional Trials[b] All Multiregional (n = 11)	Overall (N = 32)
ITT patients, no. (%)	6,268 (54.2)	3,046 (26.3)	2,253 (19.5)	5,299 (45.8)	11,567 (100.0)
Age,[c] y	39.3 (1.6)	39.4 (1.5)	36.7 (1.8)	38.4 (2.0)	39.0 (1.8)
Gender,[d] % female	22.2 (6.0)	26.8 (5.4)	40.3 (6.3)	31.7 (8.7)	25.7 (8.4)
Race					
White,[d] %	50.5 (10.1)	48.9 (13.1)	64.2 (15.5)	54.5 (15.3)	52.0 (12.1)
African,[c] %	40.3 (10.9)	37.8 (9.8)	12.7 (12.3)	28.7 (16.2)	35.9 (14.1)
Asian,[e] %	1.3 (0.8)	7.9 (9.0)	21.0 (11.7)	12.7 (11.5)	5.9 (9.1)
Weight,[c] kg	84.5 (4.5)	80.9 (3.2)	71.7 (5.4)	77.6 (6.0)	81.8 (6.1)
BMI,[f] kg/m^2	28.9 (1.4)	27.2 (0.8)	25.1 (1.3)	26.5 (1.4)	27.8 (1.8)
Dropout rate, %					
Overall	49.1 (10.1)	39.9 (6.4)	33.4 (7.9)	37.5 (7.3)	45.1 (10.7)
4-week trials[g]	37.6 (7.6)	33.5 (NA)	NA	33.5 (NA)	36.9 (7.0)
6- or 8-week trials[h]	52.6 (8.0)	41.0 (6.3)	33.4 (7.9)	37.9 (7.6)	47.0 (10.6)
Due to adverse event[d]	6.6 (3.0)	6.2 (2.2)	5.4 (2.6)	5.9 (2.3)	6.3 (2.7)
Due to lack of efficacy[i]	23.2 (11.6)	14.0 (7.8)	15.8 (6.4)	14.6 (7.0)	20.2 (10.9)
Baseline PANSS total score[d]					
Overall	88.7 (11.9)	93.4 (3.5)	93.1 (3.1)	93.3 (3.2)	90.4 (9.8)
4-week trials[g]	93.5 (4.0)	91.6 (NA)	NA	91.6 (NA)	93.2 (3.6)
6- or 8-week trials[d,h]	87.0 (13.4)	93.7 (3.7)	93.1 (3.1)	93.4 (3.3)	89.7 (10.8)

[a]All values except number and percentage of ITT patients (first row of data) expressed as mean (SD). [b]Two multiregional trials were conducted solely in foreign regions, and 9 trials were conducted in a mix of North American and foreign regions; the proportion of North America region data in the latter 9 multiregional trials was 63.0% (SD = 21.0) (range, 31.0%–91.8%). [c]Data missing from 3 North American trials. [d]Data missing from 2 North American trials. [e]Data missing from 5 North American trials. [f]Data missing from 8 North American trials. [g]Five North American trials and 1 mixed multiregional trial were 4-week trials. [h]Thirteen North American, 2 foreign, and 8 mixed multiregional trials were 6-week trials; 3 North American trials were 8-week trials. [i]Data missing from 1 North American trial.

Abbreviations: BMI = body mass index, ITT = intent to treat, NA = not available, PANSS = Positive and Negative Syndrome Scale.

of patients enrolled in North America) and "foreign-predominant multiregional trials" (trials with less than 50% of patients enrolled in North America).

Mean treatment effects and trial success rates were compared with respect to where and when the trials were conducted. Additional potential predictors of treatment effect that were explored included baseline characteristics such as mean baseline PANSS total score and mean baseline body weight, and sample size per arm. In graphical explorations, we presented individual drug-treated groups for trials that studied multiple drug groups when summarizing treatment effects.

RESULTS

Among the 32 schizophrenia trials, 21 trials (68%) were conducted solely in North America and 11 (32%) were multiregional trials. Of the 21 North American trials, 9 contained patients from both the United States and Canada (ranging from 2%–17% from Canada). Of the 11 multiregional trials, 7 were North America–predominant and 4 were foreign-predominant. Of the 4 foreign-predominant multiregional trials, 2 were conducted exclusively in foreign regions outside North America. For the trials included in the analysis dataset, almost 85% (27/32) of the trials used a fixed-dose design. Five North American trials and 1 multiregional trial were 4-week trials, and 3 North American trials were 8-week trials; the remaining 13 North American and 10 multiregional trials were 6-week trials.

Table 1 summarizes the numbers of patients, as well as the demographic characteristics, baseline disease status, and dropout percentages. A total of 11,567 patients were included in our exploratory analyses, with 54% of the patients from those 21 North American trials, 26% from 7 North America–predominant trials, and 20% from the remaining 4 foreign-predominant multiregional trials. The mean age at study entry was 39 years. Overall, 26% of the patients were female; however, a larger proportion of women (40%) were observed in foreign-predominant multiregional trials. There also appeared to be regional differences in mean body weight and BMI, ie, greater mean body weight and BMI in North American trials (85 kg and 29 kg/m^2, respectively) compared to foreign-predominant multiregional trials (72 kg and 25 kg/m^2, respectively). The dropout rate in North American trials (49.1%) was larger than in multiregional trials (37.5%). The most common reason for dropout was lack of efficacy, and the proportion of dropouts due to this reason in North American trials (23%) was also larger than in multiregional trials (15%). The proportions of dropouts due to adverse events were similar across regions. Although the mean baseline PANSS total score appears slightly smaller (ie, less severe symptoms) in North American trials, it was driven by the 3 trials with very low mean baseline PANSS total scores (range, 59–66); without these 3 trials, the mean baseline PANSS total scores were within a range of 87–100 across multiple regions.

The placebo response for each trial was measured by the mean change from baseline in PANSS total score for the placebo arm. As summarized in Table 2, the average placebo response of those 32 trials was −6.4; it was −4.3 for North American trials and −10.0 for multiregional trials. Although the observed placebo responses were generally

 © Copyright 2012 Physicians Postgraduate Press, Inc.

Table 2. Summary of Placebo Response, Drug Response, Treatment Effect, and Trial Success Rate

	North American Trials	Multiregional Trials	All Trials
Placebo response[a,b]			
Entire time span: 1991–2008	−4.3 (−12.6 to 3.6)	−10.0 (−18.8 to −2.5)	−6.4 (−18.8 to 3.6)
Period 1: 1991–1998	−2.3 (−12.6 to 3.6)	NA	−2.3 (−12.6 to 3.6)
Period 2: 1999–2008	−7.0 (−12.3 to 2.3)	−10.0 (−18.8 to −2.5)	−8.7 (−18.8 to 2.3)
Drug response[a,b]			
Entire time span: 1991–2008	−13.1 (−21.3 to −5.4)	−18.0 (−31.3 to −9.4)	−15.0 (−31.3 to −5.4)
Period 1: 1991–1998	−12.8 (−21.3 to −5.4)	NA	−12.8 (−21.3 to −5.4)
Period 2: 1999–2008	−13.4 (−17.0 to −7.0)	−18.0 (−31.3 to −9.4)	−16.2 (−31.3 to −7.0)
Treatment effect[c]			
Entire time span: 1991–2008	−9.0 (−22.2 to 5.1)	−8.1 (−18.9 to 0.5)	−8.6 (−22.2 to 5.1)
Period 1: 1991–1998	−10.8 (−22.2 to −3.6)	NA	−10.8 (−22.2 to −3.6)
Period 2: 1999–2008	−6.0 (−12.8 to 5.1)	−8.1 (−18.9 to 0.5)	−7.2 (−18.9 to 5.1)
Trial success rate[d]			
Entire time span: 1991–2008	17/21 (81.0%)	8/11 (72.7%)	25/32 (78.1%)
Period 1: 1991–1998	11/13 (84.6%)	NA	11/13 (84.6%)
Period 2: 1999–2008	6/8 (75.0%)	8/11 (72.7%)	14/19 (73.7%)

[a]Data not available from 2 North American trials.
[b]Change from baseline in PANSS total score, expressed as mean (range).
[c]Calculated as the mean (range) values for the drug-placebo difference in PANSS total score, ie, drug response – placebo response for each drug-treated group in each trial.
[d]Numerators indicate the number of successful trials, denominators indicate the total number of trials. Success rates are expressed as percentages in parentheses.
Abbreviations: NA = not available, PANSS = Positive and Negative Syndrome Scale.

smaller in North American trials, many of those trials were conducted in relatively early years, as indicated in Figure 1A. This figure also suggests increasing placebo responses over time regardless of where trials were conducted. Figure 1B describes the drug responses (ie, the mean change from baseline in PANSS total scores in the drug-treated groups) over time. Because each symbol represents response in each drug-treated group (per arm) and there were multiple drug-treated groups in some trials, there are more symbols in this plot than in the placebo response plot. This figure suggests relatively stable drug responses over time except for the 4 multiregional trials conducted in latter years with relatively larger drug responses; 2 of these 4 were conducted solely in foreign countries during the years 2004–2005 (each with 4 drug-treated groups), and the other 2 were conducted during the years 2008–2009 (each with 3 drug-treated groups). The impact of these 4 multiregional trials is reflected in Figure 1C, a display of treatment effect (ie, drug response minus placebo response for each drug-treated group in each trial) over time.

Figure 1C reveals substantially diminishing treatment effects over time for the North American trials (the result of stable drug response and increasing placebo responses), with relatively large treatment effects for a few foreign-predominant multiregional trials. Table 2 also indicates that the treatment effect in North American trials has declined over time from −10.8 PANSS units for the first period (1991–1998) to −6.0 PANSS units for the second period (1999–2008).

As displayed in Table 2, the overall trial success rate (at least 1 dose significantly better than placebo) over the almost 2 decades of observation was 78%. The success rate was higher (85%) for trials conducted before 1998 when no multiregional trials were included compared to those conducted later (74%). Similar results were observed when considering only North American trials (85% vs 75%). During the second period, the success rate of North American trials was similar to that of multiregional trials (75% vs 73%).

Potential impacts of several predictive factors on treatment effect are presented in Figures 2 and 3. One exploration focused on treatment effect and sample size per arm. Sample size allocation was equal across arms in all but 2 trials. Figure 2A shows a trend for the sample size per arm increasing over time, but this factor did not appear to have an impact on the observed treatment effect size over time (Figure 2B). Sample size per arm also did not affect trial success rate, which remained within a narrow range of 75%–80%.

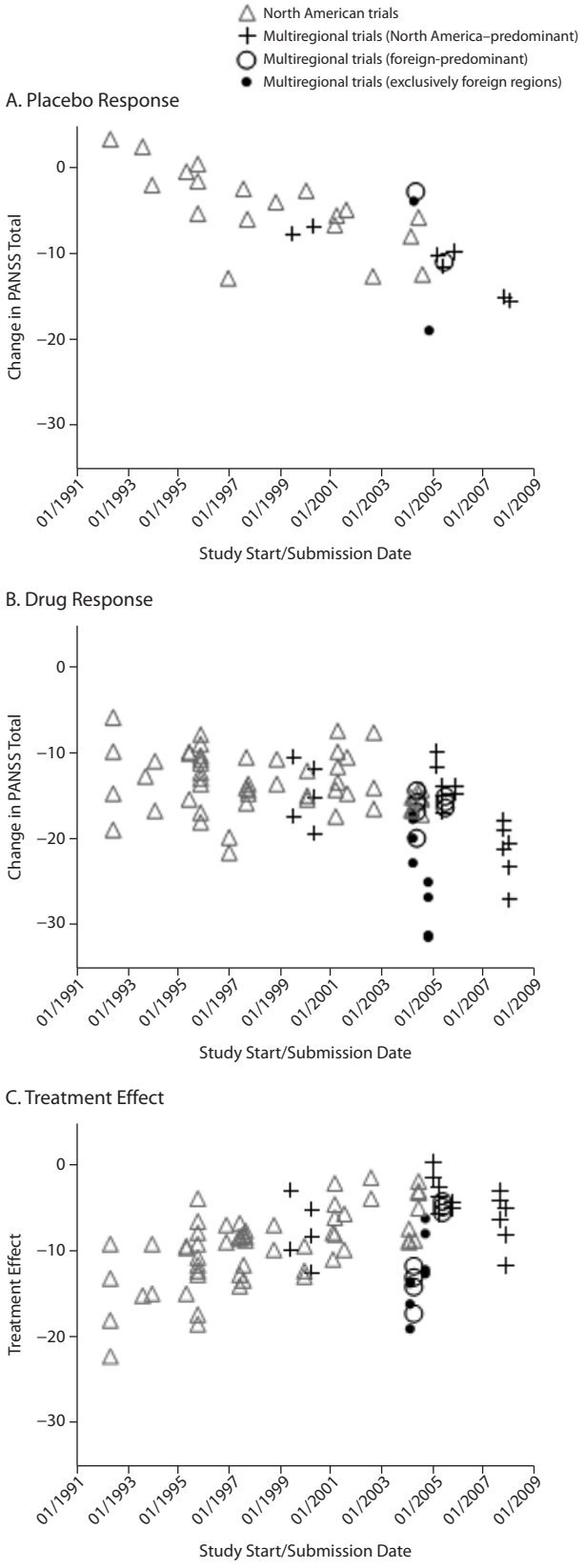
Except for a small cohort of 3 North American trials conducted in very early years, mean baseline PANSS total scores were in a narrow range over time (87–100) regardless of whether the trials were North American or multiregional (Figure 3A). Overall, the treatment effect of the small cohort (note that the 3 trials had a total of 11 treatment groups) was similar to that of the large cohort. When the 2 cohorts were considered separately, trials in patients with more severe symptoms at baseline tended to have greater improvement (Figure 3B).

Mean body weights increased over time in North American trials (range, 77 to 93 kg), while mean body weights in multiregional trials remained relatively stable. Weights were lower in multiregional trials than in North American trials, especially in the 4 foreign-predominant multiregional trials that reported a lower mean body weight (range, 64–76 kg) (Figure 3C). Treatment effects tended to decrease as body weights increased, especially in North American trials (Figure 3D).

All trials conducted in the earlier period (1991–1998) were North American trials, and these ranged in duration from 4 to 8 weeks. The mean treatment effect was largest (−13.3) for 8-week trials compared with an mean of −9.0 for 6-week and −10.0 for 4-weeks trials. All trials but 1 conducted in the latter period (1999–2008) were 6-week trials, and among those, the mean treatment effect was larger (−11.4) for foreign-predominant multiregional trials compared with an mean of −6.0 to −5.4 for North American and North America–predominant multiregional trials.

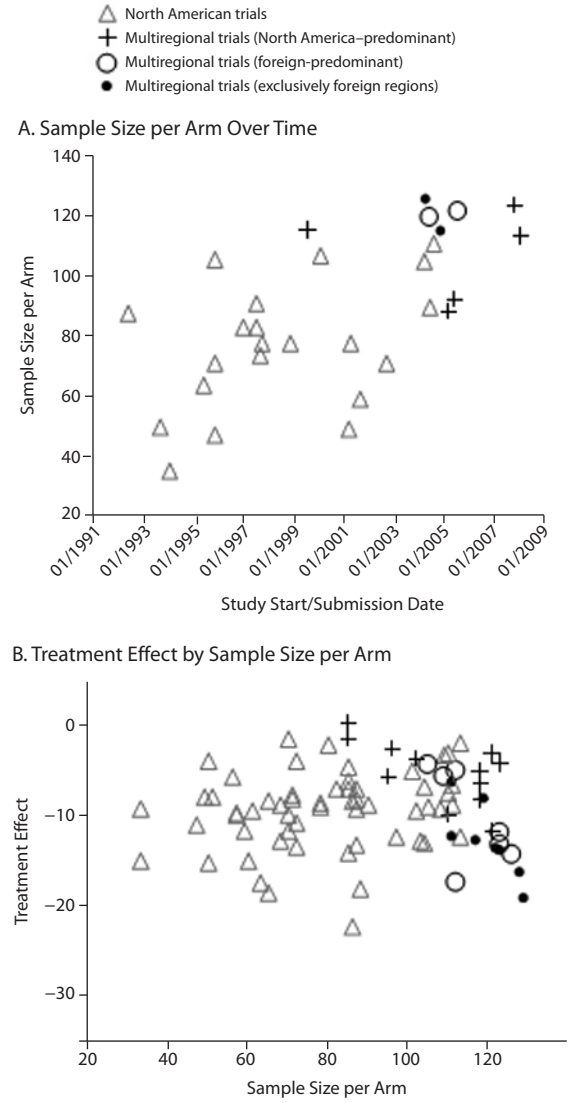
The percentage of dropouts decreased over time in schizophrenia trials (Figure 3E) in both the placebo and the drug-treated groups, but the dropout rates were higher in the placebo groups (mean of 52%; range, 26%–80%) compared to the drug-treated groups (mean of 42%; range, 17%–67%). Among drug-treated groups, the dropout percentages in

Figure 1. Placebo Response, Drug Response, and Treatment Effect Over Time[a]



[a]Treatment effect was calculated as the drug-placebo difference in PANSS total score, ie, drug response – placebo response for each drug-treated group in each trial.
Abbreviation: PANSS = Positive and Negative Syndrome Scale.

Figure 2. Impact of Sample Size on Treatment Effect[a,b]



[a]Two trials where randomization ratio was not equal across treatment arms were removed. Figure 2A was based on sample size per placebo arm, and 2B included all drug-treatment arms. [b]Treatment effect was calculated as the drug-placebo difference in PANSS total score, ie, drug response – placebo response for each drug-treated group in each trial.
Abbreviation: PANSS = Positive and Negative Syndrome Scale.

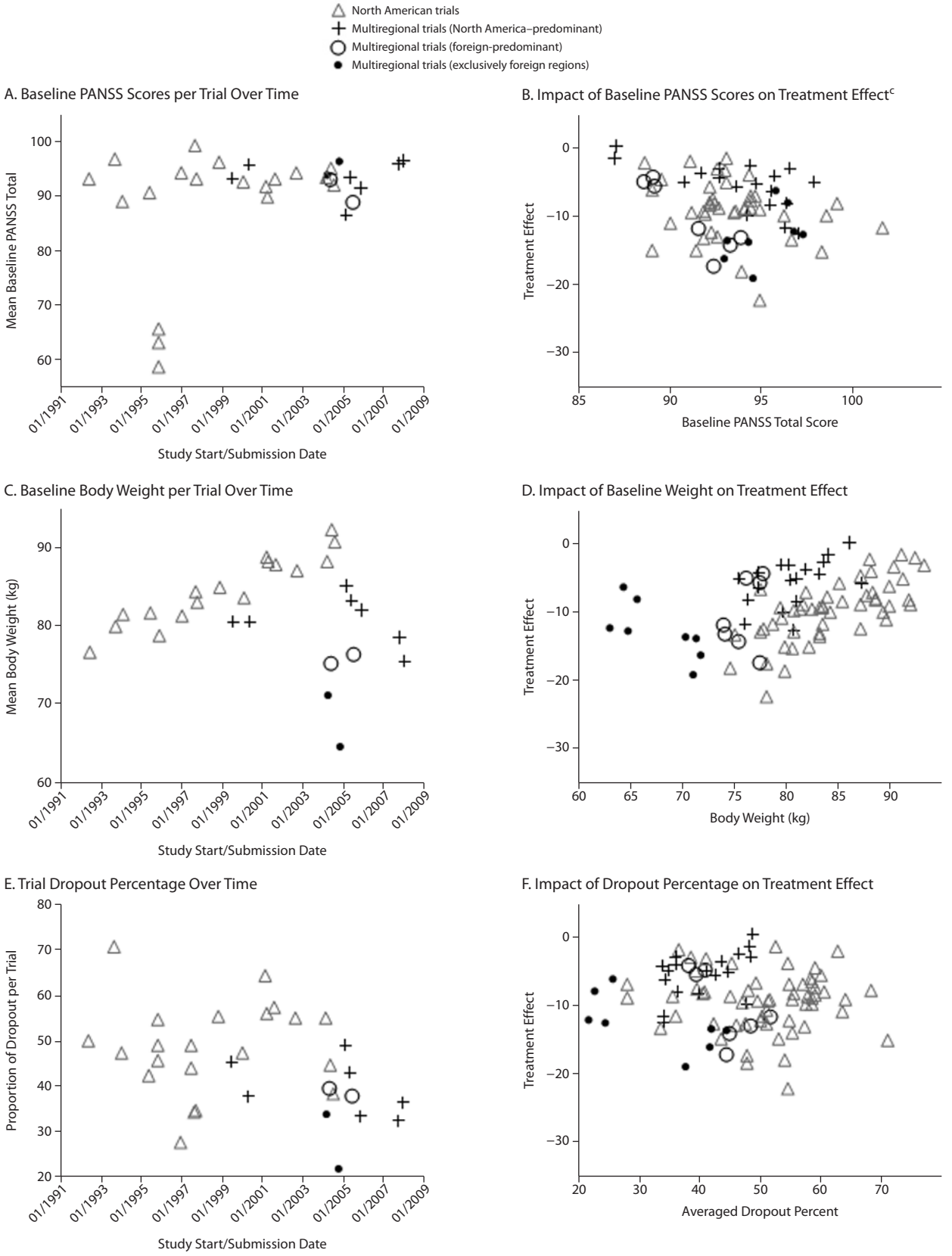
those few foreign-predominant multiregional trials were relatively lower compared with other trials. Because of differential dropout percentages between the drug-treated and placebo-treated groups within each trial (mean difference of about 10%), we explored the impact of the mean dropout percentage of the placebo arm and the associated drug-treated arm on treatment effect and found no clear impact on treatment effect in North American trials, but in multiregional trials, higher percentages of dropouts were associated with smaller treatment effects (Figure 3F).

DISCUSSION

The mean difference in change from baseline on the PANSS total score between second-generation antipsychotics

© Copyright 2012 Physicians Postgraduate Press, Inc.

Figure 3. Impact of Baseline PANSS Total Score, Baseline Body Weight, and Percentage of Dropout[a,b]



[a]Dropout percentage averaged over the placebo arm and the associated drug arm. [b]Treatment effect was calculated as the drug-placebo difference in PANSS total score, ie, drug response – placebo response for each drug-treated group in each trial. [c]Excluding the cohort of 3 trials with low mean baseline score range of 60–70.
Abbreviation: PANSS = Positive and Negative Syndrome Scale.

and placebo in a recent meta-analysis of controlled trials based on published literature was about 10 PANSS units.[6] In our analysis, responses in both placebo and drug-treated groups from multiregional trials tended to be larger than those observed in the North American trials; considering all trials, mean treatment effects (drug-placebo differences) were approximately −9 and −8 PANSS units for North American trials and multiregional trials, respectively. This finding is also similar to that in our previous publication[20] regarding major depressive disorder (MDD) in that the overall treatment effect of 2.5 Hamilton Depression Rating Scale units was observed for both North American and non–North American regions. However, it should be noted that the majority of MDD trials were conducted in the United States over the last 2 decades, while nearly half of the schizophrenia trials were conducted as multiregional trials, particularly in the most recent decade.

With regard to the effects over time, however, treatment effect sizes in North America were considerably smaller since 1999 (Figure 1C and Table 2) in schizophrenia trials. In a previous analysis based on patient-level data from trials selected using different selection criteria and methodology, and in which patients were grouped into the regions where they were enrolled,[8] we reported somewhat larger mean treatment effects (about −12 PANSS units) in studies from Asian and Eastern European regions (about 9% and 21% of the trial population) compared to a mean treatment effect of about −6 PANSS units in North America and Western Europe (about two-thirds of the trial population in that database).

Karagianis et al[21] reported similar baseline characteristics for most measures in schizophrenia patients enrolled from various regions in a large observational study. In contrast, we have observed cross-regional variations on important baseline characteristics. We found a higher mean body weight (85 kg) and BMI (29) at baseline in North American trials versus a mean body weight of 72 kg and mean BMI of 25 for patients enrolled in foreign-predominant multiregional trials. We found a higher proportion of women (40%) in foreign-dominated multiregional trials compared to 22% and 27%, respectively, in North American and North America–predominant multiregional trials. We also observed a relatively larger proportion of Asian patients (21%) and a lower proportion of Africans (13%) in foreign-predominant multiregional trials. The North American trials enrolled 1.3% Asians and 40% African Americans. Caucasian patients accounted for slightly more than 50% of the patients for both North American and multiregional trials.

It has been challenging to try to understand the decrease in treatment effect in psychiatric trials over time. High placebo response has been considered a major factor contributing to the substantial failure rate observed in schizophrenia trials.[7] We have also observed an increase in placebo response over time in schizophrenia trials, with stable response in the drug group leading to a decrease in treatment effect in North American trials. There is, however, less experience with multiregional trials, which did not begin until the late 1990s.

The overall trial success rate over the last 2 decades for schizophrenia trials in our current database was 78%, similar to results reported earlier.[5] When broken out by early and late time periods of trial conduct, the overall success rate was found to be slightly higher for trials conducted during the earlier time period (1991–1998), all of which were North American trials, compared to those conducted in the later period (1999–2008), which were mostly multiregional trials (85% vs 74%, respectively). We have observed both a rise in placebo response and a decline in treatment effect over time with a higher failure rate (50%) in MDD trials.[20] For the schizophrenia trials included in this analysis dataset, almost 85% of the trials used a fixed-dose design, as compared to two-thirds of the MDD trials, which utilized a flexible dosing regimen.

Sponsors have been using the placebo lead-in design as one approach to try to identify and exclude placebo responders in psychiatric trials. While this has not been a successful strategy in lowering the placebo response rate or increasing drug-placebo differences in depression trials,[22–24] it was recently noted that a longer duration (up to 2 weeks) of placebo run-in was associated with a smaller placebo response in schizophrenia trials as well as in MDD trials.[25] An alternative enrichment design, ie, the sequential parallel comparison design (SPCD), has been proposed as an approach to minimizing placebo response in psychiatric trials.[26,27] This design is different from a traditional parallel-group design in that it contains 2 double-blind treatment periods of the same duration. Specifically, in the first period, patients are randomly assigned to either study drug or placebo, but only those patients who are randomly assigned to placebo in the first period and do not achieve sufficient response at the end will be further re-randomized in the second period. Treatment effect would then be assessed based on data from both periods. The applicability of some common statistical analyses in this SPCD design has been evaluated.[25] Further evaluation of the strengths and weaknesses of this study design and its implications will be needed.

It is a widely held view that enrolling more severely ill patients with higher baseline scores is associated with greater clinical improvement with treatment. Targeting sicker patients by setting higher thresholds for enrolling patients might indeed be expected to increase trial success rates, but this approach must be carefully used because any benefit could be offset by score inflation at study sites, a serious concern in study conduct that would be likely to lead to larger placebo group responses.[28] Previously published results for schizophrenia trials do, however, suggest that enrolling patients with higher mean baseline PANSS total scores has been associated with larger treatment effects.[8] In our database, the mean baseline PANSS total scores were in a narrow range of 87–100 over time, except for 3 trials conducted in very early years that had very low mean baseline scores, so that a large effect would not be expected. We have, nevertheless, observed a modest trend for trials with higher mean baseline PANSS total scores to have larger treatment effects, as displayed in Figure 3B.

Several methodological approaches have been suggested to address the problem of high dropout rates and poor quality

© Copyright 2012 Physicians Postgraduate Press, Inc.

of ratings in clinical trials.[7,8,29,30] Use of centralized ratings has been proposed as one of the potential solutions to the problem of rating quality.[31,32] The overall mean dropout rate in our schizophrenia database was high (45%), with an almost 50% dropout rate for North American trials. The mean dropout rate in multiregional trials was lower, around 37%. A recent study explored several design features of schizophrenia trials for an association with dropout and identified only longer duration as consistently associated with dropout.[30] An association was not found for publication year, presence of placebo arm, or symptom level. For North American trials, we also found that longer trials were associated with a higher dropout percentage. It has been reported that, in depression trials, larger studies do not necessarily produce a better outcome.[33] We observed that larger sample size per arm was associated with higher trial success rate, but only for multiregional trials. We also noted that a smaller proportion of placebo patients (generally the result of multiple dosage groups) was associated with a larger treatment effect, but, again, only for multiregional trials.

Global drug development is inevitable, and continued efforts are needed to try to understand some regional and time differences between findings from North American and other foreign regions in schizophrenia trials. Some important factors such as access to medical care and exposure to prior medications, which might be factors potentially impacting regional differences in schizophrenia trials, were not captured in our database. Patient compliance to study drug is one of the major issues not addressed in the article. The consideration of pharmacokinetic data is beyond the scope of this article.

The increase in placebo response and decrease in treatment effect over time in North American schizophrenia trials remain a concern. Great care is needed in designing and conducting multiregional studies in schizophrenia, and close attention should be paid to possible differences in patient population, disease severity, diagnostic practices, and clinical care practices, including the use of concomitant medication.

Drug names: Trial data from all new drug applications for an antipsychotic claim regardless of approval status were included. For reasons of confidentiality, individual drugs were not reported here.

Author affiliations: Division of Psychiatry Products, Office of Drug Evaluation I, Office of New Drugs (Drs Khin and Laughren); and Division of Biometrics I, Office of Biostatistics, Office of Translational Science (Drs Chen, Y. Yang, and P. Yang), Center for Drug Evaluation and Research (CDER), US Food and Drug Administration (FDA), Silver Spring, Maryland. Dr Y. Yang is now employed by Santen, Inc.

Potential conflicts of interest: None reported.
Funding/support: None reported.

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily represent the views of FDA.

Previous presentations: A preliminary analysis of this work was presented at the following professional society meetings: the FDA symposium at the 49th Annual New Clinical Drug Evaluation Unit (NCDEU) Meeting; June 29–July 2, 2009; Hollywood Beach, Florida; the Joint American and Korean Society of Clinical Pharmacology and Therapeutics (KSCPT/ASCPT) meeting; December 1–3, 2009; Seoul, South Korea; the Pharmacokinetics United Kingdom (PKUK) annual meeting; November 1–3, 2010; Bristol, United Kingdom; and the Drug Information Agency (DIA)–sponsored Multi-Regional Clinical Trials workshop sessions; October 27, 2010, and March 9, 2011; Bethesda, Maryland.

Acknowledgments: The authors thank Robert Temple, MD, and H. M. James Hung, PhD, CDER/FDA, for their critical review of the manuscript. Drs Temple and Hung report no potential conflict of interest.

REFERENCES

1. Karlberg JPE. Trends in disease focus of drug development. *Nat Rev Drug Discov*. 2008;7(8):639–640.
2. Regier DA, Narrow WE, Rae DS, et al. The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50(2):85–94.
3. Jablensky A. The 100-year epidemiology of schizophrenia. *Schizophr Res*. 1997;28(2-3):111–125.
4. Mueser KT, McGurk SR. Schizophrenia. *Lancet*. 2004;363(9426): 2063–2072.
5. Laughren TP. The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. *Eur Psychiatry*. 2001;16(7):418–423.
6. Leucht S, Arbter D, Engel RR, et al. How effective are second-generation antipsychotic drugs? a meta-analysis of placebo-controlled trials. *Mol Psychiatry*. 2009;14(4):429–447.
7. 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(3):504–509.
8. Chen YF, Wang SJ, Khin NA, et al. Trial design issues and treatment effect modeling in multi-regional schizophrenia trials. *Pharm Stat*. 2010; 9(3):217–229.
9. Thiers FA, Sinskey AJ, Berndt ER. Trends in the globalization of clinical trials. *Nat Rev Drug Discov*. 2008;7(1):13–14.
10. US Department of Health and Human Services. Code of Federal Regulations, Title 21-Food and Drugs, Part 314, Applications for FDA Approval to Market a New Drug. 21 CFR Part 314.106: Foreign Data. http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr;sid=c4683e3b7214a3bef6f567bc293ff57c;rgn=div5;view=text;node=21%3A5.0.1.1.4;idno=21;cc=ecfr#21:5.0.1.1.4.4.1.7. Revised April 1, 2011. Updated July 29, 2011. Accessed August 2, 2011.
11. US Department of Health and Human Services. Code of Federal Regulations, Title 21-Food and Drugs, Part 312, Investigational New Drug Applications. 21 CFR 312.120: Foreign Clinical Studies Not Conducted Under an IND. http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=d199e7693c50a9bb0de86c4b652807c7&rgn=div8&view=text&node=21:5.0.1.1.3.6.1.2&idno=21. Revised April 1, 2011. Updated September 7, 2011. Accessed Sept 9, 2011.
12. Topic ICH. E5 (R1). Ethnic Factors in the Acceptability of Foreign Clinical Data. CPMP/ICH/289/95. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002842.pdf. Published September 1998. Accessed August 3, 2011.
13. Topic ICH. E6 (R1). Good Clinical Practice. Finalized Guideline: May 1996. http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html. Accessed October 27, 2011.
14. Temple R. Use of non-US data in NDAs. Special session: Regulatory Considerations of Using Non-US Data in NDAs: Focus on Efficacy, Safety and Clinical Pharmacology. Presented at the Annual Meeting of the American Society for Clinical Pharmacology and Therapeutics; March 20, 2009; Washington, DC. http://www.ascpt.org/Portals/8/docs/Meetings/2009%20Annual%20Meeting/Friday/Regulatory%20Considerations.pdf. Accessed March 10, 2011.
15. Committee for Medicinal Products for Human Use (CHMP). European Medicines Agency (EMEA): reflection paper on the extrapolation of results from clinical studies conducted outside Europe to the EU-population. EMEA/CHMP/EWP/692702/2008: 1–7. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/11/WC500013468.pdf. Published February 19, 2009. Updated October 22, 2009. Accessed March 10, 2011.
16. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276.
17. Kay SR, Opler LA, Lindenmayer JP. Reliability and validity of the Positive and Negative Syndrome Scale for schizophrenics. *Psychiatry Res*. 1988;23(1):99–110.
18. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep*. 1962;10:799–812.
19. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer-Verlag; 2000.
20. Khin NA, Chen YF, Yang Y, et al. Exploratory analyses of efficacy

data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J Clin Psychiatry*. 2011;72(4):464–472.
21. Karagianis J, Novick D, Pecenak J, et al. Worldwide-Schizophrenia Outpatient Health Outcomes (W-SOHO): baseline characteristics of pan-regional observational data from more than 17,000 patients. *Int J Clin Pract*. 2009;63(11):1578–1588.
22. Trivedi MH, Rush H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology*. 1994;11(1):33–43.
23. Faries DE, Heiligenstein JH, Tollefson GD, et al. The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *J Clin Psychopharmacol*. 2001;21(6):561–568.
24. Lee S, Walker JR, Jakul L, et al. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? a meta-analytic evaluation. *Depress Anxiety*. 2004;19(1):10–19.
25. Chen YF, Yang Y, Hung HMJ, et al. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemp Clin Trials*. 2011;32(4):592–604.
26. Fava M, Evins AE, Dorer DJ, et al. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom*. 2003;72(3): 115–127.
27. Tamura RN, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clin Trials*. 2007;4(4):309–317.
28. Landin R, DeBrota DJ, DeVries TA, et al. The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics*. 2000; 56(1):271–278.
29. Leucht S, Heres S, Hamann J, et al. Methodological issues in current antipsychotic drug trials. *Schizophr Bull*. 2008;34(2):275–285.
30. Rabinowitz J, Levine SZ, Barkai O, et al. Dropout rates in randomized clinical trials of antipsychotics: a meta-analysis comparing first- and second-generation drugs and an examination of the role of trial design features. *Schizophr Bull*. 2009;35(4):775–788.
31. Kobak KA, Kane JM, Thase ME, et al. Why do clinical trials fail? the problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol*. 2007;27(1):1–5.
32. Shen J, Kobak KA, Zhao Y, et al. Use of remote centralized raters via live 2-way video in a multicenter clinical trial for schizophrenia. *J Clin Psychopharmacol*. 2008;28(6):691–693.
33. Liu KS, Snavely DB, Ball WA, et al. Is bigger better for depression trials? *J Psychiatr Res*. 2008;42(8):622–630.