Assessing Clinical and Functional Outcomes in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Schizophrenia Trial

by Marvin S. Swartz, Diana O. Perkins, T. Scott Stroup, Joseph P. McEvoy, Jennifer M. Nieri, and David C. Haak

Abstract

Schizophrenia is a symptomatically heterogeneous disorder characterized by the presence of positive and negative symptoms, and variable impairment in community functioning. Given the diversity of symptom presentations and functioning associated with schizophrenia, one of the key challenges facing the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial was the selection of efficient assessment measures appropriate to a community-based effectiveness trial. This article describes the rationale for the measurement approach adopted for the trial, provides a brief overview of the selected measures, and describes the process of training assessment raters for a large and geographically dispersed study group.

Keywords: Outcome measures, psychopathology assessment, functional assessment, schizophrenia treatment.

Schizophrenia Bulletin, 29(1):33-43, 2003.

Schizophrenia is a symptomatically heterogeneous disorder characterized by the presence of positive symptoms, including hallucinations and delusions. Several other symptom domains are part of the schizophrenia syndrome, including negative symptoms (e.g., decreased motivation and drive, decreased social interests, and diminished emotional range and expression of emotions), disorganization of thought process and behavior, mood symptoms, and impairments in social and cognitive function. While there is continued debate over which symptoms should be included in each domain, and whether the symptom domains are measuring discrete areas of dysfunction, there is little disagreement that these areas of function are frequently disrupted in patients with schizophrenia.

The medications used to treat schizophrenia often have side effects, and these side effects may in themselves cause symptoms similar to those of schizophrenia. For example, cognitive slowing, as well as negative symptoms such as decreased motivation and emotional dulling, may occur as part of an antipsychotic-induced parkinsonian syndrome. Patients may develop side effects that are new symptoms, such as galactorrhea, disrupted sexual function, or sedation. In some instances, the subjective distress and functional impairment that result from the medication side effect may be equal to or even worse than the symptoms of schizophrenia.

Often it is the functional impairment associated with schizophrenia that is the focus of clinical attention. Schizophrenia may affect people's ability to interact with others, to care for themselves, and to work, typically resulting in significant functional disability. The illness may also affect family members, because individuals with schizophrenia may rely heavily on others for assistance.

Given the diversity of symptom presentations and functioning associated with schizophrenia, one of the key challenges facing the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial was the selection of efficient assessment measures appropriate to a community-based effectiveness trial. Arguably, a "real world" trial should employ only measures of direct relevance to clinical practice. However, the CATIE trial also offers an opportunity to test the efficiency of assessment measures and potentially introduce new assessments that might represent novel assessment domains and methods.

Numerous factors were considered in choosing the specific assessments for the CATIE project. Consideration was given to the amount of time required to complete the assessments because there is a reasonable limit to the

amount of time that a subject would be willing to contribute to the study. In addition, the costs of conducting the study are in large part directly related to the number and frequency of study assessments. Thus, the level of detail and breadth of the study assessments were balanced against the time required to collect the information from the study subject. Consideration was also given to the relevance of domains to treatment adherence, the reliability of the instrument, and the frequency with which the instrument has been previously used in clinical and community studies of schizophrenia.

The process of measure selection in the CATIE schizophrenia trial proceeded by identifying the measurement domains with the most salience to treatment outcomes, proposing available measures for these domains, and reducing areas of overlap and duplication. After initial pilot testing, the protocol assessment measures were then broadly discussed with external consultants and with the CATIE schizophrenia trial center investigators. These discussions led to recommendations for refined measurement and further reduction of unnecessary and redundant measures. In the end, the clinical and functional outcome measures selected and discussed below, although the minimum needed to assess all important and relevant outcome domains, are more comprehensive than would be utilized for a simple effectiveness trial. Certain measures not essential to evaluating effectiveness per se were retained to create a rich data depository for use by the broader scientific community.

This article will describe the CATIE schizophrenia trial's measures of assessment, which are listed in table 1 and discussed in detail below. Assessment frequency and timing are described in the article by Stroup et al. in this issue.

Clinical Outcome Measures: Primary Outcome

Treatment Discontinuation. The primary aim of CATIE is to evaluate the effectiveness of selected antipsychotic medication to treat schizophrenia. Effectiveness considers the impact of the medication on controlling the various symptoms of the illness, as well as the safety and tolerability of the medication. The overall measure of effectiveness proposed for the CATIE schizophrenia study is "all-cause treatment discontinuation"—in other words, how long the patient remained on the assigned antipsychotic, and what caused any discontinuation. It is assumed that patients and clinicians "vote with their feet" and that a decision to continue or stop medication reflects the combined evaluation of efficacy and safety/tolerability of the treatment by the patient and clinician. Causes of treatment discontinuation are assessed and rated at the point in time

when the treating clinician and/or the patient decide to, or are forced to, discontinue treatment with the current antipsychotic medication. The primary outcome analyses in the schizophrenia trial will address the proportions of patients remaining in treatment with their initially assigned antipsychotic regimen over time.

This outcome measure was selected as the primary outcome because it rates, in aggregate, multiple aspects of pharmacological effectiveness, tolerability, and other real-world circumstances that often impinge on treatment decisions. The measure is discrete, relatively straightforward to score, and clinically meaningful.

Clinician and/or patient decisions about changes in the treatment regimen are coded on the treatment discontinuation measure. In addition to capturing whether and when patients discontinue treatment prior to completion of the study, we attempt to capture why they discontinue treatment, using a sequential selection approach.

Administrative discontinuation occurs when both the clinician and the patient are satisfied with treatment and plan to continue it but an *independent external event* interferes with the logistics of study implementation to such a degree that early discontinuation becomes necessary (e.g., the patient moved to another State because the patient's parent took a new job). Other examples of independent external events include pregnancy, incarceration in a penal institution where access is not readily feasible, and unrelated death.

Other than for administrative reasons, treatment discontinuation or change in regimen occurs because of *dissatisfaction with pharmacological treatment* on the part of the clinician, the patient, or in rare circumstances, a proxy decision maker termed the subject advocate in the CATIE schizophrenia trial.

In rating treatment discontinuation, the clinician's appraisal of the treatment regimen takes precedence in the ratings. For example, if the clinician recommends discontinuing the treatment, this discontinuation is coded as a clinician determination and not as a patient decision or "subject advocate" decision even if those individuals also want to discontinue the treatment. If the clinician is dissatisfied with the assigned antipsychotic treatment, he or she must also decide whether the dissatisfaction primarily reflects inadequate therapeutic efficacy or unacceptable side effects.

A patient may decide to discontinue treatment, even if the clinician remains satisfied and would prefer to continue. When a patient refuses treatment with the assigned antipsychotic but accepts treatment with another, a clear treatment discontinuation date can be determined. In some cases, however, patients simply fail to return for followup. In such cases, a period of up to 60 days is allowed before declaring treatment discontinuation (in such cases, a patient decision), during which time the clinician will

Table 1. CATIE schizophrenia trial centers' clinical and functional outcome measures

Measure	Content	Administration
Treatment discontinuation	Queries clinician on cause of treatment discontinuation	Experienced, certified clinician
Structured Clinical Interview for *DSM–IV*	Modules A–F (mood and psychotic diagnoses, substance use disorders, and anxiety disorders)	Experienced, certified clinician
Positive and Negative Syndrome Scale	Positive symptoms: 7 items Negative symptoms: 7 items General psychopathology: 16 items Items scored 1–7	Experienced, certified clinician
Clinical Global Impressions	Overall psychopathology summary measure; items scored 1–7	Experienced clinician
Calgary Depression Rating Scale	Mood symptoms: 9 items Items scored 1–4	Experienced clinician
Quality of Life Scale	Evaluates quality of life with emphasis on clinician-rated measurement of deficit symptoms; 21 items each on a 7-point scale; 4 subscales: interpersonal relations (8 items), instrumental role (4 items), intrapsychic foundations (7 items), common objects and activities (2 items)	Experienced clinician
Insight and Treatment Attitudes Questionnaire	Assesses subject's awareness of illness and willingness to adhere to treatment; 11 items scored 0–2	Nonclinician
Drug Attitude Inventory	Assesses the patient's subjective response to medication; focuses on unpleasant and negative subjective responses using a true/false format; 10 items	Nonclinician
Alcohol Use/Drug Use Scale	Overall rating of substance use on 2 items; items scored 1–5	Experienced clinician
Dartmouth Assessment of Lifestyle Instrument	Evaluates alcohol, marijuana, cocaine, and tobacco use with 18 items (4 items not scored, 14 items included in final score)	Experienced clinician
MacArthur Abbreviated Community Violence Instrument	Assesses several categories of violent behavior in recent past, including battery that results in physical injury, sexual assault, weapon use, and threat with weapon in hand	Nonclinician
Treatment compliance	Adapted compliance measures from available instruments; use all available data	Nonclinician
SF–12 Health Survey	Adapted from longer SF–36 to assess perceived general health status; 12 items; domains: effects of health on physical functioning, bodily pain, general health, vitality (energy/fatigue), social functioning, emotional role functioning, and mental health	Nonclinician
Adverse events/side effects	Assesses the occurrence of adverse events and common side effects in patients receiving antipsychotics	Experienced clinician
Barnes Akathisia Rating Scale	Measures drug-induced akathisia; 4-item fully anchored scale; includes global rating	Experienced clinician

Table 1. CATIE schizophrenia trial centers' clinical and functional outcome measures—*Continued*

Measure	Content	Administration
Abnormal Involuntary Movement Scale	Assesses the occurrence of dyskinesia in patients receiving antipsychotics; 12-item anchored scale	Experienced clinician
Simpson-Angus Extrapyramidal Side Effect Scale	Evaluates extrapyramidal side effects (includes dystonia item); 10 items for assessing parkinsonian and related symptoms	Experienced clinician
Family/Caregiver Experience Interview	Interviews family member or caregiver about subjective and objective dimensions of family involvement with subject	Nonclinician
MacArthur Competence Assessment Tool–Clinical Research	Assesses subject's capacity to make a competent decision about research participation; assesses understanding, appreciation, reasoning, and communication about the research protocol	Experienced clinician

Note.—CATIE = Clinical Antipsychotic Trials of Intervention Effectiveness.

attempt to reconnect with the patient and to convince him or her to resume treatment.

If a patient becomes unable to make treatment decisions, the subject advocate may decide to discontinue treatment, even if the clinician and patient remain satisfied and would prefer to continue. This is appropriately scored as a subject advocate treatment discontinuation.

In addition to its importance as the primary treatment outcome measure, the reason for treatment discontinuation serves as a prompt to which pathway patients are referred to in phase II of the protocol. This means that patients who stop the medication because of lack of efficacy are referred to the efficacy pathway of phase II offering the possibility of clozapine and other atypical drugs, whereas patients who stop for reasons of tolerability are referred to the pathway where they may receive risperidone or other atypical drugs.

Clinical and Functional Outcome Measures: Secondary Outcomes

The clinical and functional outcome measures described below are secondary outcome measures in the CATIE schizophrenia trial and were chosen to evaluate diagnostic eligibility, symptom domains, functional consequences, and medication side effects that may occur in schizophrenia.

Diagnostic Eligibility. Diagnostic eligibility is determined with the Structured Clinical Interview for *DSM–IV* (SCID; First et al. 1995). The SCID is a clinician-administered semistructured interview designed to evaluate

DSM–IV Axis I diagnoses. Use of a semistructured interview such as the SCID has been shown to improve the reliability of diagnostic assessments and thus helps ensure that all patients included in the study do, in fact, meet *DSM–IV* criteria for schizophrenia (Skre et al. 1991; APA 1994; Segal et al. 1994; Ventura et al. 1998). Because the SCID is completed by a trained clinician who may rely on medical records, staff reports, and information from caretakers, an accurate diagnostic picture may be obtained even when the patient is limited in ability to provide accurate self-report, as may be true for patients that are severely disorganized or cognitively impaired. For this study, SCID modules A through F—assessing mood, psychotic, substance abuse, and anxiety disorders—are administered.

Psychotic Symptom Severity. Symptom severity and functional status are assessed using a range of assessments. The primary schizophrenia trial assessment instrument for psychotic symptoms is the Positive and Negative Syndrome Scale (PANSS) (Kay et al. 1987) total score. The PANSS contains 30 items that assess symptoms of psychotic disorders, including positive, negative, and general psychopathology (Kay et al. 1987, 1989). The PANSS was chosen because of its widespread use in clinical studies of psychosis and its demonstrated reliability in assessing psychopathology across a range of patient populations. Items from the PANSS include those from the Brief Psychiatric Rating Scale (BPRS), with additional items from the Psychopathology Rating Schedule, and a valid BPRS can be extracted from a completed PANSS (Bell et al. 1992). The PANSS was developed to improve

on the BPRS by including additional symptoms that are clinically important in assessing schizophrenia. To enhance reliability, the PANSS includes a well-developed anchor system.

General Clinical Status. The Clinical Global Impressions (CGI) (Guy 1976) is also a widely used clinician-rated measure that assesses global symptom severity and is particularly helpful for repeated evaluations of global psychopathology. The CGI is chosen for inclusion as a measure of clinically meaningful symptom severity. The CGI is a single Likert scale rating severity of psychopathology from 1 to 7.

Depressive Symptomatology. Co-occurring depression can demonstrably affect outcomes in schizophrenia. The Calgary Depression Rating Scale (CDRS) is a nine-item scale designed to assess severity of depressive symptoms in patients with schizophrenia (Addington et al. 1990, 1992). The CDRS has been shown to be a reliable and valid measure of depression in this population (Addington et al. 1992). The CDRS improves the validity of assessment of depressive symptoms in patients with schizophrenia as compared with the Hamilton Depression Rating Scale, largely because the CDRS does not contain items that overlap with negative symptoms of schizophrenia (e.g., loss of interest, social withdrawal) (Addington et al. 1994, 1996).

Quality of Life. Psychosocial functioning and quality of life are assessed using the Quality of Life Scale (QLS) (Heinrichs et al. 1984), a clinician-rated scale of social functioning, interpersonal relationships, and intrapsychic well-being. It contains 21 items with 4 subscales: interpersonal relations, instrumental roles, intrapsychic foundations, and common objects and activities. The QLS demonstrates sensitivity to change over time and is a widely validated measure used in clinical and effectiveness trials in schizophrenia. In addition, multiple items from the Quality of Life Interview (Lehman and Burns 1990; Lehman 1995) are also utilized.

Insight Into and Awareness of Illness. Insight into or awareness of illness can be understood as a self-awareness that certain unusual perceptions, thoughts, emotions, or behaviors are indicators of an illness that requires mental health treatment. Poor insight is common in patients with schizophrenia, and studies have demonstrated associations between lack of insight and poorer treatment compliance (see McEvoy et al. 1989 for a review). The Insight and Treatment Attitudes Questionnaire (ITAQ) (McEvoy et al. 1989) was designed to measure awareness of illness and insight into need for treatment in patients with schizo-

phrenia. The ITAQ is a single scale consisting of 11 items that are phrased as questions to elicit open-ended responses from patients.

Subjective Responses to Medication. The Drug Attitude Inventory (DAI) assesses the patient's subjective response to medications (Hogan et al. 1983) and is included in the trial as a potential predictor of treatment compliance and other outcomes. The instrument focuses on unpleasant and negative subjective responses that are common adverse effects of antipsychotic medications. The DAI is brief and frequently self-administered but in the CATIE trial is read to the patient and true and false responses obtained. The DAI is reliable, and previous studies have shown that it is able to predict medication compliance in patients with schizophrenia (Hogan and Awad 1992; Weiden et al. 1994; Awad et al. 1997).

Alcohol and Drug Use. Alcohol and drug use comorbidity is a key risk factor for violence, noncompliance, relapse, and other poor outcomes in schizophrenia (Swartz et al. 1998). Co-occurring alcohol and drug use is assessed in the schizophrenia trial primarily by the Alcohol Use and Drug Use Scales (AUS/DUS) (Drake et al. 1990), which are widely used and accepted clinician rating scales for this population. Each clinician rating scale utilizes all sources of available information (patient, family, case manager, other clinicians, etc.) to rate use of alcohol and drugs from abstinence to dependence requiring institutional care. In addition, use of specific drugs of abuse is also rated. The schizophrenia trial also employs the Dartmouth Assessment of Lifestyle Instrument (DALI), a self-report measure that assesses drug and alcohol use in severely mentally ill individuals (Rosenberg et al. 1998). The DALI assesses problems related to use of alcohol, marijuana, cocaine, and tobacco. The DALI is a sensitive and reliable measure of substance use and is used at baseline to improve detection of substance use.

Violent Behavior. Violent behavior and victimization are key measures of and risk factors for poor outcomes in schizophrenia. Both violence and victimization are assessed using the MacArthur Abbreviated Community Violence Instrument (Steadman et al. 1998). The MacArthur instrument asks subjects and collateral informants whether or not the subject has engaged in (or been the target of) several categories of aggressive behavior during a past period of reference. If a positive response is given, detailed probing may be conducted to obtain more specific information about each incident (i.e., regarding seriousness, context, precipitating circumstances, associated subjective states, substance use, and medication adherence at the time). These measures will be used to

longitudinally assess the relationship between violence and symptom burden, recovery, relapse, and severity of illness.

Treatment Compliance. Treatment compliance represents an expected intermediate level of outcome in the schizophrenia trial and is measured by the proportion of scheduled appointments kept. Information about appointment compliance is recorded in followup visit records. Information specific to medication compliance is obtained from the patient, the clinician, and the family according to methods modified from Kelly et al. (1987). This information is also supplemented with pill counts. While each method of measuring medication adherence is imperfect, multiple methods should dramatically improve estimates of adherence (Swartz et al. 1998). A summary rating of overall medication compliance is derived from all data available to clinical raters.

General Health Status and Functioning. The general health and functioning of persons with schizophrenia is a growing concern, given the general health of the population and concern with adverse health outcomes associated with antipsychotic medications. The 12-item questionnaire from the Medical Outcomes Study (Short form [SF]–12) was derived from the parent SF–36 and is used to measure general health status and functioning in six domains. This is an extensively validated instrument with normative data from large numbers of individuals from both clinical samples and the general population. The SF–12 includes six scales measuring the following health concepts: the effects of health on physical functioning; freedom from bodily pain; mental health; and overall perception of general health. The SF–12 has become a standard for measuring health status and quality of life in clinical research (Ware et al. 1996).

Side Effects Measures

In most schizophrenia medication treatment trials, medication side effects are determined by the frequency of spontaneous complaints, referred to as "adverse events." Adverse events that occur more commonly with the study medication than with placebo are considered side effects of that medication. For example, patients frequently complain of being "tired," whether on a drug or on a placebo. It is only when complaints of being "tired" occur more commonly with drug-treated than with placebo-treated patients that the adverse event "tiredness" is considered a medication side effect. A drawback to this method of determining the relative frequency of side effects is that patients often may not volunteer the occurrence of an adverse event, especially if the event involves a potentially embarrassing

symptom such as change in sexual functioning. In addition, patients may not spontaneously recall events when asked in a general inquiry.

Adverse Events and Side Effects. In CATIE, the relative frequencies and severity of common side effects are determined across the study antipsychotics by directly asking subjects about the presence or absence of common antipsychotic side effects, including orthostatic faintness, dry mouth, constipation, sialorrhea, menstrual irregularities, gynecomastia/galactorrhea, diminished sex drive, diminished sexual arousal, decreased likelihood of sexual orgasm, incontinence/nocturia, urinary hesitancy, skin rash, sleepiness, hypersomnia, weight gain, insomnia, akathisia, and akinesia. Both the study clinician's and the patient's assessment of severity are recorded, and subjects are asked about the impact of a side effect on willingness to continue taking the study medication. Spontaneous complaints of potential medication side effects also are recorded, as is commonly done in pharmacological treatment trials.

Akathisia. Akathisia is a common and poorly tolerated adverse effect of typical antipsychotics and is known to occur with atypical antipsychotics, especially at higher doses. The Barnes Akathisia Scale (BAS) contains four items, including objective akathisia, subjective awareness of restlessness, subjective distress related to restlessness, and a global clinical assessment of akathisia. The BAS was chosen to measure the akathisia because of its widespread use in clinical trials and demonstrated reliability (Barnes 1989; Sachdev 1994; Tracy et al. 1997). The reliability of the BAS has also been demonstrated in elderly psychiatric patients (Sweet et al. 1993).

Tardive Dyskinesia. Tardive dyskinesia is a potentially serious adverse effect of antipsychotics. While there is some indication that the atypical antipsychotics have reduced risk of tardive dyskinesia, the relative risk of tardive dyskinesia among the atypical antipsychotics and compared with perphenazine is not known. The Abnormal Involuntary Movement Scale (AIMS) has twelve items that rate severity of dyskinetic movements in various body parts and an overall severity item, on a scale ranging from 0 to 4. The AIMS was also chosen to measure severity of dyskinetic movements because of its reliability and widespread use (Sweet et al. 1993; Tracy et al. 1997).

Extrapyramidal Symptoms. Drug-induced parkinsonism is a common and poorly tolerated adverse effect of typical antipsychotics and occurs with atypical antipsychotics, especially at higher doses. There are few comparisons of drug-induced parkinsonism among atypical antipsychotics. The Simpson-Angus Extrapyramidal Side Effect

Scale (SAEPS) contains 10 items rated on a scale of 0 to 4. The SAEPS is widely used and is reliable and valid (Tracy et al. 1997).

Other Measurement Domains

Family and Caregiver Involvement With the Patient. The Family/Caregiver Experience Interview assesses involvement with the patient at baseline and the impact of schizophrenia on the objective and subjective burden on the family. The instrument was adapted from an interview developed by Deborah Franks (Franks 1990) for a survey of Alliance for the Mentally Ill families in Massachusetts and from instruments developed by Creer et al. (1982). This assessment, administered in an interview format, is completed by a family member or caregiver accompanying the patient at baseline and followup assessments. Families and caregivers are asked about instrumental support provided in the previous month, limit setting for disturbing behavior, time directly related to care for the patient, lost productivity, and financial contributions to the subject's care.

Competence Assessment. In the CATIE schizophrenia trial the MacArthur Competence Assessment Tool–Clinical Research (MacCAT–CR) is used to assess prospective participants' capacity to make a competent decision about research participation. Only persons judged to have adequate decisional capacity are allowed to enroll in the study.

The MacCAT–CR provides clinical researchers with a semistructured interview format with which to assess and rate the decisional capacity of potential research subjects related to the following four standards for competence:

- *understanding* of disclosed information about the nature of the research project and its procedures
- *appreciation* of the effects of research participation (or failure to participate) on the subjects' own situation
- *reasoning* in the process of deciding about participation, focusing on the subjects' abilities to compare alternatives in light of their consequences
- *communication* of a decision about research participation

The MacCAT–CR provides a format for disclosure of selected information that describes the CATIE trial. A standard set of questions is asked to sample subjects' abilities to understand and appreciate the information, to reason about it, and to express a choice.

A threshold score of 15 on the MacCAT-CR understanding scale is the minimum acceptable score for a person to provide informed consent for CATIE. Some people with a higher score will be judged by clinical personnel to lack the capacity to provide informed consent. Clinical judgment about competency is the final standard for study participation, but a minimum score of 15 or more on the MacCAT–CR understanding scale must be achieved before participation.

Although an assessment of a potential subject's ability to understand, appreciate, reason, and make a choice may be necessary (depending on the legal criteria for decision-making competence in effect in a particular jurisdiction), it is not sufficient to make a judgment about a potential subject's capacity to make a competent decision about participation in research. To reach valid conclusions about subjects' decision-making capacity, MacCAT–CR information must be supplemented with clinical information and knowledge of the medical and social circumstances in which the subject's decision is being made. Therefore, the MacCAT–CR is not the only information to be considered in assessing someone's capacity to make a competent decision about participation in CATIE. Clinical information is also considered. The investigator's clinical assessment is the ultimate determinant of a subject's capacity to consent to participate in CATIE.

Study Training Procedures

Rater training minimizes rater error and thus enhances the overall quality and statistical power of clinical trials (Tracy et al. 1997; Muller and Wetzel 1998; Perkins et al. 2000). Because of these important benefits, the CATIE schizophrenia trial utilizes a clinical rater training and certification process. However, the challenges to developing and administering a rater-training program for a large-scale multicenter clinical trial are many. Our initial goal was to design a training framework that would balance the time, cost, and effort required for training a large number of raters from 54 different sites with the need for high-quality reliable and accurate ratings.

Typically, rater training begins with a review of study protocol and clinical assessment procedures (Tracy et al. 1997; Salyers et al. 2001). For the CATIE schizophrenia trial, the coordinating center organized an initial 3-day training event that occurred prior to the implementation of the trial. While a study start-up meeting is an efficient strategy to initially orient site staff to the protocol and provide study procedure training, other strategies are needed to maintain reliability over time and to manage personnel turnover (Edson et al. 1997). Also, the initial training event does not take into account differing skill levels, experience, or other individual differences among the training participants (Müller et al. 1998).

Therefore, a Web-based training program was developed for the CATIE trial in an attempt to address the challenges of initial and ongoing clinical staff training for this longitudinal, multisite clinical trial, as well as to moderate the cumbersome task of material distribution to 54 clinical sites. The initial training event was digitally captured (video and audio) and then placed on a secure website. The website streams information using a wide bandwidth to allow relatively quick access times at the high resolution necessary for evaluation of taped clinical interviews. Confidential material was protected by using a nonpublished universal resource locater, password protection, real access media encoding, and, finally, Internet protocol address verification. Because of the size of this trial, it was inevitable that some sites would not be able to access the online training site; therefore a compact disc replicating the website content was created and distributed to the sites. As a result, the website became a resource for rater training, is quite efficient and effective for training new study personnel, and supports retraining of existing personnel. The online training pages are self-guiding "click-through" units that include video and audio streaming, edited transcripts, and supporting materials in downloadable Adobe print direct format files. An automatic "detect and respond" system ensures that users have the appropriate software to view the lectures and guides users through updates when necessary.

The CATIE schizophrenia trial chose to require initial and ongoing certification on the SCID, PANSS, and treatment discontinuation instruments so as to follow the empiric guidelines for ongoing interrater reliability monitoring promulgated by Tracy et al. (1997) and Warshaw et al. (2001). Use of the Web-based training program facilitated these efforts. The website and electronic mail facilitated distribution of procedural guidelines and reliability materials. To track the distribution of materials, personnel turnover, certification status, and reliability data, a partially automated data base was developed. Site personnel can use an electronic form on the website to access their certification status and update personnel listings. This information can then be directly imported into the data base, automatically updating existing information for each rater. Maintaining a data base provides quick feedback to all raters regarding their status and ratings—an important part of maintaining interrater reliability (Tracy et al. 1997; Igarashi et al. 1998).

Rater Certification and Rationale

Our rater training efforts were 3-fold: (1) initial training on all study instruments; (2) certification training on the PANSS, SCID, and treatment discontinuation measures; and (3) yearly recertification on these measures as a means

to minimize rater drift and maximize rater training over the duration of the study. Training efforts were enhanced by the fact that most of the clinical raters had a master's degree and/or significant clinical experience with schizophrenia patients.

For the PANSS, raters were required to view and rate three taped PANSS interviews, achieving a Pearson's correlation coefficient (PCC) ≥ 0.7 when compared with the expert ratings for each interview. Keeping in mind the amount of time and effort required for training and certification, we chose to start the process with the minimum number of tapes necessary to generate a PCC, providing remediation and additional PANSS interview tapes for those who did not pass certification initially. To demonstrate competency on the SCID and the treatment discontinuation instrument, raters were given a set of ten vignettes to score, requiring a correct completion rate of 80 percent for the SCID and 90 percent for the treatment discontinuation instrument. Both the SCID and the treatment discontinuation instrument had failure rates of less than 4 percent.

The training program has proven to be successful. Of the initial 176 PANSS raters, 80 percent passed with a PCC ≥ 0.7. Of that passing group, over 90 percent had an excellent PCC of ≥ 0.8 (Kay 1991; Norman et al. 1996). Less than 10 percent ($n = 13$) had a PCC between 0.7 and 0.8 and were classified as a certified PANSS rater with the understanding that additional reliability PANSS tapes would be offered over the next 12 months as a means to improve PCC scores and reliability. At the end of this 12-month period, all of the raters in this group who had continued in their positions as active raters ($n = 7$) raised their PCC score to ≥ 0.8.

To prevent rater drift and maintain training effects and thus reliable ratings, each site was sent a new PANSS interview tape to view and score every 4 months, again requiring a PCC ≥ 0.7. Through minimal training efforts we were able to remind raters of the rating rules and thereby reduce rating error. We were aware that this process did not take into account the quality of the individual raters' clinical interviews but rather emphasized that raters consistently use anchor points and scoring rules as defined by the instrument. Current PANSS raters who have completed the first-year reliability tapes all have PCC scores ≥ 0.8.

Discussion

The approach to clinical and functional outcomes assessment adopted for the CATIE schizophrenia trial is designed to meet the needs of a geographically dispersed community-based effectiveness study. The measurement approach attempts to give equal weight to clinical and

functional outcomes as efficiently as possible. While the measures adopted are more detailed than would be needed for a simple effectiveness trial, the rich data collected should prove valuable to a broad range of investigators interested in treatment outcomes in this population.

References

Addington, D.; Addington, J.; and Atkinson, M. A psychometric comparison of the Calgary Depression Scale for Schizophrenia and the Hamilton Depression Rating Scale. *Schizophrenia Research*, 19:205–212, 1996.

Addington, D.; Addington, J.; and Maticka-Tyndale, E. Specificity of the Calgary Depression Scale for schizophrenics. *Schizophrenia Research*, 11:239–244, 1994.

Addington, D.; Addington, J.; Maticka-Tyndale, E.; and Joyce, J. Reliability and validity of a depression rating scale for schizophrenics. *Schizophrenia Research*, 6:201–208, 1992.

Addington, D.; Addington, J.; and Schissel, B. A depression rating scale for schizophrenics. *Schizophrenia Research*, 3:247–251, 1990.

American Psychiatric Association. *DSM–IV: Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: APA, 1994.

Awad, A.G.; Vornganti, L.; and Heslegrave, R.J. A conceptual model of quality of life in schizophrenia: Description and preliminary clinical validation. *Quality of Life Research*, 6:21–26, 1997.

Barnes, T.R. A rating scale for drug-induced akathisia. *British Journal of Psychiatry*, 154:672–676, 1989.

Bell, M.; Milstein, R.; Beam-Goulet, J.; Lysaker, P.; and Cicchetti, D. The Positive and Negative Syndrome Scale and the Brief Psychiatric Rating Scale: Reliability, comparability, and predictive validity. *Journal of Nervous and Mental Disease*, 180:723–728, 1992.

Creer, C.; Sturt, E.; and Wykes, T. The role of relatives. In: Wing, J.K., ed. *Long-Term Community Care: Experiences in a London Borough. Psychological Medicine*, Monograph Suppl. 2, 1982.

Drake, R.; Osher, F.C.; Noordsy, D.L.; Hurlbut, S.C.; Teague, G.B.; and Beaudett, M.S. Diagnosis of alcohol use disorders in schizophrenia. *Schizophrenia Bulletin*, 16(1):57–67, 1990.

Edson, R.; Lavori, P.; Tracy, K.; Adler, L.A.; and Rotrosen, J. Interrater reliability issues in multicenter trials: II. Statistical procedures used in Department of Veterans Affairs Cooperative Study #394. *Psychopharmacology Bulletin*, 33:59–67, 1997.

First, M.B.; Spitzer, R.L.; Gibbon, M.; and Williams, J.B.W. *Structured Clinical Interview for DSM–IV (SCID-I)*

(User's Guide and Interview) Research Version. New York, NY: Biometrics Research Department, New York Psychiatric Institute, 1995.

Franks, D.D. Economic contribution of families caring for persons with severe and persistent mental illness. *Administration and Policy in Mental Health*, 18:9–18, 1990.

Guy, W. *ECDEU Assessment Manual for Psychopharmacology—Revised*. DHEW Publication No. ADM 76–338. Rockville, MD: U.S. Department of Health, Education, and Welfare; Public Health Service; Alcohol, Drug Abuse, and Mental Health Administration; NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976.

Heinrichs, D.W.; Hanlon, T.E.; and Carpenter, W.T. The Quality of Life Scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin*, 10(3):388–396, 1984.

Hogan, T.P., and Awad, A.G. Subjective response to neuroleptics and outcome in schizophrenia: A re-examination comparing two measures. *Psychological Medicine*, 22:347–352, 1992.

Hogan, T.P.; Awad, A.G; and Eastwood, R. A self-report scale predictive of drug compliance in schizophrenics: Reliability and discriminative validity. *Psychological Medicine*, 13:177–183, 1983.

Igarashi, Y.; Hayashi, N.; Yamashina, M.; Otsuka, N.; Kuroki, N.; Anzai, N.; and Kazamatsuri, H. Interrater reliability of the Japanese version of the Positive and Negative Syndrome Scale and the appraisal of its training effect. *Psychiatry and Clinical Neurosciences*, 52:467–468, 1998.

Kay, S.R.; Fiszbein, A.; and Opler, L.A. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2):261–276, 1987.

Kay, S.R.; Opler, L.A.; and Lindenmayer, J.P. The Positive and Negative Syndrome Scale (PANSS): Rationale and standardisation. *British Journal of Psychiatry*, 7(Suppl):59–67, 1989.

Kelly, G.R.; Mannon, J.A.; and Scott, J.E. Utility of the health belief model and examining medication compliance among psychiatric outpatients. *Social Science and Medicine*, 25:1205–1211, 1987.

Lehman, A.F. Measuring QOL in a reformed health system. *Health Affairs*, 45:91–101, 1995.

Lehman, A.F., and Burns, B.J. Severe mental illness in the community. In: Spiker B., ed. *Quality of Life Assessment in Clinical Trials*. New York, NY: Raven Press, 1990. pp. 357–366.

McEvoy, J.P.; Apperson L.J.; Appelbaum, P.S.; Ortlip, P.; Brecosky, J.; Hammill, K.; Geller, J.L.; and Roth, L.

Insight in schizophrenia: Its relationship to acute psychopathology. *Journal of Nervous and Mental Disease*, 177:43–47, 1989.

Müller, M.J.; Rossbach, W.; Dannigkeit, P.; Müller-Siecheneder, F.; Szegedi, A.; and Wetzel, H. Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research*, 32:151–160, 1998.

Müller, M.J., and Wetzel, H. Improvement of interrater reliability of PANSS items and subscales by standardized rater training. *Acta Psychiatrica Scandinavica*, 98:135–139, 1998.

Norman, R.M.; Malla, A.K.; Cortese, L.; and Diaz, F. A study of the interrelationship between and comparative interrater reliability of the SAPS, SANS and PANSS. *Schizophrenia Research*, 19:73–85, 1996.

Perkins, D.O.; Wyatt, R.J., and Bartko, J.J. Penny-wise and pound-foolish: The impact of measurement error on sample size requirements in clinical trials. *Biological Psychiatry*, 47:762–766, 2000.

Rosenberg, S.D.; Drake, R.E.; Wolford, G.L.; Mueser, K.T.; Oxman, T.E.; Vidaver, R.M.; Carrieri, K.L.; and Luckoor, R. Dartmouth Assessment of Lifestyle Instrument (DALI): A substance use disorder screen for people with severe mental illness. *American Journal of Psychiatry*, 155:232–238, 1998.

Sachdev, P. A rating scale for acute drug-induced akathisia: Development, reliability, and validity. *Biological Psychiatry*, 35:263–271, 1994.

Salyers, M.P.; McHugo, G.J.; Cook, J.A.; Razzano, L.A.; Drake, R.E.; and Mueser, K.T. Reliability of instruments in a cooperative, multisite study: Employment Intervention Demonstration Program. *Mental Health Services Research*, 3:129–139, 2001.

Segal, D.L.; Hersen, M.; and Van Hasselt, V.B. Reliability of the Structured Clinical Interview for *DSM–III–R*: An evaluative review. *Comprehensive Psychiatry*, 35:316–327, 1994.

Skre, I.; Onstad, S.; Torgersen, S.; and Kringlen, E. High interrater reliability for the Structured Clinical Interview for *DSM–III–R* Axis I (SCID–I). *Acta Psychiatrica Scandinavica*, 84:167–173, 1991.

Steadman, H.J.; Mulvey, E.P.; Monahan, J.; Robbins, P.C.; Appelbaum, P.S.; Grisso, T.; Roth, L.H.; and Silver, E. Violence by people discharged from acute inpatient psychiatric facilities and others in the same neighborhoods. *Archives of General Psychiatry*, 55:393–401, 1998.

Swartz, M.S.; Swanson, J.W.; Hiday, V.A.; Borum, R.; Wagner, H.R.; and Burns, B.J. Violence and severe mental illness: The effects of substance abuse and nonadherence to medication. *American Journal of Psychiatry*, 155:226–231, 1998.

Sweet, R.A.; DeSensi, E.G.; and Zubenko, G.S. Reliability and applicability of movement disorder rating scales in the elderly. *Journal of Neuropsychiatry and Clinical Neurosciences*, 5:56–60, 1993.

Tracy, K.; Adler, L.A.; Rotrosen, J.; Edson, R.; and Lavori, P. Interrater reliability issues in multicenter trials: I. Theoretical concepts and operational procedures in VA Cooperative Study #394. *Psychopharmacology Bulletin*, 33:53–57, 1997.

Ventura, J.; Liberman, R.P.; Green, M.F.; Shaner, A.; and Mintz, J. Training and quality assurance with the Structured Clinical Interview for *DSM–IV* (SCID–I/P). *Psychiatry Research*, 79:163–173, 1998.

Ware, J., Jr.; Kosinski, M.; and Keller, S.D. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34:220–233, 1996.

Warshaw, M.G.; Dyck, I.; Allsworth, J.; Stout, R.L.; and Keller, M.B. Maintaining reliability in a long-term psychiatric study: An ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. *Journal of Psychiatric Research*, 35:297–305, 2001.

Weiden, P.; Rapkin, B.; Mott, T.; Zygmunt, A.; Goldman, D.; Horvitz-Lennon, M.; and Frances, A. Rating of Medication Influences (ROMI) scale in schizophrenia. *Schizophrenia Bulletin*, 20(2):297–310, 1994.

Acknowledgment

The Authors

Marvin S. Swartz, M.D., is Professor, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC. Diana O. Perkins, M.D., M.P.H., and T. Scott Stroup, M.D., M.P.H., are Associate Professors, Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, NC. Joseph P. McEvoy, M.D., is Associate Professor, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center. Jennifer M. Nieri, M.S.W., L.C.S.W., is Research Instructor, and David C. Haak, M.S., is Research Coordinator, Department of Psychiatry, University of North Carolina School of Medicine.