

What Is Causing the Reduced Drug-Placebo Difference in Recent Schizophrenia Clinical Trials and What Can be Done About It?

Aaron S. Kemp^{1,2}, Nina R. Schooler³, Amir H. Kalali^{4,5},
Larry Alphas⁶, Ravi Anand⁷, George Awad⁸,
Michael Davidson⁹, Sanjay Dubé¹⁰, Larry Ereshefsky¹¹,
Georges Gharabawi¹², Andrew C. Leon¹³, Jean-
Pierre Lepine¹⁴, Steven G. Potkin², and An Vermeulen¹⁵

²Psychiatry and Human Behavior, University of California, Irvine, CA; ³Psychiatry, Downstate Medical Center, State University of New York, Brooklyn, NY; ⁴Quintiles Inc, San Diego, CA; ⁵Psychiatry, University of California, San Diego, CA; ⁶Ortho-McNeil Janssen Scientific Affairs, Titusville, NJ; ⁷Anand Pharma Consulting, Oberwil, Switzerland; ⁸Clarke Institute of Psychiatry, University of Toronto and Mental Health Program, Humber River Regional Hospital, Toronto, Canada; ⁹Psychiatry, Chaim Sheba Medical Center, Tel Aviv, Israel; ¹⁰Lilly Development Laboratories, Eli Lilly and Company, Indianapolis, IN; ¹¹California Clinical Trials, Los Angeles, CA; ¹²Hoffman-La Roche, Medical Affairs, Nutley, NJ; ¹³Psychiatry and Public Health, Weill Medical College of Cornell University, New York, NY; ¹⁴Service de Psychiatrie, Hôpital Lariboisière Fernand Widal, Paris, France; ¹⁵Johnson and Johnson Pharmaceutical Research and Development, Beerse, Belgium

On September 18, 2007, a collaborative session between the International Society for CNS Clinical Trials and Methodology and the International Society for CNS Drug Development was held in Brussels, Belgium. Both groups, with membership from industry, academia, and governmental and nongovernmental agencies, have been formed to address scientific, clinical, regulatory, and methodological challenges in the development of central nervous system therapeutic agents. The focus of this joint session was the apparent diminution of drug-placebo differences in recent multicenter trials of antipsychotic medications for schizophrenia. To characterize the nature of the problem, some presenters reported data from several recent trials that indicated higher rates of placebo response and lower rates of drug response (even to previously established, comparator drugs), when compared with earlier trials. As a means to identify the possible causes of the problem, discussions covered a range of methodological factors such as participant characteristics, trial designs, site characteristics, clinical setting (inpatient vs outpatient), in-

clusion/exclusion criteria, and diagnostic specificity. Finally, possible solutions were discussed, such as improving precision of participant selection criteria, improving assessment instruments and/or assessment methodology to increase reliability of outcome measures, innovative methods to encourage greater subject adherence and investigator involvement, improved rater training and accountability metrics at clinical sites to increase quality assurance, and advanced methods of pharmacokinetic/pharmacodynamic modeling to optimize dosing prior to initiating large phase 3 trials. The session closed with a roundtable discussion and recommendations for data sharing to further explore potential causes and viable solutions to be applied in future trials.

Key words: schizophrenia/clinical trials/placebo response/signal detection/meeting summary/ISCTM/ISCDD

Introduction

Following the “Decade of the Brain,” the climate of scientific understanding regarding the pathophysiological mechanisms underlying diseases and dysfunctions of the central nervous system (CNS) had never been better for the development of therapeutic breakthroughs. Nevertheless, presently approved treatments are, for the most part, uniformly unable to address the underlying etiology of the diseases they are aimed to treat, providing, at best, palliative symptomatic relief.¹ Moreover, despite increased scientific understanding and tremendous industrial investment in the development of CNS therapeutics, only 9% of the compounds that have entered phase 1 trials in recent years have survived to launch.² Over 50% of these failures, according to Hurko and Ryan,² were directly attributable to an inability to demonstrate efficacy during phase 2 trials. This is an increase of 15% in phase 2 failures over the previous decade. Clearly, there is a need for addressing the methodological obstacles that are impeding progress in this critical sector of therapeutic development.

The development of pharmacologic interventions to treat CNS disorders is fraught with unique challenges that often require highly specialized methodologies and trial designs. In particular, existing forums have not

¹To whom correspondence should be addressed; tel: 714-456-8690; fax: 714-456-5967, UCI Neuropsychiatric Center, 101 The City Drive South, Orange, CA 92868; e-mail: akemp@uci.edu.

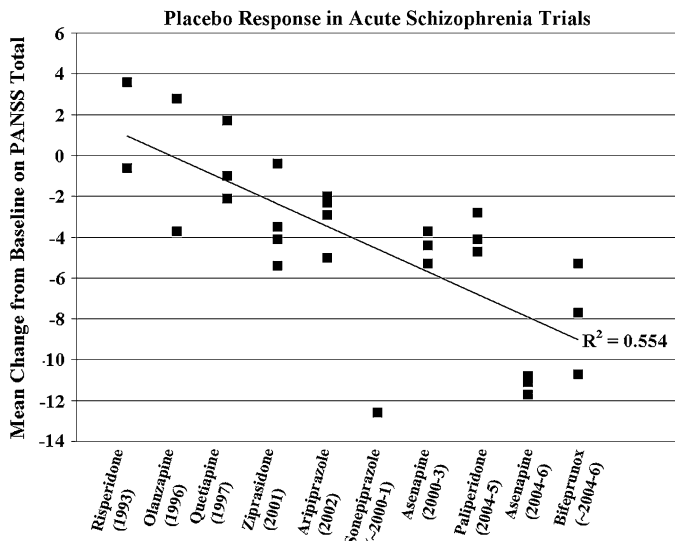


Fig. 1. Mean Change From Baseline in Total Positive and Negative Syndrome Scale (PANSS) Scores for Subjects Receiving Placebo Across Randomized, Double-Blind, Placebo-Controlled, Clinical Trials Has Increased in the Direction of Greater Improvements, Which Is Correlated With the Year That the Studies Were Conducted.

adequately addressed the gaps or disconnects between regulatory demands and requirements and clinical feasibility in the CNS area. In contrast, the fields of oncology, immunology, and cardiology have developed extensive interactions among academics, regulators, and industry scientists, and this process has significantly improved clinical research in these disciplines. Therefore, in order to address the obstacles confronting researchers pursuing experimental CNS treatment modalities, 2 complimentary organizations (the International Society for CNS Clinical Trials and Methodology [ISCTM] and the International Society for CNS Drug Development [ISCDD]) have been formed in recent years.

The focus of this joint session was on examining the apparent diminution of drug-placebo differences (even with previously established comparator drugs) that has been observed in recent clinical trials of antipsychotic

medications for schizophrenia. Because the development of antipsychotic medications is now entering a “third generation,” this issue must be exhaustively reviewed prior to launching additional multimillion dollar trials, or new chemical entities may suffer from the same methodological shortcomings that have apparently limited the potential for accurate determinations of efficacy over placebo. The session was organized to present data regarding drug-placebo differences over time, review probable causes, and consider potential solutions that could be implemented in future trials. The following is a summary of the presentations and discussions from the session.

Nature of the Problem

The session chairs, A.H.K. and N.R.S., provided an overview of the issue and introduced its importance. To place the issue within historical context, L.A. provided a review of how recent randomized controlled trials (RCTs) stand in contrast with earlier studies particularly regarding placebo response levels. Drawing on data from trials conducted between 1991 and 2006 (figure 1), the mean reduction from baseline on the total score of the Positive and Negative Syndrome Scale (PANSS³) for participants receiving placebo has increased across RCTs. There is a significant correlation ($R^2 = 0.55$) between the amount of reduction and the year the study was conducted.

L.A. also compared a phase 3 development program conducted during the early to mid-1990s with a more recent phase 3 program. This comparison showed that, by and large, participants in the later RCTs had a significantly larger response to placebo on both PANSS total scores and Clinical Global Impression-Severity ratings at 6 weeks, relative to baseline. As shown in figure 2, the percentage of placebo-treated patients falling into respective categories of change from baseline on total PANSS scores shows a significant overall shift toward greater improvement/less worsening at 6 weeks in later trials compared with earlier trials. Paralleling this increase in placebo response, L.A. also reported a reduction in active drug response, further contributing to an overall

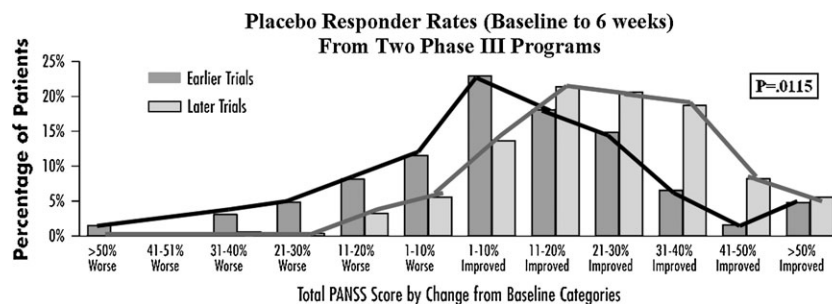


Fig. 2. The Percentage of Placebo-Treated Patients Falling into Respective Categories of Change From Baseline on Total Positive and Negative Syndrome Scale (PANSS) Scores Shows Fairly Large Differences at Many Specific Levels and a Significant Overall Shift Toward Greater Placebo Improvement and Less Placebo Worsening at 6 wk in Later Trials Compared With Earlier Trials.

Table 1. Results From a Phase 3, Placebo-Controlled, Antipsychotic Trial

	Placebo	Compound A	Compound B
Number of patients	89	85	85
Mean PANSS total score at baseline	85.9	87	87
Mean change from baseline in PANSS	-11.7	-10.7	-12.2
<i>P</i> value for change (vs placebo)		NS	NS

Note: PANSS, Positive and Negative Syndrome Scale; NS, not significant.

diminishing of drug-placebo differences. These observations raise many questions: Have the populations included in the trials changed? Are recent studies poorly designed or executed? Has the disease itself or its diagnosis changed? Are the newer drugs simply less effective than older compounds studied?

Further evidence for a diminishing of drug-placebo differences in recent schizophrenia trials was also presented by S.G.P. The results of a recent phase 3, randomized, double-blind, placebo- and positive-controlled, multicenter, efficacy, and safety trial provided evidence that the reduction in drug-placebo difference is not only due to increases in placebo responses but that diminished responses to active compounds has also contributed to the effect. As shown in table 1, the overall mean change from baseline in total PANSS scores for subjects receiving both active compounds was not distinguishable from that of subjects receiving placebo. S.G.P. pointed out that both compounds had shown an efficacy signal in earlier acute exacerbation trials. One, the active comparator, was a widely used antipsychotic that is well accepted as safe and efficacious.

There were no obvious demographic differences among the randomized groups. However, one potentially important distinction within the groups may be relevant as a potential source of the problem. Table 2 shows that the subjects at sites outside of the United States had

slightly lower placebo response coupled with a notably larger drug response to both active compounds. Subjects from these sites (5 in Ukraine and 2 in Russia) also displayed higher (more severely ill) PANSS scores at baseline. However, baseline severity does not account for the difference in drug-placebo difference between US and ex-US sites. Duration of illness, on the other hand, did appear to have an impact; subjects whose current episode of exacerbation exceeded 4 weeks at baseline displayed consistently lower response to active drug, though placebo responses were mixed. Additional data from 3 RCTs were also presented to assess the impact of whether the studies were conducted in the United States or other locations. These data did not illustrate any consistent differences in placebo responses.

In discussing the implications of these findings, meeting attendees suggested additional participant characteristics that should be examined. Prior exposure to newer atypical antipsychotics may be relevant because the prescription prevalence of these may differ within and outside of the United States. An additional, related possibility raised was whether having more patients who are pleased with their current medications might alter the potential subject pool because fewer participants may seek research opportunities in hopes of finding better treatment. This also raised the question that “professional” research participants, increasingly common in the United States and whose motivations may not stem solely from a desire to find a more efficacious medication, may represent a potential source of higher placebo response due to an apparent eagerness to please the investigators, in hopes of returning for additional trials.

Findings of diminished drug-placebo differences are not unique to schizophrenia RCTs, as was illustrated by S.D., who presented an overview of efforts to resolve similar challenges in the area of antidepressant drug development. Many of the previous findings that have served to draw attention to this question in antidepressant RCTs and proposals for potential design or analytic solutions were credited to the pioneering work spearheaded by William Potter.⁴⁻⁶

Table 2. Results From a Phase 3 Antipsychotic Trial by Sites Within and Outside of the United States (US)

	Sites outside of the US			Sites within the US		
	Placebo	Compound A	Compound B	Placebo	Compound A	Compound B
Number of patients	11	8	9	78	77	76
Mean PANSS total score at baseline	99	96	98	84	86	86
Mean change from baseline in PANSS	-9.2	-22.8	-27.7	-12.0	-9.5	-10.3
<i>P</i> value for change (vs placebo)		.040	.093		NS	NS

Note: PANSS, Positive and Negative Syndrome Scale; NS, not significant.

Overall, depression is an illness that is inherently “placebo responsive” and often characterized by fluctuating severity and spontaneous remission. As such, antidepressant RCTs have always been particularly bedeviled by poor drug-placebo differentiation. Similar to the findings presented for schizophrenia trials, this phenomenon has become increasingly more problematic in recent years. Among the underlying issues that have been considered in antidepressant trials are the increasing expectations for greater efficacy on the part of study participants, statistical “regression to the mean” for participants who enter a study when fluctuating symptom severity is at its highest, findings that higher dosing frequency results in greater placebo response, and the fact that participants who fail to improve are more likely to drop out of a study early. An additional issue is the question of whether newer compounds, with fewer side effects, may be less “detectable” to participants, thus decreasing the inherent positive bias toward the active compound and overall drug-placebo differences in response. In conclusion, S.D. suggested that the placebo response must be considered in the broader context of “signal detection” and that the focus must remain on methods to improve the determination of efficacy (the “signal”) independent of the “noise” introduced by placebo response.

Search for the Causes

In the interest of identifying contributing or “causal” factors, session chairs A.H.K. and N.R.S. presented an overview of the relevant issues raised in the earlier discussions and introduced additional factors for consideration. Among the factors discussed were specific participant characteristics (eg, gender, treatment resistance and previous exposure, medication adherence, diagnostic accuracy, illness severity), site characteristics (eg, academic or commercial, experience and training of personnel, recruitment pressures and procedures, “recycling” of subjects), trial design issues (eg, entry requirements, timing of assessments, double-blind placebo lead-in, allowable concomitant medications), clinical treatment setting (inpatient or outpatient), and the reliability/generalizability of existing outcome measures and assessment instruments. Discussions among the session participants about these issues were primarily directed at the imperative to overcome methodological obstacles but were also well balanced with ethical concerns for study participants, minimization of risks, and optimization of the “real-world” clinical applicability/generalizability of study outcomes (for additional in-depth discussions of some of these issues, see Leucht et al,⁷ Reidel et al,⁸ or Hoffer et al⁹)

R.A. highlighted the importance of study design factors. Specifically, the issue of allowable concomitant treatments and “rescue” medications was reviewed within the context of a recent RCT. The use of benzodiazepines, in particular, warrants careful consideration be-

cause their use may introduce confounds in interpreting efficacy. The exclusion of benzodiazepines appears to lead to higher dropout rates due to lack of efficacy in the placebo group but simplifies the interpretation of study outcome by eliminating the potential for synergistic (or unpredictable) interactions with the experimental compound. R.A. also commented that the equivalence of trials conducted within and outside of the United States must be evaluated more thoroughly before reaching conclusions about whether the differences that have been previously discussed will persist. For example, if differences are a function of greater use of atypical antipsychotics, these differences may diminish in time because clinical practice trends currently seen in the United States may rapidly spread to other regions throughout the world.

These discussions did not lead to a general consensus among participants regarding a single or prominent cause. However, there was a consensus that additional progress could be made if data from more trials could be evaluated. The presentations on which the evidence of diminished differences rested were based on anecdotal evidence or data from 1 or 2 trials. The general recommendation from session participants was that the field should collaborate to share data from multiple failed and successful trials (appropriately deidentified to protect commercial interests) so that questions such as these can be answered using robust data from different programs. By definition, some trials may fail due to statistical chance alone (1 in 20 given a .05 α), but the majority of failed trials are probably attributable to a combination of methodological factors. Thus, the focus of the final portion of the joint session was on presenting statistical and methodological perspectives that could be meaningfully applied in future RCTs.

Potential Solutions

A.V. provided a detailed presentation on the use of advanced pharmacokinetic (PK) and pharmacodynamic (PD) modeling and simulation as a strategy to determine the dose of antipsychotic medication that should maximize separation between side effects (eg, extrapyramidal symptoms, prolactin increases, and other adverse events) and efficacy (eg, PANSS improvement or dopamine receptor occupancy level as an intermediate indicator) while still differentiating from placebo responses. Specifically, population-based PK/PD modeling and simulation can provide a priori projections to guide the design of RCTs (ie, dose selection, timing of efficacy measures, etc) in a manner that should optimize the probability that a given study will detect the “signal” of the compound, despite the “noise” of placebo response. Furthermore, A.V. also reviewed how advanced mixed-effects models that take many relevant aspects into account (eg, participant characteristics, disease progression, treatment effects, placebo effects, differential

dropout rates) can also be applied post hoc to provide invaluable insight into the parameters that contribute most directly to response rate differences among participants. Collectively, advanced methods of modeling and simulation are relatively underutilized analytic tools that could contribute directly to overcoming existing obstacles in detecting drug-placebo differences in antipsychotic clinical trials.

The refinement of assessment instruments may also provide a way to improve detection of differences. This has become a familiar topic in forums addressing clinical trial methodology. In a review of the importance of such issues, A.C.L. presented a statistical perspective on how improved assessment procedures directly translates into increased power to detect changes across time and between treatment arms. According to the *Guidelines for Statistical Practice from the Committee on Professional Ethics* (American Statistical Association, 1999), researchers should “avoid the use of excessive or inadequate numbers of research subjects by making informed recommendations for study size.” Such informed recommendations stem from statistical power analyses, which for most clinical trial designers means increasing sample sizes until the power is sufficient to detect statistically significant change.

Alternatively, substantial improvements in the reliability of assessment procedures can result in decreased within-group variability, increased between-group effect sizes, and consequently smaller sample size requirements to achieve acceptable statistical power.¹⁰⁻¹² This precept was illustrated in a poster, presented at the session by A.S.K., which suggested that the improved reliability afforded by computerized administration of neurocognitive assessments could result in a 28% reduction in the sample size required to detect a 10% improvement on these measures. This estimate was derived from the respective means and SDs obtained in a direct comparison of test-retest reliability and concurrent validity between standard and computerized administration of a representative battery of neurocognitive tests, including

those selected by the Clinical Antipsychotic Treatment Intervention Effectiveness (CATIE) and Measurement And Treatment Research to Improve Cognition in Schizophrenia (MATRICS) consortia.¹³ A.C.L. also provided evidence that within-group variance could be substantially reduced by enhancing interrater reliability and ratings validity using a limited cadre of highly trained raters who were blinded not only to treatment but also to time point in the study. He described a method for such assessments that uses raters at a central site who are connected to study participants via a secured video internet connection and showed data indicating improved ability to detect drug-placebo differences.

The importance of site characteristics and potential solutions for improving site performance was presented by L.E. Among the many issues reviewed, the concern that “professional” participants or rater inflation are a unique problem for US-based sites was depicted as somewhat premature, particularly because entrepreneurs throughout the world will inevitably follow capital investment in this market. Therefore, the critical determining factors influencing site selection should be based on individual site and investigator characteristics that indicate the investigators’ commitment to ensure ethical, nonbiased execution of study protocols. Included in the attributes that L.E. suggested a “quality” site must possess were staff with considerable clinical experience working with the patient population and assessment instruments employed, demonstrated “in-house” training procedures and quality assurance metrics, ongoing programs to prevent rater drift and insure consistency as staff turnover, a reliable source of participant recruitment across a variety of settings, and facilities that are appropriate to fully service the clinical needs of the participants and requirements of the study. An additional issue raised in discussion following this presentation was the need for enhancing dialogue between the sponsor and the participating sites to induce greater involvement by the investigators in the planning and design aspects of the study. In closing, L.E. stated that the “culture” of

Table 3. Summary of Possible Contributory Problems and Potential Solutions Proposed at the Meeting

Contributory Problems	Potential Solutions
Subjects with longer duration of current acute episode show poorer response to treatments.	Exclude subjects whose current acute episode exceeds 1 mo prior to enrollment in study.
Use of “rescue medicines” such as benzodiazepines can confound drug-placebo differences.	Exclude subjects who are concurrently “stabilized” on benzodiazepines and prohibit concomitant use.
Variability in outcome assessments decreases statistical power to detect drug-placebo differences across the study.	Increase reliability through improvements in assessment administration/instrument design and rater training/qualification procedures.
Insufficient bases for selecting doses to be tested in phase 3 studies may limit potential findings.	Use of pharmacokinetic/pharmacodynamic modeling to optimize dosing regimes in trials.
Subjective reports of medication compliance are notoriously unreliable in this population.	Explore new methods to measure medication compliance and increase protocol adherence.

a site is best judged by the involvement of the principal investigator, which, in turn, is a critical determining factor in the quality of the data that will result from the study as a whole.

Following formal presentations, roundtable discussions were conducted among the session participants and a panel composed of L.E., R.A., M.D., J.-P.L., and A.C.L. These discussions were moderated by session chairs, A.H.K. and N.R.S., and served as a platform for debating the overall implications of the issues raised throughout the session among a range of expert attendees from industry, academia, clinical sites, and governmental agencies. A poster session also served as an additional outlet for the sharing of findings with direct relevance to the issues discussed.

Conclusions

The first joint session between the ISCTM and the ISCDD was brought to a conclusion by N.R.S. who reviewed the relevant issues concerning the problem of diminishing drug-placebo differences in acute schizophrenia RCTs and potential solutions discussed (summarized in table 3). Overall, shorter duration of current exacerbation of symptoms (<1 month) seems to be a relevant factor resulting in greater drug responses, as does inpatient settings. Reducing the use of benzodiazepines appears prudent, as do strategies to reduce sample size requirements by improving assessment reliability, particularly because the latter represents an ethically responsible method to reduce placebo exposures to research participants. Further factors that seem reasonable to pursue include improved methods to enhance and/or measure protocol adherence and medication compliance by participants, increased site investigator involvement and commitments to scientific rigor, and the credentialing of raters to ensure interviewing skills sufficient to gauge the full breadth of symptoms that characterize patients with schizophrenia. In closing, N.R.S. appealed to the 2 societies' respective members to develop larger pooled data repositories that could be examined to identify potential sources of placebo responses on failed trials and facilitate more in-depth meta-analyses such as that recently reported by Leucht et al.¹⁴ Because the interest instigated by the industry's needs for improved methodologies may be leveraged into greater scientific understanding with much larger implications, the pursuit of solutions to this problem will serve to benefit both the patients to be treated and society, as well.

Acknowledgments

Proceedings of the First Collaborative Session between the ISCTM and the ISCDD, Brussels, Belgium,

September 18, 2007. Rapporteur: Kemp; Session Chairs: Schooler and Kalali; Speakers: Alphas, Anand, Awad, Davidson, Dubé, Ereshefsky, Gharabawi, Kalali, Leon, Lepine, Potkin, and Vermeulen. Open access charges provided by the International Society of CNS Clinical Trials and Methodology (ISCTM).

References

1. Pangalos MN, Gallen CC. Drug discovery for disorders of the central nervous system. *Neurotherapeutics*. 2005;2:539–540.
2. Hurko O, Ryan JL. Translational research in central nervous system drug discovery. *Neurotherapeutics*. 2005;2:671–682.
3. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale for schizophrenia. *Schizophr Bull*. 1987;13:261–276.
4. Faries DE, Heiligenstein JH, Tollefson GD, Potter WZ. The double blind variable placebo lead-in period: results from two antidepressant clinical trials. *J Clin Psychopharmacol*. 2001;21:561–568.
5. Mallinckrodt CH, Sanger TM, Dube S, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biol Psychiatry*. 2003;53:754–760.
6. Liu KS, Snavely DB, Ball WA, Lines CR, Reines SA, Potter WZ. Is bigger better for depression trials? [published online ahead of print September 07, 2007]. *J Psychiatr Res*. 2008;42(8):622–630.
7. Leucht S, Heres S, Hamann J, Kane JM. Methodological issues in current antipsychotic drug trials. *Schizophr Bull*. 2008;34:275–285.
8. Riedel M, Strassnig M, Muller N, Zwack P, Moller HJ. How representative of everyday clinical populations are schizophrenia patients enrolled in clinical trials? *Eur Arch Psychiatry Clin Neurosci*. 2005;255:143–148.
9. Hofer A, Hummer M, Huber R, Kurz M, Walch T, Fleischhacker WW. Selection bias in clinical trials with antipsychotics. *J Clin Psychopharmacol*. 2000;20:699–702.
10. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry*. 1995;52:867–871.
11. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry*. 2000;47:762–766.
12. Kraemer HC. To increase power in randomized clinical trials without increasing sample size. *Psychopharmacol Bull*. 1991;27:217–224.
13. O'Halloran JP, Kemp AS, Gooch K, et al. Psychometric comparison of computerized and standard administration of the neurocognitive assessment instruments selected by the CATIE and MATRICS consortia among patients with schizophrenia [published online ahead of print December 21, 2007]. *Schizophr Res*. 10.1016/j.schres.2007.11.015.
14. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials [published online ahead of print January 08, 2008]. *Mol Psychiatry*. 10.1038/sj.mp.4002136.