Feasibility of Reducing the Duration of Placebo-Controlled Trials in Schizophrenia Research

Robert P. McMahon², Deanna L. Kelly², Douglas L. Boggs², Lan Li³, Qiaoyan Hu⁴, John M. Davis⁴, and William T. Carpenter, Jr.¹,²,⁵

¹Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland Baltimore Medical School, Baltimore, MD; ²Department of Psychiatry, University of Maryland Baltimore Medical School, Baltimore, MD; ³Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; ⁴Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; ⁵VISN 5 Mental Illness Research, Education and Clinical Center, Veterans Administration, Baltimore, MD

Use of placebo-controlled trials in medical and psychiatric research has been controversial, although a consensus is emerging about conditions under which placebo-controlled trials are ethical. In schizophrenia research, the paradigm of slow onset of antipsychotic effects has led to a model in which placebo-controlled trials of 6–8 weeks duration have been used to demonstrate efficacy. Recent evidence that the largest symptom reductions are typically seen in the first weeks of treatment suggests that shorter placebo-controlled studies to demonstrate antipsychotic efficacy are possible. In a pilot study of the feasibility of shortening placebo-controlled studies, we reanalyzed data from placebo-controlled registry trials of olanzapine and risperidone and found that trials as short as 4 weeks could have similar power to longer term 6–8 week studies, given the estimated time course of treatment effects. Although fuller evaluation is required, the results suggest future placebo-controlled studies to demonstrate antipsychotic efficacy are possible. In a pilot study of the feasibility of shortening placebo-controlled studies, we reanalyzed data from placebo-controlled registry trials of olanzapine and risperidone and found that trials as short as 4 weeks could have similar power to longer term 6–8 week studies, given the estimated time course of treatment effects. Although fuller evaluation is required, the results suggest future placebo-controlled trials could be shortened from 6–8 weeks to 3–4 weeks with a relatively low increase in sample size requirements. Shortening placebo-controlled trials would reduce patient burden and ethical objections to prolonged administration of placebo and reduce potential bias due to high dropout rates in longer clinical trials.

Key words: antipsychotic/research design/ethics

Background

The placebo-controlled trial is the standard method to demonstrate efficacy and safety of antipsychotic drugs for the treatment of schizophrenia. Currently, superiority over placebo is the only acceptable and reliable proof of efficacy of new antipsychotic medications.¹,² Suggested alternatives cannot prove that improvements in symptoms do not represent nonspecific treatment effects or the natural course of the disease. Also, in trials intended to show that a new drug is equivalent to established drugs, a variety of problems³,⁴ may reduce power to detect clinically important differences between the new drug and the established agents. Most importantly, without a placebo comparator, we cannot tell whether the apparent equivalence of standard and novel antipsychotic has occurred because the clinical population recruited for a study is nonresponsive to both drugs. Also, if the experimental drug is ineffective and/or unsafe, the larger sample size required for noninferiority testing may expose more patients to harmful side effects or ineffective treatment compared with placebo-controlled designs. Studies which fail to conclusively establish efficacy of medications risk exposing future patients to treatments that offer no benefit and some potential harm.

Nevertheless, over the past decade there has been increasing attention and concern regarding the ethics of conducting placebo-controlled trials. Patients randomized to placebo in double-blind clinical trials would be foregoing an aspect of accepted treatments⁵ and may experience symptom deterioration or delayed recovery⁶ and suboptimal treatment for periods of time ranging from 4 to 12 weeks in length.¹ Further, some have questioned whether patients truly understand the potential consequences of placebo treatment when giving informed consent.⁷,⁸ The growing ethical controversy over the use of placebo in various fields of medicine⁹,¹⁰ led in part to the latest revisions to the Declaration of Helsinki. Changes to this international accepted code of ethics were made to help clarify specific circumstances when placebo-controlled trials are ethically acceptable.⁹,¹¹ Nevertheless, the field continues to struggle with the ethical dilemmas surrounding placebo-controlled trials of people with schizophrenia.¹² While placebo-controlled
trials are necessary at times, it is desirable to minimize the duration of exposure to ineffective treatment.

In addition to these ethical concerns, recruitment into placebo-controlled trials is often challenging and may prolong study duration and costs. While the FDA in the United States continues to support placebo-controlled registration trials for antipsychotic drugs, Institutional Review Boards, particularly in developed countries, limit their approval of placebo trials. This has led to most premarketing studies being completed in international patient samples, which may be less generalizable to the United States and European Community population. Fleischhacker and Burns reported less than one-third of schizophrenia researchers polled stated they would participate in an acute placebo-controlled trial. Additionally, over half of all outpatients with schizophrenia surveyed said they would not be willing to participate in a placebo-controlled trial. More recently, Roberts and colleagues reported an inverse relationship between perceived potential for harm and willingness to participate in schizophrenia research protocols and that placebo treatment was perceived as moderately harmful. Mohr and Czobor examined the hypothesis that inclusion of a placebo treatment arm would bias the selection of study subjects. They reported that based on a total of 296 studies, participants within placebo-controlled trials were on average significantly older and had a longer duration of illness and lower symptom severity than participants in active comparator-controlled trials.

Study dropout rates are reported to be higher in studies with a placebo arm as compared with comparator-controlled clinical trials. Likewise, a recent meta-analysis of trials of second-generation antipsychotics reported weighted mean dropout rates in the active treatment arms were significantly higher in placebo-controlled trials than in active control trials (48.1% vs 28.3%). Higher placebo dropout rates, which may disproportionately remove patients who are experiencing lack of efficacy or symptom exacerbations, make it more difficult to develop unbiased estimates of treatment effects.

Designing placebo-controlled antipsychotics trials that minimize the duration of exposure to potentially ineffective treatment requires an understanding of how rapidly substantial antipsychotic effects will happen. The paradigm of a substantial delayed onset between the start of antipsychotics treatment and the observation of clinically significant treatment effects has influenced psychiatric research and treatment recommendations for decades and is reflected in clinical trial designs requiring at least 6–8 weeks of double-blind treatment to attain the study end point. The delayed onset paradigm is contrary to clinical experience and has not been supported in recent analyses. In 2003, Agid et al conducted a meta-analysis of 42 double-blind clinical trials to compare symptom reductions observed for placebo (262 patients) vs various doses of olanzapine (3750 patients), haloperidol (2447 patients), risperidone (896 patients), and chlorpromazine (95 patients). Without significant variation across antipsychotics, they found—contrary to expectations under the delayed onset hypothesis—that relative reductions in total symptom scores and psychosis scores in the first 2 weeks were more than double those observed in the next 2 weeks. This observation of early onset of symptom response to D2 antagonist antipsychotics subsequently has been extended to other agents in this class: quetiapine (Small et al meta-analysis of 3 placebo-controlled trials demonstrating detectable reductions in Brief Psychiatric Rating Scale (BPRS) total score and psychosis score at week 1), amisulpride (Leucht et al pooled analysis of individual patient data from 7 active-controlled studies comparing amisulpride to control groups receiving risperidone, haloperidol, or flupenthixol, but not placebo, showing a time vs symptom reduction curve similar to Agid et al but extending the findings to 1 year); and ziprasidone (Daniel et al, 1999, placebo-controlled trial). This cumulative evidence suggests that early onset of action is a predictable property of D2 antagonists as a class. Interestingly, early trials of aripiprazole, a D2 partial agonist, and LY2140023 a selective agonist of the metabotropic glutamate 2/3 receptors also showed detectable onset of effects on total and positive symptom scores within 2 weeks, with effects similar to comparator antipsychotics at 4 weeks.

Because the emerging body of evidence suggests that antipsychotics have an earlier onset of antipsychotic efficacy than previously thought, it may be feasible to shorten placebo-controlled antipsychotic trials, at least for agents that act through D2 receptor antagonism. Before such a change is accepted as feasible, it is necessary to establish that the early antipsychotic effects can be detected without recruitment of excessively large numbers of patients. The purpose of this article is to conduct a retrospective power analysis, using individual patient data from placebo-controlled trials of olanzapine, risperidone, and haloperidol, examining the impact on sample size requirements if the duration of follow-up were shortened. This analysis is intended to offer initial pilot data to open discussion of changing our model for clinical trials to establish antipsychotic efficacy. Further reanalyses of multiple placebo-controlled studies with a variety of antipsychotic agents will be needed to confirm our findings on power and to address important issues such as the variability in placebo response and individual symptom response to active treatment and how well it can be assessed in short-duration studies.

Methods

Study Overview

To assess how changing the duration might affect the power of placebo-controlled trials, we performed case
studies using individual patient data from 3 large double-blind, placebo-controlled registry trials of risperidone and olanzapine, to examine: (1) the magnitude of treatment effect differences by study week; (2) the distribution of time to response by placebo and antipsychotic drug; and, (3) how shortening the duration of these trials would have affected the sample sizes needed to detect the differences observed. Intensive analysis of individual patient data from a few trials, rather than a meta analysis of published data across a greater number of trials and antipsychotics, was selected as the basis for this power analysis in order to (1) estimate within-patient correlation parameters required for power analysis which are not available in published data and (2) consider statistical analysis methods not employed in the original trial reports.

Data Included

Risperidone data were analyzed using the combined North American double-blind risperidone trials\textsuperscript{28,29} (2 trials, 1 with 6 sites in Canada, 1 with 19 sites in the United States) and the North American double-blind olanzapine trial\textsuperscript{30} (3 sites in Canada, 20 in the United States). The 2 risperidone trials had identical protocols with 8 weeks of double-blind treatment, with symptom assessments at baseline and weeks 1, 2, 4, 6, and 8. Patients with DSM-III-R diagnoses of schizophrenia and total scores on the Positive and Negative Syndrome Scale (PANSS)\textsuperscript{28,29} between 60 and 120 were randomly assigned to double-blind treatment with placebo, haloperidol 20 day, or 1 of 4 doses of risperidone (2, 6, 10, 12.5, 15.0, 17.5 mg/day). Patients completed a 1-week placebo washout phase before starting double-blind treatment.

The North American double-blind olanzapine trial\textsuperscript{30} had symptom assessments weekly during 6 weeks of double-blind treatment. Patients were randomly assigned to placebo, haloperidol (10, 15, or 20 mg/day), or 1 of 3 dose ranges of olanzapine (L = 2.5, 5.0, 7.5; M = 7.5, 10.0, 12.5, or H = 12.5, 15.0, 17.5 mg/day). Patients on active treatment started at the midpoint of their assigned dose range, and the dose was adjusted up or down within the range as clinically indicated. For simplicity, we will identify the olanzapine treatment groups by their initial doses (5, 10, or 15 mg/day). Patients were required to meet DSM-III-R criteria for schizophrenia with an acute exacerbation, established by interview and chart review, plus a minimum BPRS\textsuperscript{30} score of 24. Patients were required to have been withdrawn from their neuroleptics for at least 2 days and to complete a 4- to 7-day placebo run-in before starting double-blind treatment.

In keeping with recommended practice for analysis of mixed models for repeated measures,\textsuperscript{31} data analysis was restricted to the observed data, without imputation of scores for dropouts or missing ratings. The data on mean scores and differences by treatment and study week presented here therefore differ from those presented in Chouinard et al.,\textsuperscript{28} Marder et al.,\textsuperscript{29} and Beasley et al.,\textsuperscript{30} which used the last observation carried forward method for imputing missing values. In the risperidone trial, symptom data were collected on the PANSS. For comparability to the olanzapine trial, PANSS items corresponding to the BPRS 18-item total score and the BPRS psychosis factor (conceptual disorganization, hallucinations, delusions, and suspiciousness) were used to calculate BPRS scores in the risperidone trial. Data are presented in the text for results from the BPRS total score from both trials, with Supplementary Material on BPRS psychosis scores.

All patients in these studies or their legal guardians provided written informed consent. The study protocols were approved by the Institutional Review Boards of the participating institutions.

Magnitude of Mean Treatment Differences

At each week, the mean change from baseline in BPRS total score was estimated for the 2 treatment groups, and then the average difference between treatments in the size of these changes, week by week. For each study, the week with the largest between-treatment mean difference in change from baseline was identified, and mean differences for other weeks were expressed as a percentage of that maximum. Mean change from baseline by week and treatment was estimated using the mixed model: $D_{ikt} = \text{week}_t + \text{treatment}_i + \text{week}_t \times \text{treatment}_i + \text{error}$, where $D_{ikt}$ is the change from baseline for the $i$-th participant in treatment group $k$ at week $t$, $\text{error}$ is an indicator variable for the $k$-th week of follow-up, $\text{treatment}_i$ an indicator for active treatment vs placebo, and the error term is estimated from an unstructured covariance matrix. Because loss to follow-up is commonly associated with worsening of symptoms, it is assumed that the means of the observed data will be biased estimators of the true treatment effects expected at each week. The mixed model estimates adjust the observed means at each week for the expected values for participants with missing observations, conditional upon their prior observed values, and thus may reduce this bias.\textsuperscript{20}

Distribution of Time to “Response”

Response was defined as a $\geq 20\%$ reduction in symptom scores relative to baseline score. The Kaplan-Meier method\textsuperscript{32} of survival analysis was used to analyze the cumulative distribution of time until patients first achieved these response criteria for the BPRS total score. Using this method, the cumulative probability of patients achieving response by visit $k$ is estimated successively for each new visit as $1 - \left[ \left( 1 - d_k/n_k \right) \left( 1 - d_{k-1}/n_{k-1} \right) \ldots \left( 1 - d_1/n_1 \right) \right]$, where $d_k$ is the number of patients who are new responders at visit $k$ and $n_k$ is the total number of patients observed at visit $k$ who have not yet responded at previous visits, and the product of all the terms inside

Downloaded from https://academic.oup.com/schizophreniabulletin/article-abstract/34/2/292/1925547 by guest on 07 March 2019
the brackets [·] is the estimated probability of failing to respond at all visits up to \( k \). Limitations to this analysis include (1) this only estimates the time to 1st response by visit \( k \), not the probability of remaining in response at that visit and (2) the method assumes patients remaining in the study at week \( k \) are a random sample of all participants. If dropouts who have not yet responded have a lower probability of meeting the response criterion after dropping out, then the estimates of the cumulative probability of response will have an increasingly “optimistic” bias as the percentage of dropout increases. For this reason, the estimated cumulative percentages of patients who had responded by later stages of the trials should be treated with caution.

**Sample Size Estimation**

We used statistical methods developed by Hedeker et al\(^{33}\) modified appropriately to allow different dropout rates between groups, to estimate power for mixed models for repeated measures with dropouts over time. These methods require (1) specification of the mean differences between treatments at each time point; (2) specification of the variances of the study outcome at each time point, and the within-patient correlations among time points; (3) specification of the dropout pattern within each treatment group; and (4) expressing the primary measure of treatment difference as a 1 degree of freedom contrast involving the differences between the treatment group means at each time point. For example, suppose we are looking at weeks 0, 1, 2, and 3, with corresponding mean \( d_0, d_1, d_2, \) and \( d_3 \). A test for a linearly increasing trend in the differences would be given by testing the hypothesis that the sum \(-3d_0 - d_1 + d_2 + 3d_3 = 0\). (Technically, the coefficients for the mean differences are chosen to give an orthogonal polynomial contrast for testing a linear trend among 4 equally spaced visits.) A test for difference in average change from baseline would be given by testing the hypothesis that the sum \( d_0 - d_1/3 - d_2/3 - d_3/3 = 0\). Note that the same set of mean difference estimates at each visit, \( d_0, d_1, d_2, \) and \( d_3 \), are used in the power calculations for the 2 ways of defining a treatment effect (difference in slope vs difference in average change from baseline).

To apply these methods to placebo-controlled trials of antipsychotics, required information for items (1) to (3) was obtained by reanalysis of data on BPRS total scores from the registry trials. We performed these sample size calculations for a range of potential study durations from 1 week to the full duration of each study. For purposes of sample size estimation, we considered the 2 alternative contrasts discussed above as candidates for defining a primary end point for testing treatment differences over time: (1) difference in linear trends in the means (SLOPE contrast) (equivalent to a difference in slopes or linear rate of change in the means) or (2) mean difference from baseline across all follow-up visits (MEAN CHANGE contrast). If differences between placebo and active treatment increase gradually over 6–8 weeks, a linear trend contrast could provide a good approximation to this pattern and have high power; if differences between groups are manifest in a relatively short period and then show only limited further increases, as suggested by the Agid meta-analysis,\(^{22}\) the average difference over time could be a more appropriate summary of the between-group differences in the means than a straight-line trend contrast. All power calculations were performed for 2-sided tests at alpha = 0.05.

Estimates of the means by treatment group at each visit, along with the variances and correlation structure of the data from each trial, were obtained from SAS® PROC MIXED. The fixed effects were fitted with the model: \( Y = \text{time} + \text{treatment} + \text{treatment} \times \text{time} \), with time treated as a categorical variable, and the variance-covariance matrix was estimated without restrictive assumptions about pattern of between-visit correlations (unstructured model). To calculate sample size requirements for trials shorter than the original length of a study, we dropped the estimated means for later visits, along with corresponding elements of the estimated variance-covariance matrix, when applying the Hedeker et al\(^{33}\) sample size equations. The observed dropout patterns by treatment group in each trial were used in sample size estimation.

**Results**

**Dropout Patterns and Mean Differences**

**Risperidone Trials** Initially, 88 patients were assigned to placebo, 87 to haloperidol, 87 to risperidone 2 mg, 86 to risperidone 6 mg, 87 to risperidone 10 mg, and 88 to risperidone 16 mg. By 8 weeks of follow-up, BPRS total scores were available in 49% of placebo patients, 61% of haloperidol patients, 63% of risperidone 2 mg, 68% of risperidone 6 mg, 67% of risperidone 10 mg, and 74% of risperidone 16 mg (figure 1). Some patients...
were not assessed for symptoms at week 6 but were at week 8. Mean change from baseline in BPRS total scores by week and treatment are shown in table 1 and figure 3. By week 4, the placebo vs active treatment differences were 57%–76% of their maximum values over the 8-week trial. The active treatments were at very high percentages of their maximum change from baseline over 8 weeks (haloperidol: 94%, risperidone 2 mg: 100%, risperidone 6 mg: 91%; risperidone 10 mg: 81%; risperidone 16 mg: 100%). The placebo group had its largest reduction compared with baseline in week 2, after which it began to rise, with a near-zero change from baseline in week 4 and a 3-point increase from baseline at week 8.

Olanzapine Trial. Initially, 68 patients were assigned to placebo, 65 to olanzapine 5 mg, 64 to olanzapine 10 mg, 60 to olanzapine 15 mg, and 69 to haloperidol. By 6 weeks of follow-up, BPRS total scores were available in 35% of patients assigned to placebo, 43% to olanzapine 5 mg, 42% assigned to olanzapine 10 mg, 53% assigned to olanzapine 15 mg, and 44% assigned to haloperidol (figure 2). Mean change from baseline BPRS total score, by week and treatment, are shown in table 2 and figure 4. By week 4, differences between placebo and the active treatments were at 83%–100% of their maximum over the course of the 6 week trial, and the active treatments were at a high percentage of their maximum change from baseline: haloperidol, 94%; olanzapine 5 mg, 93%; olanzapine 10 mg, 71%; and olanzapine 15 mg, 81%. Notably, the placebo group continued to decrease until week 6 and had only achieved 58% of its maximum reduction from baseline at week 4.

Table 1. Combined North American Risperidone Trial: BPRS Total Score Mean Changea From Baseline by Treatment and Week and Time Course of Estimated Treatment Effects

<table>
<thead>
<tr>
<th>Week</th>
<th>Placebo</th>
<th>Haloperidol</th>
<th>Risperidone 2 mg</th>
<th>Risperidone 6 mg</th>
<th>Risperidone 10 mg</th>
<th>Risperidone 16 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>Baseline</td>
<td>54.2</td>
<td>1.2</td>
<td>54.9</td>
<td>1.3</td>
<td>52.1</td>
<td>1.2</td>
</tr>
<tr>
<td>1</td>
<td>-1.5</td>
<td>1.1</td>
<td>-2.9</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>2</td>
<td>-2.0</td>
<td>1.4</td>
<td>-3.6</td>
<td>1.5</td>
<td>-4.3</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>-0.4</td>
<td>1.5</td>
<td>-5.2</td>
<td>1.5</td>
<td>-4.8</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>2.8</td>
<td>1.8</td>
<td>-5.5</td>
<td>1.7</td>
<td>-3.4</td>
<td>1.7</td>
</tr>
<tr>
<td>8</td>
<td>3.1</td>
<td>2.0</td>
<td>-4.8</td>
<td>1.9</td>
<td>-3.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes from baseline</th>
<th>Diff</th>
<th>%max</th>
<th>Diff</th>
<th>%max</th>
<th>Diff</th>
<th>%max</th>
<th>Diff</th>
<th>%max</th>
<th>Diff</th>
<th>%max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>17.7</td>
<td>1.8</td>
<td>28.0</td>
<td>5.8</td>
<td>36.5</td>
<td>1.6</td>
<td>15.6</td>
<td>4.2</td>
<td>33.2</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
<td>20.1</td>
<td>2.3</td>
<td>34.9</td>
<td>8.3</td>
<td>52.9</td>
<td>3.4</td>
<td>32.8</td>
<td>6.4</td>
<td>51.0</td>
</tr>
<tr>
<td>4</td>
<td>4.8</td>
<td>57.9</td>
<td>4.5</td>
<td>68.2</td>
<td>11.4</td>
<td>71.9</td>
<td>5.5</td>
<td>53.1</td>
<td>9.5</td>
<td>76.0</td>
</tr>
<tr>
<td>6</td>
<td>8.3</td>
<td>100.0</td>
<td>6.2</td>
<td>95.1</td>
<td>15.8</td>
<td>100.0</td>
<td>9.3</td>
<td>89.5</td>
<td>12.4</td>
<td>98.6</td>
</tr>
<tr>
<td>8</td>
<td>7.9</td>
<td>96.2</td>
<td>6.5</td>
<td>100.0</td>
<td>14.9</td>
<td>94.3</td>
<td>10.4</td>
<td>100.0</td>
<td>12.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Time Course of Response to Treatment. In the risperidone trials, 50%–71% of patients in the various treatment groups met a ≥20% response criterion for the BPRS total score by week 8 (figure 5). Among responders, the percentage in each treatment group who had met the ≥20% response criterion by week 4 ranged from 77% to 97%. The patterns of time to response by treatment group were generally similar in the olanzapine trial (figure 6), with 66%–93% in each group meeting the ≥20% response criterion by week 6. Among responders, the estimated percentages who met the response criterion by week 4 ranged from 77% to 96%. Thus, for both placebo and active treatments, 66%–97% of the maximum percentage of patients estimated to respond over the course of a 6- to 8-week follow-up will have responded by week 4.
Table 2. North American Olanzapine Trial: BPRS Total Score Mean Change* From Baseline by Treatment and Week and Time Course of Estimated Treatment Effects

<table>
<thead>
<tr>
<th>Week</th>
<th>Placebo</th>
<th>Haloperidol</th>
<th>Olanzapine 5 mg</th>
<th>Olanzapine 10 mg</th>
<th>Olanzapine 15 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>Baseline</td>
<td>58.7</td>
<td>1.4</td>
<td>60.1</td>
<td>1.3</td>
<td>58.9</td>
</tr>
<tr>
<td>1</td>
<td>-3.2</td>
<td>1.6</td>
<td>-11.6</td>
<td>1.6</td>
<td>-5.9</td>
</tr>
<tr>
<td>2</td>
<td>-3.8</td>
<td>1.7</td>
<td>-14.8</td>
<td>1.6</td>
<td>-8.2</td>
</tr>
<tr>
<td>3</td>
<td>-4.7</td>
<td>1.9</td>
<td>-14.8</td>
<td>1.8</td>
<td>-9.1</td>
</tr>
<tr>
<td>4</td>
<td>-5.0</td>
<td>2.0</td>
<td>-15.0</td>
<td>1.8</td>
<td>-10.0</td>
</tr>
<tr>
<td>5</td>
<td>-6.7</td>
<td>2.2</td>
<td>-15.2</td>
<td>2.0</td>
<td>-9.1</td>
</tr>
<tr>
<td>6</td>
<td>-8.6</td>
<td>2.2</td>
<td>-15.9</td>
<td>2.0</td>
<td>-10.7</td>
</tr>
</tbody>
</table>

*Estimated from mixed model for incomplete repeated measures: \( D_{tk} = \text{week}_t + \text{treatment}_k + \text{week}_t \times \text{treatment}_k + \text{error} \). Observed means and SEs are presented for baseline BPRS total score.

Sample Size Requirements by Trial Duration

Sample sizes needed to detect the observed treatment differences by study week are presented in table 3. Sample size requirements are presented for 2 methods for comparing treatment effects (as defined in the methods): a SLOPE and a MEAN CHANGE method. Mixed model analysis of variance comparing BPRS total scores for active treatment vs placebo at 2-sided alpha = 0.05 and power = 0.80 were used in all data sets. Using the mixed model estimates of mean differences and covariances from both data sets, we note the following: (1) In the olanzapine trial, sample size requirements for the SLOPE contrast varied irregularly with the duration of the studies, reflecting the absence of a consistently increasing trend in differences in the means; in particular, there were very large increases in required sample sizes at 6 weeks when the placebo-active treatment differences narrowed sharply. Sample size requirements were generally smaller and more stable for the MEAN CHANGE than the SLOPE contrast. (2) In the risperidone trial, sample size requirements were smaller for the SLOPE than the MEAN CHANGE contrast for the haloperidol, risperidone 2 mg, and risperidone 10 mg groups, which had generally smaller differences from placebo than risperidone 6 and 16 mg but longer time to maximum difference from placebo; for the risperidone 6 and 16 mg groups, there was little difference in the sample size requirements for the 2 methods of analyzing treatment differences. (3) In the olanzapine trial, by week 4, sample size requirements...
were at or near their minimum value for the MEAN CHANGE contrast, with little gain in power from further increases in the length of the trial. In the risperidone trial, the most effective treatments (risperidone 6 and 16 mg) were at or near the minimum sample size requirement by week 4 for both analysis methods. For the remaining active treatments (haloperidol, risperidone 2, and risperidone 10 mg), required sample sizes decreased by 19% to 55% with an increase in trial duration from 4 to 8 weeks.

Similar results were obtained examining power and sample size requirements for the BPRS psychosis score, another common outcome in antipsychotic trials, using SLOPE or MEAN CHANGE contrasts (Supplementary Table S3).

**Discussion**

The results of this study document that symptom reduction from baseline occurs early in the course of treatment with most of the reduction in symptoms antipsychotic response occurring by the 4th week in the study. This is consistent with the results in the meta-analysis of placebo-controlled antipsychotic trials by Agid et al., reviewed above, which found early onset of action within the first 4 weeks (with the largest effect in the first 2 weeks) is a common property of D₂ antagonists with no detectable differences between drugs in timing of onset of symptom reduction. Other recent reports using smaller samples have also suggested that response to antipsychotic treatment are evident in over 80% of subjects by 4 weeks and that nonresponse to antipsychotics can be identified as early as 2 weeks. Our findings add to the mounting evidence that the mechanism of antipsychotic action may be more proximally related to the blockade of dopamine transmission than originally thought. This evidence for earlier antipsychotic response suggests the feasibility of shortening clinical trials intended to establish treatment efficacy. Placebo response is also seen early, but differentiation between drug and placebo is seen in the early phases of these clinical trials.

We also found that a substantial proportion of patients who achieve a clinically significant (>20%) reduction in symptoms over 6–8 weeks of follow-up will have done so by week 4; however, a nontrivial fraction will not achieve such a reduction until later in the course of treatment. Thus, the clinical question of how long an individual patient should be tried on a given agent before she or he is judged a “nonresponder” may not be fully answered in short-duration studies. Longer term follow-up, possibly in designs with active rather than placebo controls, will be required to provide data on this question.
This pilot study found relatively little impact on the sample size required to detect the evidence of efficacy if the duration of placebo-controlled trials of second-generation antipsychotics were substantially shortened, to 3–4 weeks. Exceptions occurred in the combined risperidone trial data set with drug/dose combinations that showed relatively small effects and whose limited effectiveness vs placebo was largest at late stages in the study when potential for bias due to dropout is highest. Further reanalysis of other placebo-controlled studies is desirable to confirm our result that the efficacy of drug/dose combinations with effects comparable to usual clinical doses of risperidone olanzapine can be confirmed with a 4-week study. Factors contributing to this result may include (1) early onset of antipsychotic action renders effects detectable within relatively short periods; (2) high levels of dropout observed from the active treatment and placebo arms make later visits less informative for the statistical analyses considered in this report; and (3) increasingly high between-visit correlations in BPRS scores as these trials progressed (data not shown)—possibly reflecting stabilization of symptoms among patients continuing in the studies—lead to less “new” information being provided by later visits. Aside from power considerations, shortening the duration of antipsychotic trials would have the added statistical benefit of reducing bias due to high rates of dropout, particularly among patients assigned to placebo (>50% by the last visit of the 2 studies considered).

One limitation of our current study is this potential for biases due to informative dropouts. Use of mixed models to estimate treatment differences over time, to study the time course of treatment effects and to provide inputs to estimate sample size requirements, may reduce but not eliminate this potential for bias. Use of the Kaplan-Meier method of estimating distribution of time to response also is potentially biased by the presence of informative dropouts (patients who leave due to exacerbation of symptoms which reduce their probability of responding compared with those who stay in). This bias would be larger at later visits and would have the effect of overestimating the cumulative response rate at those visits. To the extent (eg, in the risperidone trial) that informative dropouts are more common in the placebo group, this bias would result in underestimating the placebo-active treatment difference in response rates at later stages of these trials. Thus, consideration of the likely impact of biases in these results reinforces the conclusion that the majority of evidence for placebo-active treatment differences in response is seen by 4 weeks. We note further that shortening the duration of placebo-controlled trials would reduce the potential adverse impact of informative dropouts on study validity.

The short-duration design for placebo-control antipsychotics trials that we propose may serve a variety of purposes in the drug development process, both in phase 2 studies, which seek to demonstrate proof of efficacy while selecting the most effective dose for further development, and later in phase 3 registration trials, particularly where short-duration placebo-controlled trials might satisfy requirements of some countries for studies conducted within their own borders. Relatively short-duration studies of 4 or even 8 weeks may not be sufficient to fully identify some treatment emergent side effects, such as metabolic abnormalities. Demonstration of long-term safety, tolerability, and effectiveness in phase III studies might require longer term placebo-controlled follow-up, although some of these purposes might be addressed by active control parallel group designs comparing novel compounds to several already licensed antipsychotics selected to have well established and somewhat different adverse effect profiles (eg, olanzapine, risperidone, and/or haloperidol). The recently concluded CATIE trial might serve as one model for such studies. Active control studies of long duration are attractive from an ethical stand point and may be less prone to the scientific problems that come with high dropout rates in placebo-controlled studies.

Another critical question for evaluating the proposed new design for dopamine antagonist antipsychotic trials...
is whether a shorter study duration increases the risk of false-positive trials, in which a compound with early evidence of efficacy vs placebo loses efficacy later in patient follow-up, or of false-negative (failed) trials, in which a compound whose efficacy is otherwise proven fails to show a difference from placebo in one study. Future investigation of failed trials would be helpful to clarify this point. In the studies considered in the current analysis, the answer appears different in the olanzapine vs the risperidone groups, primarily due to differences in the behavior of the placebo groups at the later stages of the trials. In the olanzapine trial, the placebo-active treatment group tends to narrow at later weeks, especially for the haloperidol and olanzapine 5 mg/day groups. This is not due to a loss of antipsychotic action in the olanzapine and haloperidol arms—there is little reduction in the magnitude of the actual change from baseline at later weeks compared with week 4. But the mean symptom score for the placebo group also shows continuing improvement through the final week, at the same time as the number of patients remaining in the placebo group goes from 51% at week 4 to 35% at week 6. This unexpected behavior of the placebo group could be due to selective withdrawal of patients who experienced exacerbation or unsatisfactory clinical response to placebo. In the combined risperidone trials, the magnitude of the symptom change from baseline in the haloperidol, risperidone 2 mg/day and risperidone 10 mg/day likewise shows little change between 4 and 8 weeks. However, the placebo group worsens substantially from week 4 to week 8 (while retention drops from 63% to 49%), so that sample sizes required to detect differences vs placebo for the 3 least efficacious treatment groups in this trial decreased with longer study duration. It is difficult to explain why the observed time trends in symptoms among the placebo patients were so different in the olanzapine and risperidone trials, but these results reinforce caution about the difficulty in inferring treatment effects in long-duration placebo-controlled trials with high dropout rates. Variation in placebo response patterns, and their relationship to differential attrition, need to be investigated across a wider range of studies.

These data and conclusions relate to a sharply defined problem: initial testing of an efficacy hypothesis for a dopamine antagonist antipsychotic drug. Caveats are (a) a short trial duration can only detect acute safety issues; (b) a short trial duration will not establish duration of efficacy or prevention of exacerbation; (c) comparisons of efficacy among compounds based on meta-analyses would need to take account of the varying duration of follow-up and use active vs placebo controls and (d) these data may not support shortening trial duration for compounds with a different mechanism of action than currently marketed antipsychotic drugs or injectable forms of existing drugs with a long half-life that may take longer to first attain therapeutic levels.

The short trial duration has advantages including (a) increasing safety by reducing time of exposure to placebo and to experimental drug; (b) reducing subject attrition; and (c) facilitating recruitment of subjects. If an efficacy hypothesis is supported, further studies with different designs will address duration of action, relapse prevention, and safety. It seems likely that short trial results will predict duration of action and relapse prevention because all dopamine antagonist antipsychotic agents to date share the early effect, prolonged effect, and prevention effect.

Based on the data from this study, further investigation is warranted to establish whether placebo-controlled trials of 3- to 4-week duration may provide the optimal balance between establishing efficacy of new antipsychotic agents and reducing exposure to an ineffective treatment regimen (placebo) or unexpected adverse effects of the experimental drug. The results of this pilot study suggest that shortening randomized placebo-controlled clinical trials for antipsychotic drug efficacy testing may be feasible, responds to ethical concerns about extended use of placebo controls, and may improve scientific validity.

Supplementary Material
Supplementary materials are available online at http://schizophreniabulletin.oxfordjournals.org.

Funding
Advanced Center for Intervention and Services Research (P50 MH40279) from National Institute of Mental Health.

Acknowledgments
We are grateful to Eli Lilly and Co and Janssen Pharmaceuticals for providing the data from, respectively, the North American double-blind olanzapine trial (Beasley et al, 1996) and the combined North American risperidone trials (Marder et al 1994, Chouinard et al 1993), which were used in these analyses. These analyses would not have been feasible without the dedicated efforts and staff of the North American double-blind olanzapine trial and combined North American double-blind risperidone trials. Preliminary results were presented at the American College of Neuropsychopharmacology, December 2006.

References


