Implementation Considerations for Multisite Clinical Trials with Cognitive Neuroscience Tasks

Richard S. E. Keefe\textsuperscript{1,2} and Philip D. Harvey\textsuperscript{3}

\textsuperscript{2}Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Box 3270, Durham, NC 27710; \textsuperscript{3}Department of Psychiatry, Emory University School of Medicine, 101 Woodruff Circle, Suite 4000, Atlanta, GA 30322, US

Multisite clinical trials aimed at cognitive enhancement across various neuropsychiatric conditions have employed standard neuropsychological tests as outcome measures. While these tests have enjoyed wide clinical use and have proven reliable and predictive of functional disability, a number of implementation challenges have arisen when these tests are used in clinical trials. These issues are likely to be magnified in future studies when cognitive neuroscience (CN) procedures are explored in these trials, because in their current forms CN procedures are less standardized and more difficult to teach and monitor. For multisite trials, we anticipate that the most challenging issues will include assuring tester competence, monitoring tester performance, specific challenges with complex assessment methods, and having resources available for adequate monitoring of data quality. Suggestions for overcoming these implementation challenges are offered.

Key words: schizophrenia/clinical trials/cognition

Previous work on the effects of pharmacologic interventions on cognition has generally involved standard neuropsychological (NP) tests with a long history of use in neuropsychiatric populations including schizophrenia. In fact, a history of previous use in clinical populations is one of the criteria for selection of cognitive outcome measures recommended by several experts in this field.\textsuperscript{1–5} These tests have been chosen due to the ease with which they can be used in multisite clinical trials, and many of these trials have been completed to date. The possibility of similar multisite trials involving cognitive neuroscience (CN) tasks, which may require more attention to methodological detail, more complex instrumentation, and greater expertise with complex assessment methodology, presents a series of new challenges for our field. It is the intention of this article to provide a description of lessons that have been learned from previous implementation of standardized NP tests in multisite clinical trials and to apply these lessons to future studies involving CN tasks. We define CN tasks here as procedures that allow the assessment of very specific cognitive procedures. These tasks are almost always developed in a single laboratory, although different versions may be developed at multiple laboratories simultaneously. Usually, considerably more work has been done with healthy subjects performing these procedures. In almost all cases, CN procedures involve computerized stimulus presentation and subject response data collected with electronic data capture. However, many computerized presentations of standard tasks do not qualify as CN tasks. Further, while many CN tasks have been used in the context of psychophysiological or neuroimaging methods such as electroencephalography or functional magnetic resonance imaging, we will limit our comments here to those tests that focus primarily on the collection of performance data. Naturally, some of the arguments presented here will apply to imaging and psychophysiological tests as well.

An example of a relatively simple CN task is the AX Continuous Performance Test (AX CPT) procedure. This task requires several practice trials so that patients can demonstrate their ability to distinguish between left mouse-button responses to pairs of stimuli that follow the sequence “A” then “X” and right mouse-button responses to all other sequences, such as “A” then “not X” and “not A” then “X.” The pattern of response errors across trial types indicates disruption of specific cognitive processes but is only meaningful if the patient demonstrates understanding of the task rules. Thus, it is essential that patients learn the task correctly during an extensive training period.

Most previous studies testing schizophrenia patients with CN tasks have involved single sites. This methodology allows great attention to detail and essentially no need for clear communication of assessment procedures to other investigators who may have reduced levels of assessment (or even research) experience. Multisite clinical trials differ from these single-site investigations in a variety of ways. First, the collection of data from different
testers, sites, and patient populations and the diffusion of responsibility associated with multisite clinical trials can reduce data quality and produce heterogeneity of results, including increased variance in scores across sites. Data quality from sites with little experience administering cognitive tests may be particularly jeopardized. Frequently, schizophrenia clinical trial sites that have a history of completing symptom rating scales are included in multisite cognitive enhancement trials. Raters from these sites may have no previous training or experience in the performance of NP assessments. While some of these site effects (eg, variation in characteristics of patient samples) may be controlled statistically, the systematic error that is due to administration errors across different sites cannot be controlled statistically and needs to be minimized. Control over the systematic error arising from inaccurate administration of NP tests needs to be addressed early on in the research design process.

One of the ways in which data quality may vary is the ability of sites and testers to complete the selected tasks in a manner consistent with the standardized instructions for NP test administration. While many multisite studies with standard NP tests have used testers with minimal previous testing experience, it may be essential for all sites in multisite clinical trials with a CN outcome to have experience with the administration of similar tasks. At the outset, this may be a challenge because many of the clinical trial sites that have added NP assessments to their clinical trials tool kit may have almost no experience with more sophisticated tasks. If this is the case, greater education, training, and certification procedures beyond what is normally employed for a clinical trial will be required. Attention will need to be paid to the ability of a site facility to provide an assessment space that meets the requirements (eg, ambient sound and light levels) for administration of the CN procedure and for the processing and storage of accurate data. In addition to an evaluation of each site, each tester must also be carefully selected and trained. It will be important for any study to identify the minimal educational and experience requirements for any task used in a clinical trial. Tester screening and evaluation should occur prior to any central investigators’ meeting. While these procedures have been employed previously in NP assessment studies, the concerns described below suggest that there has been less than optimal success even when testing is routinized and standardized. These concerns, summarized in table 1, have the potential to be magnified in trials with complex administration procedures, as may occur with the use of CN tasks.

Rater Training and Certification

One of the crucial foundations to designing a study that is sensitive to a real pharmacologic effect is the establishment of test-retest and inter-rater reliability of assessment procedures, ranked by the Measurement and Treatment

<table>
<thead>
<tr>
<th>Table 1. Concerns from Schizophrenia Cognitive Enhancement Clinical Trials Using Standard Neuropsychological Tests That May be Compounded in Multisite Trials Using Cognitive Neuroscience Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater training and certification are essential</td>
</tr>
<tr>
<td>Are testers qualified?</td>
</tr>
<tr>
<td>Excluding unqualified testers</td>
</tr>
<tr>
<td>Educating testers prior to certification</td>
</tr>
<tr>
<td>All testers will require certification</td>
</tr>
<tr>
<td>Are the necessary procedures supported by sponsors?</td>
</tr>
<tr>
<td>Is the importance of these procedures acknowledged by site investigators?</td>
</tr>
<tr>
<td>Data review processes</td>
</tr>
<tr>
<td>When cognition is the primary outcome measure, all data reviewed centrally</td>
</tr>
<tr>
<td>Less intensive data review is risky and must include random checks throughout trial</td>
</tr>
<tr>
<td>Intervention during a trial</td>
</tr>
<tr>
<td>Prior to study initiation, procedures must be in place for adding testers and sites</td>
</tr>
<tr>
<td>Task considerations for clinical trials</td>
</tr>
<tr>
<td>Increased task complexity can increase missing data rate</td>
</tr>
<tr>
<td>Simplify instructions for testers and patients</td>
</tr>
<tr>
<td>Additional concerns with computerized tests</td>
</tr>
<tr>
<td>Automatized procedures can hide problems indigenous to</td>
</tr>
<tr>
<td>schizophrenia clinical trials</td>
</tr>
</tbody>
</table>

Research to Improve Cognition in Schizophrenia (MATRICS) experts as the most important feature of a cognitive battery for clinical trials. Maintenance of high reliability of assessment during a clinical trial will greatly enhance the sensitivity of a task to treatment effects. There are a variety of sources of reduced reliability.

Sponsors

In the past, the research designs and rater training procedures for cognitive enhancement clinical trials in schizophrenia have been limited by the absence of a significant shift in attention toward cognitive outcomes in the training process. Most schizophrenia trialists have come from a tradition in which symptom rating scales such as the positive and negative syndrome scale are the primary outcomes. Thus, even investigators’ meetings that are directed toward cognitive outcomes often have more time allotted for symptom rating training than for training in NP assessment. In many ways, researchers and sponsors alike have not gone past the old strategies for addressing the new challenges associated with cognitive enhancement trials.

Sponsors will need to have ample warning for the training and certification processes that will be required for establishing test reliability. CN test designers must be prepared to teach sponsors the importance of reliability and why low reliability can be produced by poor standardization and will result in a design that has reduced sensitivity to treatment effects that will cost the sponsor valuable financial resources. Test designers should be prepared to cancel their involvement in a trial if procedures
for adequate standardization are not established. Several previously completed studies failed due to inadequate resources directed toward the establishment of good research practice; compromised standardization procedures result in useless data because a missing or invalid baseline score precludes collection of any change score data. The reason that some trials have failed is because sponsors have sometimes chosen not to monitor the quality of data as they were collected, have allowed secondary training of new raters at sites without recertification by a central source, or have not controlled who actually administered the assessments through a certification program.

**Principal Investigators**

Physicians and other professionals associated with the research activities of a site within a multisite trial are often busy individuals who allot precious little time to the identification of individuals who can complete cognitive testing. While this is certainly a nuisance for cognitive enhancement trials that involve standard NP tests, it could be catastrophic for studies using CN tasks. Some principal investigators may delay the identification of testers indefinitely if not forced to choose ahead of time, perhaps because they want to avoid the necessity of paying testers’ salaries prior to the collection of fees associated with assessments of patients. Investigators may also be indecisive about identifying specific study personnel if they are not aware of the level of detail in the training process involved. Because more extensive training and certification processes will be required for clinical trials involving CN tasks, it will be essential that principal investigators are required to identify testers early in the process of site selection and setup. Once testers have been chosen, they should only be changed in extreme circumstances and the new testers should also be certified in the same manner as the previous ones. It will be essential that the sponsor and the research design team keep the principal investigator aware of the training process so that the principal investigators do not inadvertently interfere with it.

**Testers**

The individuals who are chosen to complete the CN tasks at their site will probably need to have hands-on experience with CN tests and will benefit greatly from previous clinical experience with patients with psychotic disorders. Even the experience associated with performing CN tests with healthy undergraduates is a far cry from engaging a psychotic individual into a difficult cognitive paradigm. Screening, education, and certification are essential elements for allowing cognitive testers to collect data for a clinical trial in schizophrenia.

**Screening.** It will be important to gather evidence of previous experience with similar tasks. A process needs to be established that will identify testers who exaggerate their previous testing experience. It is our experience that testers who claim to have administered all tests on a list, including tests that do not exist, will need additional scrutiny.

**Education.** It is essential to be sure that testers demonstrate “evidence” of having completed, practiced, and understood necessary educational sequences that will prepare them for certification at an investigators meeting.

**Certification.** It is very helpful if testers are sent necessary equipment, software, and instruction and scoring manuals well ahead of an investigators meeting (1 month is ideal) so that they can practice their performance prior to certification. Researchers need to eliminate the possibility of a claim from testers that they were not informed of the necessity of full preparation. It is essential that there is personal contact between the personnel from any central coordinating site and each potential tester. We recommend that individuals from a central coordinating site have a telephone conversation with each potential tester in which they indicate that they will be meeting the tester at the investigators meeting and that they will be going through specific steps in such a way that the tester will need to demonstrate his or her competence in test administration and scoring. The tester should also be instructed that the investigator’s meeting procedures will not be for educational purposes only but rather will have a certification component that can be passed or failed depending on the tester’s level of practice and demonstrated competence.

There are a variety of ways in which this system itself can fail. Busy individuals who have reduced motivation for being a cog in the very large machine of a clinical trial will often not put forth the necessary effort to establish the systematic standardization procedures for these tasks. It is essential that sponsors give central coordinating centers the power to evaluate the performance of the testers and the power to reject testers who will not collect reliable data. Especially, during this time of increasing criticism of pharmaceutical companies, industry physicians and project leaders may place a great emphasis on the positive impression that clinicians and academic researchers have of their company. However, these public relations’ concerns should not be allowed to undermine the decision to prohibit incompetent testers from collecting cognitive data for a trial. Such concerns may result in the collection of unclean data and lead to confusion about the effects of a potentially important drug for schizophrenia cognition.

Various types of sites will be available for cognitive enhancement clinical trials. Previously, professional clinical trial sites have largely been devoted to the collection of symptom data. Recently, however, several of these sites have been able to collect excellent NP data for
neurocognitive trials. It is an open question whether these sites will be able to run experimental paradigms with sophisticated CN tasks. However, despite their limited expertise, many of these investigators are very eager to be involved in these research efforts and often have flexible staff who are easily coached. The potential of their personnel combined with the large samples of patients that are available to some of these sites makes this type of professional clinical trial site a viable option for trials involving CN tasks. On the other end of the spectrum are academic sites that usually have access to fewer patients, have far slower institutional review board processes, and yet have greater expertise. While the greater expertise of academic sites may seem to be an advantage, it may be a mixed blessing. On the one hand, it is helpful for testers to know the importance of attention to detail with regard to CN tasks, and testers at academic sites usually have been steeped in this knowledge. However, testers with a long history of experience with similar tasks may think that the agreed-upon methodology is insufficient. It is crucial that these testers from academic sites are taught the necessity of all testers testing in a systematic manner. Reliability is a function of standardization, and “better” testing performed at one site may actually reduce the aggregate reliability of the overall assessment procedure.

Data Review Processes

In general, the larger the number of sites, the less the testers and principal investigators feel personally responsible for the data that will be analyzed during a clinical trial. Because some clinical trials in schizophrenia with CN tasks may involve a large number of sites, it will be essential to assure that some single entity takes responsibility for data accuracy. These large trials will require ongoing data quality review. In cases where a CN task measure is the primary outcome for a phase III trial, each data point for every patient will need to be reviewed. This review process should be applied to earlier phase trials also because less oversight will endanger data quality. However, if resources for adequate review are not available, at a minimum, the first few assessments should be reviewed in detail by a central source. Random test review following these initial reviews will also help to assure that testers will maintain high-quality standards based upon the expectation that any one of their assessments could be reviewed. Catastrophic continuing inaccuracies can be avoided by reviewing data immediately following test administration.

Another mechanism for maintaining data quality is to establish a network of sites and supervising psychologists to identify and describe problems of test administration or scoring or computer/task malfunction. This network of individuals, which can be maintained through conference calls and occasional in-person meetings, will also serve to establish a community to which testers will feel accountable. These networks also raise the camaraderie of the group and further facilitate the motivation for collecting good quality data. Several successful trials have used multiple supervisory psychologists, which has the benefit as well of making it clear to the testers that high-quality work is not simply an irrational request of a single supervisor.

Smaller studies with select sites of high-quality testers that also have academic reputations to lose may require less review from central coordinating centers. However, each site must have adequate ongoing data quality-checking procedures. Comprehensive or random test review will still be necessary to keep all testers wary of careless mistakes or failures of the computer systems to store data accurately. Clear accountability is an important component of any multisite data collection process.

The importance of data quality issues will increase markedly in studies with complex dependent variables that are often collected from CN tasks. It is less of a challenge to monitor valid ranges of scores on many standard NP tests. For instance, if when administering verbal fluency the tester is instructed to begin timing after subjects make their first response, scores of 0 could not be valid. Similarly, scores of 0 on total learning in multi-trial list-learning tests also mean that the participant did not meaningfully participate. In contrast, a score on a complex hybrid variable such as the ones commonly obtained from CN tests may not be as easily screened for validity by a nonexpert observer. The consistency with which a CN task can produce reliable data will be further challenged by the tendency of these tasks to confine patients’ responses to timed mouse-button presses with little or no opportunities for mid-test feedback. Test administration procedures and patients’ understanding of the test must thus be flawless.

Intervention during the Course of a Trial

During the course of even the most elegantly designed trials, there will be challenges and crises that require response. One of the most frequent of these, of course, is low enrollment of patients. It is a rare cognitive enhancement clinical trial that exceeds the expected pace of enrollment. Most ongoing clinical trials require the addition of sites during the course of the trial. Thus, a plan for mobile evaluation of new sites and testers should be in place prior to the initiation of the trial. Even if new sites are not added during the course of a trial, personnel changes can be expected. Adding new testers to a site with experience can be less demanding than adding testers to a brand new site. While an investigators meeting cannot be scheduled each time an additional tester is added, it is possible to develop processes by which certified testers at individual sites are deputized to train new testers. However, final approval of all testers should be central so that the consistency of test administration
and scoring procedures can be maintained throughout the course of the trial at all sites. Eventually, despite the best intentions, some sites may not be able to participate in this type of trial. It is far better to identify these sites early in the process and to discontinue their involvement before additional resources are wasted.

**Task Considerations**

In addition to the specific implementation challenges of multisite clinical trials with CN tasks described above, the practical constraints of multisite trials will require investigators to make decisions early in the course of study design regarding task duration and complexity. Some of the lessons that have been learned from clinically oriented NP batteries used in multisite cognitive enhancement trials for schizophrenia can be applied to CN paradigms. While the level of sophistication required for CN tasks will certainly be greater, there are some general rules that may apply. First, in multisite trials, if tests measure the same thing with equal reliability, shorter and easier are likely to be better. Cognitive science researchers who complete single-site studies may not be accustomed to sacrificing conceptual focus for brevity and practicality. Unfortunately, multisite clinical trials often require this type of sacrifice. Simplicity and efficiency are key features in multisite assessment. Task designers will need to consider carefully whether a construct can be measured in a simpler, cheaper, and more parsimonious manner. While the notion of a specific cognitive task that measures a cleanly described construct is appealing, task designers should question whether the cognitive construct they have in mind is really as specific as assumed. Any task design features that can be simplified can greatly enhance the consistency with which tasks are administered across sites.

Complex tasks cause confusion not only among testers but possibly among patients, investigators, and the eventual recipients of this clinical information. One of the major problems in translating complex laboratory tests used on healthy populations to the schizophrenia clinic is that the patients being tested may have difficulty learning and remembering what they are supposed to do for successful performance. Assuring that testers understand the instructions for administering the test is a crucial first step; it is obviously unlikely that testers who cannot understand the instructions will be able to teach patients. Multiple practice examples are helpful and shorter blocks of stimuli with repeated repetition of instructions will be very helpful as well. Many patients have reported subjectively that they forgot what they were supposed to be doing on even simple NP tests like animal naming. On an empirical basis, over 25% of the first-episode schizophrenia patients tested by Harvey et al received a score of 0 categories on the 6-category Wisconsin Card Sorting Test (WCST). It is difficult to determine why patients perform this poorly. Either they were unable to match 10 consecutive stimuli on the basis of color, as required in this first category; they forgot the demands of the test, which does not allow repeated instructions; or they were completely unable to adapt to the test’s problem-solving demands. Complex tests like the WCST with limited instructions and no opportunities for mid-test feedback may frequently result in ambiguous outcome measures. Clear instructions, practice trials, and opportunities for feedback are essential elements of CN tasks to be utilized in clinical trials.

Missing data are, of course, greatly deleterious to the estimated statistical power of a clinical trial. Any efforts that can be made to reduce task complexity (for both testers and patients) and thus potentially reduce missing data will enhance the sensitivity of a task to measure drug effects. Finally, designers of a clinical trial with CN tasks need to address the practical question of whether the expected size and/or specificity of a treatment effect is substantial enough to justify the resultant scientific and financial constraints.

Recent data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project support the notion that additional task complexity should be justified by empirical data. While the battery of tests that were completed in the 56-site CATIE project required about 90 min of assessment time and 91% of patients were able to complete enough cognitive data at baseline to provide investigators with a reliable composite score, the more complex tasks had the lowest completion rates. Furthermore, tasks that were easier to administer and score and had the lowest amount of expected testing time contributed the greatest variance to the composite score in the assessment battery. All tests with high rates of completion had written instructions that the testers read to the patients, often directly from the source document forms. Thus, tests that had easily accessible instructions and paper and pencil format were associated with higher degrees of test completion.

Thus, with regard to measuring the generalized deficits of patients with schizophrenia in multisite trials, more complicated is not always better and may sometimes be worse. In fact, any CN task that is included in a multisite clinical trial will need to have a compelling rationale as to either what aspect of cognition is being measured with this task that is not being measured with a generalized cognitive test battery or how this CN task may be doing a better job of measuring that aspect of cognition. The superiority of unifactorial models of cognition in schizophrenia and the importance of general cognitive tasks to measure the deficits in schizophrenia have underscored the explanatory power of brief, simple testing. In fact, the Wechsler Adult Intelligence Scale digit symbol subtest, a brief simple test, was completed by the highest proportion of cases in the CATIE study, took the least time to administer, and accounted for
more variance in the cognitive composite score than any other test.\textsuperscript{7} Recent comparisons between short batteries of tests and batteries of moderate length suggest that the more extensive cognitive assessments do not contribute unique variance to the discrimination between schizophrenia patients and controls.\textsuperscript{10,11} It could be argued that the simpler tasks in these multisite studies explain more variance in overall performance because they are more likely to be administered properly in multisite studies. Given that the CN tasks in multisite trials may be more complex than the most difficult of tasks in these previous trials, such as the Continuous Performance Test—Identical Pairs (CPT IP),\textsuperscript{12} it will be essential that great care is given to the accurate administration and scoring of these tasks during the course of the trial and that missing data points are minimized. Attention to these details as described in this article may help to enable maximal data quality. The assumption inherent in the use of CN tasks in clinical trials is that they will be sensitive to specific deficits or treatment effects beyond these general NP measures. This assumption can only be adequately tested empirically if sufficient attention is paid in these trials to the quality of the data collected.

### The Promise and Challenge of Computerized Testing

While computerization of test procedures, including instructions, stimulus presentation, and data collection, appears to be likely to be associated with increased data quality and higher completion rates of psychological assessments, the data on this issue are contradictory. As noted above, in the CATIE study, the 3 tests with the most missing data were all computerized. In fact, by the third assessment of the patients, 407 of the 1332 (31\%) cases were missing at least one data point on the computerized CPT IP, as compared with 40 out of 1332 cases (3\%) with a missing assessment on the Hopkins Verbal Learning Test. Table 2 lists the various reasons that computerized and noncomputerized tests were missing at the baseline assessment of the CATIE schizophrenia trial and the number of patients for whom data were missing. The most likely reasons for cognitive test data to be missing were the patient's inability to understand the task demands, computer malfunction, and patient's refusal, all of which were far greater in the computerized tests.

Although an attempt was made to recruit “all comers” in the CATIE trial, allowing for inclusion of patients with substance abuse and including sites with minimal previous research experience, the finding of more missing data on computerized assessments is consistent with previous, less inclusive trials. In a study of first-episode patients with schizophrenia randomized to treatment with risperidone or haloperidol, 16\% of the patients were missing their CPT assessment at the baseline and another 8\% were missing their 3-month follow-up. In contrast, the rates of missing data for the multi-trial list-learning test in that study were 4\% at baseline and an additional 1\% at the 3-month follow-up.\textsuperscript{8} Similar rates of attrition were reported in another first-episode study\textsuperscript{13} comprised solely of academic sites, where 12\% of cases were missing the CPT and 4\% were missing their list-learning performance. In a schizophrenia study examining an abbreviated cognitive assessment battery,\textsuperscript{10} there were no missing data for either patients with schizophrenia or healthy controls on a list-learning test, but 27\% of patients with schizophrenia and 8\% of healthy controls subjects had missing data on the CPT, even though this testing was completed at an academic site with well-trained testers and patients with schizophrenia who were clinically stable.

It may be instructive to note that these testing challenges are not limited to patients with schizophrenia. Very similar rates of missing data were found in a study of the treatment of traumatic brain injury with cholinesterase inhibitors.\textsuperscript{14} In that study, every case record form for every patient was monitored for accuracy within 24 h of assessment and the rate of missing data on a paper and pencil verbal learning test was 0\%. However, data were missing from the computerized Cambridge Neuropsychological Test Automated Battery assessment on up to 14\% of cases across the different dependent variables.

These data are not presented to suggest that computerized assessments are unsuitable for use in clinical trials in schizophrenia. In fact, changes in the identical pairs version of the CPT administration system, including more practice trials and a change in the subject response required, are likely associated with the substantial improvement in completion rates from the earlier studies employing the CPT. Rather, we wish to emphasize

<table>
<thead>
<tr>
<th>Reason for missing data</th>
<th>Number of Patients Missing Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient refused test</td>
<td>38</td>
</tr>
<tr>
<td>Patient did not understand instructions</td>
<td>4</td>
</tr>
<tr>
<td>Computer malfunction</td>
<td>0</td>
</tr>
<tr>
<td>Tester error</td>
<td>14</td>
</tr>
<tr>
<td>Unknown</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>Computerized tests</td>
<td>407</td>
</tr>
<tr>
<td>Noncomputerized tests</td>
<td>40</td>
</tr>
</tbody>
</table>

Note: Computerized tests included bundled versions of the 64-card Wisconsin Card Sorting Test (WCST), Visuospatial Working Memory Test, and Identical Pairs Continuous Performance Test as described in Keefe et al.\textsuperscript{7} CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness.

\textsuperscript{a}Only 2 patients did not understand computerized WCST test instructions.

Table 2. Number of Patients (of 1331 total) in the CATIE Schizophrenia Trial for Whom Data Were Missing at Baseline Based upon the Reasons and Whether the Test was Computerized.
that CN test developers will need to ensure that their computerized assessments are “participant” friendly, as well as “tester” friendly in order to obtain suitable completion rates in clinical trials. Although test batteries that are completely computerized may enable patients and testers to become more easily accustomed to this method than test batteries that combine computerized and pen-and-paper tests, it would be an error to assume that completely automated data collection procedures will assure that testing will be completed on all patients.

Summary

Multisite clinical trials of patients with schizophrenia that include CN tasks will contain more complex instrumentation and require significant attention to methodological detail and greater expertise with complex assessment methodology. These issues present a series of new challenges for our field. If accurate data are to be collected, it will be important for study designs to allot time and resources to various issues that are not normally considered in studies involving CN tasks. It will be essential that these issues are addressed proactively prior to the initiation of a trial and continually during a trial by investigators, sponsors, and study personnel alike. These issues include the following:

Task design—tests must have optimal brevity and simplicity while maintaining a sufficient sampling of the constructs of interest;

Documentation—detailed description of task administration and scoring procedures will reduce disconnection between the task design and implementation;

Site selection—facilities, resources, and personnel competencies must allow the collection of complex computerized data;

Testers—individuals who will be collecting CN data must possess relevant educational background, the ability to take responsibility for accurate data collection, experience with the relevant clinical population, and previous experience with CN tasks;

Tester training—testers will need to complete the training and certification procedures necessary to ensure that data are collected accurately and identically across multiple sites;

Data review—ongoing review of data collection accuracy, both centrally and locally, especially at the beginning phases of a trial, will prevent systematic or sporadic errors and missing data.

Acknowledgments

Conflict of Interest: P.H. has served as an advisor or consultant to Astra-Zeneca Pharmaceuticals, Bristol-Myers Squibb, Eli Lilly and Company, Johnson and Johnson, Inc., Memory Pharmaceuticals, Novartis Pharmaceuticals, Pfizer, Inc., Solvay-Wyeth Alliance, and the Sanofi-Aventis group. He has current or immediately prior grant or contract support from Astra-Zeneca Pharmaceuticals, Bristol Myers Squibb, Johnson and Johnson, Inc., and Pfizer, Inc. R.S.E.K. has received grant/research support from Astra-Zeneca, Eli Lilly, Johnson & Johnson, Pfizer, and NIMH, as well as providing educational services to Astra-Zeneca, Eli Lilly, Johnson & Johnson, Organon, and Pfizer. He has also served as a consultant and on advisory boards for various pharmaceutical companies as follows: Abbott Pharmaceuticals (advisory board), Acadia (consultant), Astra-Zeneca (advisory board, consultant), BiolineRx (consultant), Bristol Myers Squibb (advisory board, consultant), Cephalon (consultant), Cortex (consultant), Cyberonics (consultant), Dainippon Sumitomo Pharma (consultant), Eli Lilly (advisory board, consultant), Gabriel Pharmaceuticals (consultant), Johnson & Johnson (advisory board, consultant), Lundbeck/Solvay/Wyeth (advisory board, consultant), Memory Pharmaceuticals (advisory board, consultant), Merck (advisory board, consultant), Orexigen (advisory board, consultant), Organon (advisory board, consultant), Otsuka (consultant), Pfizer (advisory board, consultant), Saegis (advisory board, consultant), Sanofi/Aventis (advisory board, consultant), and Xenpo (consultant). In addition, R.S.E.K. receives royalties from the Brief Assessment of Cognition in Schizophrenia (BACS) testing battery and the MATRICS Battery (BACS Symbol Coding).

References


