

Assessment of Community Functioning in People With Schizophrenia and Other Severe Mental Illnesses: A White Paper Based on an NIMH-Sponsored Workshop

Alan S. Bellack^{1,2}, Michael F. Green³, Judith A. Cook⁴, Wayne Fenton⁵, Philip D. Harvey⁶, Robert K. Heaton⁷, Thomas Laughren⁸, Andrew C. Leon⁹, Donna J. Mayo¹⁰, Donald L. Patrick¹¹, Thomas L. Patterson¹², Andrew Rose¹³, Ellen Stover¹⁴, and Til Wykes¹⁵

²University of Maryland School of Medicine, 737 West Lombard Street, Suite 551, Baltimore, MD 21201; ³University of California, Los Angeles; ⁴University of Illinois at Chicago; ⁵National Institute of Mental Health; ⁶Mt Sinai School of Medicine; ⁷University of California, San Diego; ⁸Director, Division of Psychiatry Products, FDA; ⁹Weill Medical College of Cornell University; ¹⁰National Institute of Mental Health; ¹¹National Institute of Mental Health; ¹²University of California, San Diego; ¹³American Institutes for Research; ¹⁴National Institute of Mental Health; ¹⁵Institute of Psychiatry, Kings College, London

People with schizophrenia frequently have significant problems in community functioning. Progress in developing effective interventions to ameliorate these problems has been slowed by the absence of reliable and valid measures that are suitable for use in clinical trials. The National Institute of Mental Health convened a workgroup in September 2005 to examine this issue and make recommendations to the field that would foster research in this area. This article reports on issues raised at the meeting. Many instruments have been developed to assess community functioning, but overall insufficient attention has been paid to psychometric issues and many instruments are not suitable for use in clinical trials. Consumer self-report, informant report, ratings by clinicians and trained raters, and behavioral assessment all can provide useful and valid information in some circumstances and may be practical for use in clinical trials. However, insufficient attention has been paid to when and how different forms of assessment and sources of information are useful or how to understand inconsistencies. A major limiting factor in development of reliable and valid instruments is failure to develop a suitable model of functioning and its primary mediators and moderators. Several examples that can guide thinking are presented. Finally, the field is limited by the absence of an objective gold standard of community functioning. Hence, outcomes must be evaluated in part by “clinical significance.” This

critterion is problematic because different observers and constituencies often have different opinions about what types of change are clinically important and how much change is significant.

Key words: schizophrenia/assessment/outcomes

Schizophrenia is frequently conceptualized in the context of severe psychotic symptoms, but the only symptom that DSM-IV requires of all cases is a deterioration (or failure to achieve adequate levels) of social functioning. As many as two-thirds of people with schizophrenia are unable to fulfill basic social roles, such as spouse, parent, and worker, even when psychotic symptoms are in remission. Fewer than one-third work regularly, and the majority are underemployed (based on premorbid functioning) even when they can work. Only a small percentage of persons with schizophrenia marry, and marriages often end in divorce. Most patients have significant impairments in social relationships, and they often are socially isolated. When they do interact with others, they often have difficulty maintaining appropriate conversations, expressing their needs and feelings, achieving social goals, or developing close relationships. People with schizophrenia have increased medical morbidity and early mortality, and it has been hypothesized that an important factor in their poor medical status is difficulty effectively relating to health care providers. More than 50% of people with schizophrenia meet criteria for substance abuse or dependence, and a major factor driving use of illicit drugs is the desire to seem like other people. Overall, social dysfunction is one of the most important factors in the disability associated with the illness and is a source of great distress for patients and family members.

Assessing Social Competence

While much is known about the impact of social dysfunction, much remains to be learned about its causes and potential interventions, especially pharmacological, that can significantly enhance social role functioning. Progress in this area has been slowed by lack of a well-established, widely accepted, and practical way of assessing functioning in the community. Progress in

¹To whom correspondence should be addressed; tel: 410-605-7383, fax: 410-605-7739, e-mail: abellack@psych.umaryland.edu.

both research on mechanisms and treatment development would be fostered by a comprehensive but relatively economical instrument for the assessment of social functioning that could be used both as a treatment outcome measure and as a valid standard of community functioning in laboratory studies of social competence or functional capacity.

The absence of a satisfactory scale is not a function of lack of effort in the field. A plethora of instruments have appeared in the literature in the last 20 years, including the Behavior and Symptom Identification Scale¹ (BASIS-32), the DSM-IV Social and Occupational Functioning Assessment Scale,² the Social Functioning Scale,³ the Social Performance Schedule,⁴ and the Multidimensional Scale of Independent Functioning.⁵ Most of these instruments, and variants that have been employed in individual studies, rely on either clinician ratings or patient report. In psychiatry, limited information is available from laboratory tests or physiologic measures, and clinician assessment has been and remains the foundation for outcomes assessment in severe mental illness. Clinician-rated outcomes may be evaluated by trained raters or health care providers involved in patients' care. These outcomes may be assessed through physical examination, patient interviews, and/or global impressions taking into consideration all sources of evidence and experience with the patient population. To date, patient-reported outcomes (PROs) refer to any reports coming directly from patients about their health conditions and treatment. PROs vary in the degree to which they can be verified or observed by others and in the nature of the event, behavior, experience, or feeling that is being reported. Patient behaviors or events are sometimes observable and theoretically reliably reported by others. Ratings of subjective worries or feelings of satisfaction, in contrast, are known only to patients themselves. These subjective ratings or reports may or may not be communicated or made visible to others, including families and clinicians.

Whether clinician or patient rated, all extant instruments have limitations, including problems in scope (breadth of coverage), reliability, validity, floor and ceiling, suitability for diverse populations (eg, both inpatients and outpatients), and applicability for both clinical trials and mechanism studies. As will be discussed further below, a satisfactory instrument for assessing social functioning in schizophrenia must satisfy a variety of difficult criteria. It should be applicable to the range of settings in which people with schizophrenia live, sample the full range of low to high performance in each item so that even relatively small changes can be detected, be designed with few "skip outs" (items rated not applicable or not observed), integrate information from a number of sources, and be straightforward enough to be completed by raters without sophisticated clinical training. We are not aware of an instrument for assessment of

social functioning with a proven track record that has these characteristics.

National Institute of Mental Health Workshop: Innovative Approaches to Evaluate Performance, Therapeutic Outcomes, and Recovery in People With Mental Disorders

In light of the importance of developing an effective measure (or measures) of social competence, the National Institute of Mental Health (NIMH) convened a workgroup on September 13, 2005 to identify the key issues in defining and assessing community functioning and begin to formulate a plan to develop suitable instruments. Panel members were selected to represent diverse areas of interest and expertise, including neuropsychology, rehabilitation, clinical trials, psychometrics, biostatistics, and behavioral assessment. The purpose of this article is to summarize the issues and positions discussed, report on recommendations that were made, and stimulate research in the area.

NIMH Rationale for Convening the Workshop

The mission of NIMH is to reduce the burden of mental illness through research on mind, brain, and behavior. In relation to this mission, a better understanding of functioning in individuals with schizophrenia and better methods of measuring functioning is important from a variety of perspectives. First, while most treatments for schizophrenia target symptoms, there is increasing recognition that deficits in functioning in the form of social isolation, unemployment, and impaired self-care represent a significant component of illness burden. While symptom control is an important treatment outcome, both patients and families consistently identify better social and occupational functioning as important self-defined treatment goals. From the patient and family perspective, enhanced functioning may be the most meaningful and valued outcome of treatment. Given the societal costs of poor functioning, it is also a priority for society at large.

Creating a new generation of interventions requires better models of the determinants of functioning. Similarly, testing these interventions requires more nuanced measures of functioning. Finally, it is increasingly clear that available and foreseeable interventions for schizophrenia have costs, both monetary and in terms of side-effect burden. Particularly, when resources are scarce or side-effect burdens are heavy, new treatments should provide real clinical, financial, and personal value. The ability of a treatment to enhance functioning must be a central part of the value risk-benefit or cost-effectiveness analysis.

To help NIMH achieve a better understanding of functioning in individuals with schizophrenia and determine better methods of measuring functioning, NIMH convened the workshop reported here. This workshop also responded to recommendations from the Behavioral

Table 1. Organizing Themes and Questions

A. Measuring community functioning	
1.	What dimensions of community functioning need to be assessed to effectively represent outcomes for people with schizophrenia?
2.	Are there reliable and valid measures of community functioning?
3.	How does one validate a measure of functioning?
4.	What methodological and statistical issues need to be considered in validating a new instrument?
5.	How much variance should these measures account for to be useful?
B. Assessment issues	
6.	Is it possible to develop measures of generalized functioning or do measures have to be situation/skill specific?
7.	Can proxy (clinic/laboratory) measures effectively represent community functioning?
8.	How do we address the need for repeated testing in clinical trials?
9.	What other innovative approaches have other fields (eg, human immunodeficiency virus) used to assess functioning?
C. Theoretical models	
10.	What are the current theoretical models for relationships among cognition, symptoms, motivation, capacity, and performance?

Science Workgroup of the National Advisory Mental Health Council. These recommendations, as reported in the NIMH document “Translating Behavioral Science Into Action,” called for increased research on understanding “how mental illnesses and their treatments affect the abilities of individuals to function in diverse settings and roles (eg, carrying out personal, educational, family, and work responsibilities).” Council also recommended increased research to “apply methods from basic behavioral science to the development of tools to assess functioning,” including “the assessment of functioning as an outcome in intervention, services, and risk-factor research.” The assessment strategies and basic research findings from these areas are then to be used to facilitate theory-driven research on the development of behavioral and psychosocial interventions—prevention and treatment—aimed at improving behavior change and functional outcome and reducing disability, morbidity, and mortality in people with mental disorders.

Organization of the Discussion

The workshop was organized around 3 themes and 10 questions that reflect important practical, methodological, and conceptual issues involved in assessment in general and assessment of social functioning specifically. The questions and themes are presented in table 1. The first question was intended to identify the key parameters of functional assessment, which would set the context for answering subsequent questions. Surprisingly, it proved to be the most difficult question to answer, sparking considerable debate among group members. First, there was disagreement about exactly what the term “community functioning” implies. Some members of the group were inclined to limit the term to overt behaviors and role function: what the person does and how well she/he does it. Examples would include work, activities of daily living (ADLs), and social roles such as friend, spouse, and parent. Another view was that it should include mechanisms or factors that influence behavior, especially including neurocognition and social cognition. A third domain that generated some controversy was subjective appraisal:

what the person feels about himself and his life. This domain is often conceived of as “quality of life.” As will be discussed further below, there are important distinctions between the person as a reporter of more or less objective phenomena (eg, “How many hours a week have you worked during the last month?”) vs the person’s subjective level of satisfaction with circumstances and behavior (eg, “How satisfied are you with how much you have worked”). Often these 2 dimensions are blurred, with questions such as follows: “On a 5-point scale, how would you rate your work over the last month?”

One major reason for the difficulty in reaching consensus on a definition for community functioning is that any specific definition necessarily reflects the scope and purpose of assessment. For example, an assessment to evaluate the efficacy of a skills training intervention would likely examine specific behavioral skills that were trained and role functioning that would be expected to improve if skills were increased. In contrast, assessment for a trial of a new antipsychotic expected to decrease negative symptoms would likely examine social drive or interest, anhedonia, and verbal output, rather than specific social skills. Similarly, a short-term efficacy trial of a new medication might not assess aspects of role functioning such as work and independent living, which would only be expected to change over time. In the context of this discussion, the workgroup deferred to NIMH to guide the scope and goals of assessment. In sponsoring the workshop, the NIMH was particularly interested in assessing treatment outcomes. Specifically, 2 central outcome criteria or domains were identified: the ability of treatments to (1) reduce the burden of mental illness (on the person, the family, and the community) and (2) increase the person’s independence. By restricting the focus to these domains, specific behavioral aspects of functioning that are frequently assessed, such as social skills and ability to perform ADLs, are only relevant to the extent that they have an impact on one or both of these outcomes. These behavioral dimensions can be considered proxies of everyday functioning, as was suggested in the NIMH-sponsored Measurement and Treatment

Research to Improve Cognition in Schizophrenia initiative to develop a consensus cognitive battery for use in clinical trials. Alternatively, they can be conceptualized as “mediators and moderators” in the language of theory and measurement.⁶ Focusing on treatment outcome, rather than behavior per se, establishes a number of important parameters for instrument development. For example, in order to be useful for clinical trials, a measure must be suitable for repeated assessments (ie, baseline, posttreatment, and follow-up), either by having minimal practice effects, by having practice effects that do not yield performance levels close to ceiling, or by having alternate forms. Given concerns about length, ease of administration, as well as subject burden for assessment batteries in clinical trials, a practical measure must be cost efficient and require a modest amount of time to administer.

Another critical dimension is that a measure in a clinical trial must be sensitive to change, often over a relatively short period of time. This criterion often creates significant measurement problems. First, many, if not most, important aspects of community functioning are slow to change, and changes in the short term are apt to be small. Second, changes in many aspects of functioning (eg, work) depend at least in part on the environment. For example, finding a job is substantially influenced by the local economy and job market, employer prejudice, and the person’s willingness to risk the loss of disability benefits when accepting a job. Similarly, maintaining a job is affected by the local economy and the employer’s business acumen, as well as the person’s job performance. A measure might accurately reflect significant changes in work capacity resulting from treatment, but there might not be an actual change in hours worked or income if the person cannot find or will not accept a job by the time of posttreatment and follow-up assessments. It was also noted that different types of interventions target different domains of functioning and/or attempt to alter different mediators and moderators of behavior. For example, a pharmacological agent that purported to alleviate cognitive impairment might be expected to have an impact on a broad range of community behaviors, while an agent that putatively influenced social drive (ie, the motivation to engage in social activities) might increase social interactions but not affect work or ADLs. Similarly, social skills training that targeted work behaviors would not be expected to improve safe sex behaviors. Consequently, no single measure would be suitable for all types of trials. In that regard, it will be important to determine the generality of any measure that is employed.

It is also important to consider that the universe of clinical trials and treatment research more broadly is not a monolithic enterprise. There are important differences in conceptualization, strategies, goals, and outcome targets that belie a single approach to assessment. The following sections highlight some of the issues involved from the perspective of rehabilitation trials and pharmacological trials.

Rehabilitation Trials. Psychiatric (or psychosocial) rehabilitation encompasses a range of social, vocational, educational, behavioral, psychological, cognitive, and pharmacologic interventions to enhance the community functioning and recovery of individuals with severe and persistent mental illness.⁷ These programs are diverse in nature but share a common set of features designed to foster improvement in performance of adult roles in the realms of work, education, socialization, independent living, and avoidance of rehospitalization.⁸

Cook and Jonikas⁹ reviewed outcome measures used in randomized trials and quasi-experimental evaluations of psychiatric rehabilitation and noted that they constitute a mix of behavioral outcome, self-report, and clinician assessments. For example, outcomes commonly used in the vocational domain include employment status, earnings, hourly salary, and job tenure. Residential rehabilitation outcomes typically include level of clinical supervision and support for “instrumental” activities of daily living (IADLs) provided on-site at the residence, nature of housing as agency-owned and controlled vs available on the open housing market, days of stable housing (vs homelessness), neighborhood safety, and quality of housing stock. Measures of quality of life include multidimensional standardized instruments designed specifically for use with this population such as the Quality of Life Interview,¹⁰ the Quality of Life Index,¹¹ and the Quality of Life Enjoyment and Satisfaction Questionnaire.¹²

There is broad consensus among investigators about the need for “real-world” measures of objective, behavioral changes in service recipients’ lives that result from rehabilitation services.^{13–16} Barton’s⁷ review of the literature in this field identifies a number of outcome constructs commonly measured in efficacy trials, including vocational outcomes, residential outcomes, income, hospital use, self-esteem, and life satisfaction. At the same time, assessment of psychiatric rehabilitation outcomes is complicated by a number of issues.¹⁷ First, there is the diversity of different program models and their intended outcomes. Second, even within particular models, different clients receive different service mixes corresponding to their personalized rehabilitation plan, a hallmark of psychiatric rehabilitation. Third, measures of efficacy must assess the patient’s use of skills in settings beyond the training environment, as well as changes in functional status and role performance that result from skill acquisition such as changes in labor force participation and residential status. In response to these challenges, tools have been designed to assess outcomes in multiple domains across different types of psychiatric rehabilitation programs. One example is the International Association of Psychosocial Rehabilitation Services Toolkit¹⁸ designed to monitor progress toward recovery across multiple dimensions, including employment, education, financial status, residential status, legal

system involvement, hospitalization, perceived quality of life, empowerment, and client satisfaction.

Drug Treatment Trials and Regulatory Interest in Functional Improvement. Registration trials (ie, trials that serve as the basis for gaining regulatory approval) for psychiatric drugs have traditionally focused on measures of symptomatic improvement even though a clear goal of treating psychiatric illness is to improve a patient's level of functioning (social, vocational, academic, etc). This is in part a practical matter. First, it seems unreasonable to expect much improvement in function in the typically short-term trials that support approval of most psychiatric drugs. Second, adding a measure of functional improvement along with a measure of symptom change as a coprimary outcome increases the risk of the trial failing to succeed in showing a drug effect.¹⁹ Finally, there is not an abundance of instruments for measuring functional improvement associated with drug treatment that can be conveniently utilized in drug treatment trials. Despite these obstacles, there has been some movement toward requiring in registration trails some measures that might tap into the functional domain better than the usual symptomatic change measures do.

Within the area of psychiatric drugs at the Food and Drug Administration (FDA), FDA has endorsed both cognitive impairment and negative symptoms of schizophrenia as legitimate targets for drug development, targets for which drugs may receive specific approval.²⁰⁻²² With cognitive impairment of schizophrenia, however, FDA has taken the position that a functional measure must be utilized as a coprimary outcome for such trials. There are 2 reasons for this requirement. First, there is concern that marginal changes in cognitive abilities might be demonstrated that are difficult to interpret clinically and may have no functional significance. Second, given FDA's position that drug trials for cognitive impairment in schizophrenia can and should be of longer duration than drug trials for many psychiatric disorders, there was the perception that there might therefore be an opportunity for measurable functional improvements to occur. FDA has not yet taken a position on whether or not such functional measures might be required in drug trials for negative symptoms of schizophrenia. A difficulty in both situations is that there has not been adequate development of convenient and effective instruments for measuring functional improvement in drug treatment trials for these indications.

Assessing Recovery-Related Issues. The consumer-driven recovery movement²³ has sensitized the scientific community about the importance of subjective appraisal of functioning and satisfaction with life. While scientific definitions of recovery emphasize the level of symptomatology, social role functioning, and ability to live independently, consumer definitions focus on feelings of

hope, empowerment, and fulfillment. These subjective feelings can be more important to consumers than residual symptoms and ability to work or live independently. While acknowledging the importance of recovery and consumer perceptions, the panel decided that it was distinct from objective measures of functioning. Moreover, self-perceptions of recovery are determined by so many diverse factors that it will often not be suitable as an outcome in clinical trials. Hence, the group agreed to exclude this domain from consideration.

Psychometric Issues

Broadly speaking, the field has placed more emphasis on face validity, consensual agreement among test developers, and content validity than on more objective and important issues in test development. While the fundamentals of test development are well known, some general background information is important to insure that a common standard is employed in measures of community functioning. It is well known that validity is a measure of how well a test is measuring what it is intended to measure. Validity may be indicated by an association between the test and other scales that unequivocally measure the desired attribute (and have themselves been "validated"). Thus, concurrent validity is the association of an individual's scores with another measure that is accepted as an agreed-upon criterion; predictive validity associates scores at one time with scores on a relevant criterion at another time. These validity measures are indicators of our confidence in the inferences we make from the test scores: the degree to which we are willing to assert that the 2 scores are the result of the same underlying attribute or process; that if the test were administered to different group in different circumstances, it would again measure the same attribute; or the degree to which the test will measure the same attribute when administered to the same individuals at a later date. This conception of validity is reasonable when a criterion gold standard exists (and can be reliably measured within acceptable tolerances of measurement error). However, the situation is much more difficult when developing a measure of an attribute or process where no such standard exists, as is the case with most mental health measures, including community functioning.

Validity As a Process. Most psychometricians define validation more broadly than as a property of a measure, assessed only in terms of its relation to an external, criterion-related outcome. Rather, validity-based decisions regarding functioning rely on inferences about performance demands, situational requirements, the skills, knowledge, and abilities of individuals, theories of dysfunction, and relationships among these factors. Validation is the process that tests the viability of those

inferences. Current standards hold that a strong validation program is one that builds and weighs evidence about each of the inferences that lead to a final decision.^{24,25} Validating individual inferences within the decision system (1) increases the probability that the ultimate outcome of the decision-making system will be accurate, defensible, and explainable, (2) allows the isolation of different decision-making components for systematic development of the entire system, and (3) builds on existing theory and empirically based knowledge, thereby enhancing our understanding of the viability of the decision system.

Moreover, viewing the validation of measures as a process of finding, building, and documenting relevant evidence prior to and over the lifetime of the measure's use—rather than solely as the result of a single crucial study—is both realistic and practical. Emphasizing that validity is a judgment regarding the appropriateness of inferences drawn from measures^{26,27} implies that the practical importance of valid inferences accrues, not to a single instance of behavior or to a single study, but to the verified consistency of measured behavior and of its consequences. This definition also encompasses Cronbach's^{28(p4)} assertion that validation is an argument linking “concepts, evidence, social and personal consequences, and values.”

Situational Validity. The points made above underscore that validity is tied to the purpose of the measurement and that validity of inferences made as the result of a measure is likely to be different for different individuals in different circumstances. Three related points warrant further consideration.

Issues With Longitudinal Studies. The course of illness and severity of symptoms often wax and wane over time; hence, we can experience difficulty in tracking severity of illness or treatment effects over time. Not only might the symptomatology itself be nonlinear or discontinuous but the measures of functioning could also display discontinuities. As a result, we might observe irregularities in both the illness (the predictor) and individual functioning (the criterion). From the perspective of Generalizability Theory, an implication for researchers and developers of measures is that, for longitudinal studies, an additional term needs to be added to the statistical model that represents time or the “occasion of measurement.” Thus, the sources of variance in observed functional performance include the symptom, the treatment, the person, the occasion of measurement, and the 2- and 3-way interactions. More generally, unless these irregularities and discontinuities are taken into account, they may exert major effects on the robustness of the tests and therefore on their validity.

Distribution Characteristics of Community Functioning. For a given individual, we can characterize

community functioning over time in a number of ways in addition to the standard indices of mean and standard deviation, including “Modal” or most typical day, consistency or day-to-day variance, and functional level range or “peakedness” of the distribution (kurtosis). For analyzing tests of functioning to determine treatment effects, these measures might be more sensitive or diagnostic than more standard measures—or, as we have been using the term, they might be more valid indicators of what we are trying to measure.

“Mini-Chaos” Theory.

When attempting to measure functioning “validly,” an individual's performance is a function of “person” characteristics and “demand” characteristics of the situation. Individuals have unique combinations of actual and perceived strengths and disabilities and have developed individualized coping mechanisms that are reflected in behavior. These coping mechanisms may be more or less successful in the particular measurement situations. Likewise, the individual's perception of the situation affects motivation, which in turn may not only affect behavior but may also interact with symptoms to distort functioning. Thus, at an individual level, a particular measure may or may not be valid in the sense of enabling one to make accurate inferences about a given person. The conclusion is as follows: think of the validity of measurements and scales in terms of their ability to add evidence that supports decisions, rather than as a property of the measurement itself.

Floor, Ceiling, and Dynamic Testing. The discussion above highlights some of the key conceptual issues that need to be considered in developing and validating a measure of functioning. Several other practical issues need to be addressed as well. One of the most important concerns “floor” and “ceiling” characteristics. Floor refers to the scale's ability to tap low levels of performance, while ceiling refers to adequate range to tap reasonably good performance. Inadequate floor or ceiling results in a constricted range and, frequently, a non-normal distribution of scores. While most people with schizophrenia are impaired in reference to premorbid capacity, there is a wide range of functioning both within people over time and between individuals. A cohort of long-term inpatients would be very different from a cohort of well-stabilized outpatients, and individuals who were recruited for a clinical trial in the midst of an acute episode would function very differently after they became stabilized. An effective instrument would be able to cover the expected range for a particular trial. This issue is complicated when the goal is to measure performance in the context of normal functioning, rather than when the question is restricted to persons with schizophrenia. Studies examining change in reference to a normal or community sample often have a high ceiling, which most patient subjects cannot approach.

Hence, their scores tend to be lumped at the low end of the distribution, and the overall data set is bimodal.

A creative approach to this problem involves the use of “dynamic testing.” Applications of modern test theory—computer-adaptive testing (CAT), item banking, and item response theory (IRT)—represent one direction toward individualizing outcomes assessment. IRT has been used for many years in educational testing to develop achievement tests and entrance exams that relate item difficulty in a test to a person’s ability to answer questions correctly. Conceptualized as probabilistic Guttman scaling, IRT relates characteristics of items (item parameters) and characteristics of individuals (latent traits) to the probability of a positive response. A variety of IRT models have been developed for dichotomous (ie, yes/no or true/false) and ordinal polytomous data (ie, excellent, very good, fair, poor). In each case, the probability of endorsing a particular response category can be represented graphically by an item (option) response function. Perhaps the most important application of IRT in the health field is CAT, a measurement approach in which the selection of items is tailored for each respondent.^{29,30} The development of a CAT requires several steps that are not required in the development of a traditional measure, including identification of “starting” and “stopping” rules. CAT’s most attractive advantage is its efficiency. Greater measurement precision can be achieved with fewer items. For example, a patient who is unable to walk would skip out of items pertaining to walking and answer only those items related to his or her baseline status. CAT might be an attractive approach for persons with severe mental illness by locating initially the area of social functioning for each individual in which to focus assessment, ie, work, independent living, social interaction. Item banks contain health status and quality of life items that are “cross-calibrated.”^{31,32} When used in dynamic testing, 2 individuals could be compared who answer different items because the items have all been put into a measurement system that provides scoring for all items on a similar metric. The National Institutes of Health (NIH) in the United States is sponsoring a large effort as part of the NIH ROADMAP for Medical Research to develop a Patient-Reported Outcomes Measurement Information System network. This trans-NIH initiative aims to use IRT, item banks, and CAT to measure patient-reported symptoms such as pain and fatigue and aspects of health-related quality of life across a wide variety of chronic diseases and conditions (<http://www.nihPROMIS.org/>).

Assessment Strategies

There has been considerable variability across instruments for measuring functional outcomes in regard to (1) who provides the information and (2) how it is

collected or rated. Four different approaches have been widely employed, each with numerous variations: self-reports, informant reports (eg, parent, caregiver), clinician ratings, and behavioral observation (observer rating). Behavioral observation has employed both observations of behavior in the natural environment and performance-based assessments in standardized, simulated environments. Regardless of the source of information, ratings in each domain have included both subjective appraisals of targeted behaviors and more objective reports of whether or how often a behavior occurs. Examples of subjective appraisals include ratings of how well a behavior is performed and how satisfied a person or caregiver is with the behavior. In some cases, ratings are linked to a time frame, such as the last week or month. In other cases, current functioning is compared with premorbid status or with a normative standard. More objective scales typically query the rater about quantitative dimensions, such as frequency of occurrence of some behavior, hours worked or wages earned on a job, and number of friends a person has.

In some cases, strategies have been selected for theoretical reasons. For example, behavioral or social learning models emphasize direct observation as the best (sometimes only) way to determine what an individual actually does in a particular situation. More frequently, cost and practicality are primary determinants. For example, multisite trials with large assessment batteries generally cannot afford to conduct behavioral observations. Informant ratings cannot be employed for subject samples in which many participants do not have informants, but are almost always used in studies of family interventions. Each assessment method and rating scheme has advantages and disadvantages, some of which are inherent to the approach and some of which are specific to the question being asked. The following sections provide an overview of key issues pertinent to each method.

Self-report

Self-report is perhaps the easiest to obtain and most widely used method for evaluating functioning. Some instruments address more than one area of dysfunction, while others address one primary domain. Advantages of using self-report instruments include their relative low cost, given that they require little staff time and do not require highly trained raters. Also, information regarding functioning is obtained directly from the individual in question. This is particularly important in regard to subjective experiences and self-appraisals such as quality of life, where the patients’ own perception of their functioning is central to the construct.^{33,34} Frequently, the person being assessed is the only one privy to the information being sought, as is the case with sexual behaviors and friendships. Some generic self-report instruments designed to provide a broad view of functioning of general

patient populations include the SF-36 health survey questionnaire³⁵ and the Quality of Well-being Scale.³⁶ Two instruments, the Self-reported Quality of Life Measure for People with Schizophrenia³⁷ and the Independent Living Skills Survey (C. J. Wallace, N. Kochanowicz, and J. Wallace, unpublished data, 1985) were developed specifically for patients with chronic mental illness and address common issues relevant to that population.

There is, however, some controversy regarding the validity of the self-report modality for assessing functioning in severely mentally ill individuals, particularly those with psychosis.³⁸⁻⁴¹ First, people with severe mental illness may not be accurate observers of their own behavior. Generally, the more specific the behavior being assessed and the more historical the report, the less likely that accurate information will be provided. Second, cognitive impairment may make it difficult for people to understand abstract questions and make objective self-appraisals or to accurately judge or rate performance. Third, self-report may be influenced by lack of insight, personal values or perceptions of others values, and situational events.³⁸ For example, grandiosity may distort appraisal of effectiveness. In other cases, the person may not be willing to acknowledge personal failings. Self-reports may also be distorted by current psychopathology (eg, thought disorder) and emotional functioning (eg, depressive symptoms).^{38,3-45} These problems may result in a measurement of functioning that does not objectively reflect the actual experience of the patient.

The complexity of this issue is illustrated in research on self-report measures of disability, neuropsychological performance (NP) and performance-based measures of functional skills.⁴⁶ Self-reports of disability on the part of outpatients with schizophrenia were correlated with self-reports of reduced quality of life, but neither of these self-reported variables was correlated with ratings of the clinical symptoms of schizophrenia, with NP, or with scores on a performance-based measure of functional capacity. Informant reports of patients' cognitive impairments were recently found to be more strongly convergent with both patients' neuropsychological scores and patients' functional skills performance than were patient self-reports of their own cognitive impairments.⁴⁷ While the finding that self-report in schizophrenia is uncorrelated with objective measures of functional outcomes argues against using these reports as exclusive measures of outcome, the fact that systematic relationships were detected between aspects of self-report on the part of people with schizophrenia suggests that it may be possible to identify the factors that influence the discrepancy between self-reports and objective measures. It should also be noted that this problem is not limited to people with schizophrenia. In any case, these data demonstrate that multiple sources of data should be collected both to provide cross-validation and to cover areas where only one person has access to the information.

Informant Reports

Collateral or caregiver reports have frequently been employed as a second, independent source of information on many measures in an attempt to accommodate to the uncertain reliability of self-report. Examples include the Social Functioning Scale,³ the Social-Adaptive Functioning Evaluation⁴⁸ (SAFE), and the Independent Living Skills Survey Informant version (C. J. Wallace, N. Kochanowicz, and J. Wallace, unpublished data, 1985). Other measures rely solely on the report of a collateral or caregiver (eg, Bayer Activities of Daily Living Scale⁴⁹). A particular limitation of collateral reports is that a substantial number of outpatients with schizophrenia, particularly those who are middle aged and older, are unable to identify a person who can report on their everyday functioning.⁵⁰ In addition, some studies,^{51,52} but not others,^{46,50} indicate that collateral reports of patient functioning may be unreliable (due to issues related to the rater's own cognitive, emotional, and psychiatric functioning), further compounding problems with sole reliance on this method of data collection. Informant reports may also be influenced by the specific relationships between caregiver and patient. For instance, in the Harvey *et al*⁴⁸ study cited above, the informants were trained mental health professionals providing care to long-stay psychiatric inpatients. Thus, the level of contact with these patients is regular and intensive. Given the difficulty in securing informants, frequently more effort is devoted to finding informants than to determining their ability to provide valid information. It cannot be assumed that anyone who has a relationship to a patient has access to the information being assessed or can accurately report it.

Clinician Ratings

Clinical ratings of patient behavior are a component of many symptom assessments (eg, Brief Psychiatric Rating Scale,⁵³ Geriatric Depression Scale⁵⁴). With regard to functioning, clinician evaluations of performance of skills needed to function in the community are commonly secured for patients in institutional settings (eg, Rehabilitation Evaluation of Hall and Baker,⁵⁵ SAFE⁴⁸). Ratings can be based on observation of the patients' behavior during an interview, interviews with informants, or on the patient's self-report. There is an important distinction between ratings based on clinician judgment after an interview or period of observation and ratings that primarily reflect patient self-reports to the clinician. In the latter case, the rater is primarily a scribe, and validity of the data depend primarily on the accuracy of patient reports. Instruments that rely primarily on clinician judgment are subject to interrater variability and within rater variability over time. These instruments require considerable rater training both before and during a project. It is also essential to document interrater

agreement (reliability) by having a sample of patients (usually 15–20%) rated by an independent rater. Consequently, these instruments can be very costly to employ. It is possible to devise instruments that achieve highly reliable interinformant and interrater scores.^{4,56} However, it should be noted that high interrater agreement is necessary but not sufficient to assure validity. Raters may agree on a score but may not have enough accurate information about actual community behavior to make valid judgments.

Behavioral Observation

Behavioral observation entails use of trained raters to score/rate behavior based on observation of the person performing the behavior, rather than based on retrospective reports and subjective evaluations of the person or an informant. It is generally assumed that this is the most valid approach because it circumvents several sources of inaccuracy or bias.^{57,58} Observations can be conducted in the natural environment or in simulated situations. Observation in naturalistic settings is considered to be the most robust approach because it allows the rater to evaluate whether the skills are actually implemented in the environment.⁵⁹ Such data can also provide a measuring stick by which to evaluate the validity of other measures of everyday functioning. Although this method of assessment is technically the ideal, it is rarely practical or cost efficient in outpatient environments.

Performance-based measures based on simulations or artificial environments have been developed as a practical alternative.^{45,60,61} This measurement modality requires that the participant perform a target skill in a contrived testing environment that may mirror the real world. For example, they may be asked to perform basic ADL tasks, such as combing their hair, to more complex tasks, such as engaging in social interactions, comparison shopping, or paying bills with checks. Patients may interact with an examiner in a role-play or simply demonstrate the implementation of a skill in question. The examiner may then score the patient with respect to his or her performance of that particular skill.

Two measures that have been used successfully in multisite trials are illustrative. The University of California, San Diego Performance-Based Skills Assessment (UPSA)⁶² examines a person's ability to perform IADLs in 5 areas as follows: (1) general organization, (2) finance, (3) social communication, (4) transportation, and (5) household chores. The UPSA involves role-play tasks similar in complexity to situations that a community-dwelling person is likely to encounter, including planning a trip to the beach, using a bus schedule, and balancing a checkbook. The UPSA yields both domain-specific scores and an overall score. It has been shown to be reliable, to have good discriminant validity, and to be only modestly correlated with severity of illness and neurocog-

nitive impairment. The Maryland Assessment of Social Competence⁶³ measures the person's ability to solve common problems in an interpersonal context (eg, interacting with a health care worker). It requires the person to engage in a series (usually 3–4) of 3-min conversations with a confederate. It was empirically developed and has proven to be reliable in several studies, to have good discriminant validity, and to be relatively independent of changes in symptomatology. Each scenario is coded on 3 dimensions that reflect different aspects of social skill: Verbal Skill (a content measure), Nonverbal Skill (a measure of paralinguistic style, eye contact, and gestures), and Overall Effectiveness (ability to maintain focus and achieve the goal of the scenario).

One attractive feature of performance-based assessments is that they rely on observation of patients' performance and are, thus, less subject to limitations of self-report in patients with schizophrenia. Another attractive feature of performance-based measures is that by focusing on real-life skills they provide information about patient functioning and deficits that may be targets for interventions. Despite these advantages, performance-based instruments are, themselves, not without limitation. External validity of these instruments cannot be automatically assumed. Reliance on contrived environments to conduct assessments may create assessment demands that differ from those in the real world. They also ignore the many environmental factors that may facilitate or inhibit participation in real-life situations. Community functioning depends on the person's ability or capacity to perform a particular behavior, willingness to perform the behavior, and environmental supports or constraints for performance. Behavioral assessments in simulated settings can provide excellent measures of the patient's capacity to perform behaviors under ideal circumstances, but they may not always accurately reflect the quality or rate or occurrence of the behavior in the community.^{63,64}

Conclusions About Assessment Modalities

In considering the strengths and weaknesses of the various assessment strategies, the workgroup agreed on a number of issues. In regard to self-report, there was agreement that (1) people with schizophrenia can report reliable and valid data under some circumstances, (2) their reports are important and may be the only source of information on some behaviors, and (3) information is lacking about the circumstances in which they can provide reliable and valid reports. A reliable and valid self-report measure that is sensitive to change would be of enormous value to the field and should be developed.

The group also agreed on several issues concerning behavioral observation. Specifically, (1) it is possible to collect reliable and generalizable samples of behavior in simulated settings, (2) such data can be important

and valid for some purposes, and (3) we need to know how, when, and for what purpose the data are collected in order to determine validity. It is a psychometric truism that no measure is valid for all people under all conditions: validity is a property of a test by person by situation interaction, not of the test per se. This principle is inherent to behavioral observation. It is especially important to understand the “demand characteristics” of a simulated interaction vis-à-vis the real-life environments the behavior sample is intended to represent (eg, level of stress, real vs simulated negative consequences). Differences in demand characteristics may not be highly important when the individual lacks the skill to perform effectively but could be a rate-limiting factor when the person has the requisite skills.

The group concluded that clinician/professional rater data: (1) are potentially helpful, (2) they can be reliable and valid for some behaviors under some circumstances, (3) clinician ratings need to be differentiated as raters/judges vs data collectors (transcribers), and (4) the sources of information clinicians use to make judgments and the manner of data collection are critical, yet most existing measures do not describe or standardize data collection strategy (eg, structured interview, collection of data from multiple sources). A particular concern in regard to clinician ratings is what other scales or measures the rater administers in a particular study. For example, it is not clear whether a rater can make unbiased ratings of community functioning after rating symptoms. Global ratings, such as the Global Assessment of Functioning or Global Assessment Scale, present special problems in regard to lack of specificity, potential for bias, and unknown source of information. In addition, it is essential to specify how well the rater knows the person being rated and whether she/he has access to other information needed to make ratings. In particular, a blinded rater who has not met the person before the assessment interview is very different from a clinician who has observed the person over time and/or been part of a clinical or research team in which the person is discussed. Failure to have adequate background information is a problem for ratings made by a blinded assessor. Conversely, having too much information about prior functioning creates the potential for bias and presents a considerable problem for assessing change in a clinical trial. Finally, an important issue that has not always been considered is cultural sensitivity of the rater and the extent to which either bias or lack of adequate information about cultural differences can affect the validity of ratings.

Several issues need to be considered in the use of informant ratings. Informants may be reliable and valid sources of information for some behaviors and may be the only source of some information. More information is needed on which behaviors and under what circumstances informant ratings are likely to be valid. Little is known about person characteristics (ie, individual dif-

ferences) that affect reliability/validity and what kinds of information different informants can provide. The literature rarely employs objective criteria to determine the amount of face-to-face contact and what kind of relationships informants must have with the patient in order to be considered as potentially valid sources of information. Family informants may be unduly biased by history and may not be sensitive to change; conversely, they may also be the best source of information because they know the person best. It is generally recommended that concrete data be requested (how often? how recently?), rather than qualitative ratings (how well?). In addition, the presence of an informant may not be a random factor and may need to be taken into account in data analysis. Patients with and without an informant may differ on a variety of important characteristics such as homelessness, severity of illness, age, etc.

Models of Functional Outcome in Schizophrenia

Considerable discussion during the workshop was devoted to question 10: What are the current theoretical models for relationships among cognition, symptoms, motivation, capacity, and performance? There was agreement on the importance of a theoretical model for developing and selecting assessment measures, and several examples were mentioned. Importantly, there is increasing focus on formal statistical testing of these models, and examples of statistical approaches are described below. The process of model generation is highly iterative. Models can be first proposed based on a theory and a pattern of bivariate correlations among a group of key variables. Model generation that is based on observed patterns of bivariate relationships can be highly creative but can also be difficult to disprove. In contrast, formal model testing with statistical approaches either supports the model or it forces one to give up certain components, connections, and constructs in the models that do not fit the data. Based on such testing, models can be modified and reevaluated.

In addition to enabling one to reject components of models, another advantage of the statistical testing is that it enables one to test “mechanisms” of proposed effects. Given the highly complex interplay of factors that determine functional outcome, it is assumed that many of the effects occur through mediating variables. For example, it is widely accepted that neurocognition has a relationship with functional outcome, but these relationships may be influenced by mediators that act between neurocognition and functional outcome. Hence, a full explanation of the effects requires mediators to be included in the model. Of course, the challenge for investigators is to know which likely mediators to include in the data collection.

A third advantage of constructing and formally testing models is they can suggest specific therapeutic targets. If

a component (eg, social cognition, metacognition, or functional capacity) is a key mediator of functional outcome, one starts to consider it as a possible target for intervention. The question is particularly relevant if components are thought to be closer to the outcome of interest. For example, if a mediator lies closer to community outcome or vocational success than basic neurocognition processes, it may be a promising target. The goal at this stage is not to decide that one level of intervention is better than another, rather it is to map out the key connections to help interpret treatment effects and guide new interventions. The following sections describe relevant statistical issues and representative models.

Statistical Issues

There are important statistical issues that need to be addressed in evaluating theoretical models and developing effective measures. Hypothesized models of functioning can be empirically evaluated using path analysis, confirmatory factor analyses, or structural equation modeling. Path analysis, initially developed by Wright,^{65,66} is used to examine the strength of direct and indirect linear relationships among observed variables. Path analysis, as with multiple linear regression analysis, can examine multiple independent variables, yet only the former can include multiple dependent variables. (The dependent variables, also called the endogenous variables, in a path model are those variables that are recipients of path arrowheads.) Confirmatory factor analysis, on the other hand, is used to estimate the number of constructs, or latent factors, that are measured by a set of observed variables and which of those variables is most highly related to each factor. Consider, eg, measures of functioning at home, in the workplace, and interpersonal relations with friends and family. Confirmatory factor analyses could be used to test whether these measure 4 distinct constructs or perhaps all 4 are observed measures of a single construct, namely “functioning.”

A structural equation model essentially combines the path and factor models in that it examines structural linear relations among latent variables. Both path models and structural equation models are referred to as “causal” models, yet they simply examine the strength of associations among variables. These statistical models cannot be used to infer causality, particularly with observational data. Confounding variables, both measured and unmeasured, could account for the hypothesized relations among the variables. Hypothesized path models, confirmatory factor models, or structural equation models can be tested using maximum likelihood estimation as implemented by several different software packages including LISREL,⁶⁷ EQS Software,^{68,69} AMOS⁷⁰ (SPSS), and CALIS.⁷¹

The consistency between a hypothesized model and the observed data is evaluated using various goodness-

of-fit indices. The indices generally evaluate how closely the variances and covariances that are estimated with parameters from the specified model correspond with the variances and covariances that are observed in the sample. There are numerous goodness-of-fit indices that are described in detail by Bollen and Long,⁷² including chi-square goodness-of-fit, ratio of chi-square goodness-of-fit to degrees of freedom, adjusted goodness-of-fit index, standardized root mean square residual, and the Akaike information criterion.

None of these indices can be used to determine if the hypothesized model is the best that can be posited. However, many of them can be used to compare the plausibility of one model relative to another. For instance, one can determine if the addition of one or more parameters (eg, paths) significantly improves the fit of the model to the data. These are referred to as nested models (ie, that with fewer parameters is nested in the other model). In contrast, the fit of 2 models cannot be compared if they each contain at least one unique path. Furthermore, 2 models can be compared only if the observations included in each model are identical; therefore, the model builder must be aware of the problems with missing data between the 2 models that are being compared.

An additional criterion to consider in evaluating a model is the proportion of variance (R^2) in each dependent variable (ie, the endogenous variable) that is explained by the variables included in the model. The explained variance includes that accounted for by direct and indirect (ie, mediated) effects of the other variables. A very good fitting model might explain only a trivial amount of variance in the primary outcome. Other important aspects of the model include the magnitude and direction of each parameter estimate and the statistical significance of those estimates. Finally, none of these indices can offset the problems faced if a critical variable was not assessed in a study in the first place.

We present brief descriptions of 4 illustrative models. All the models share certain key features but also have distinct elements. Importantly, all the models have been, or can be, formally tested by the statistical methods described above.

*Model 1: Heaton et al.*⁷³ An illustrative model proposed by Heaton et al⁷³ is shown in figure 1. It was conceived as a potential guide for studies of pharmacologic treatments targeting cognitive impairments associated with schizophrenia. However, the model also should apply to other disorders that have neurocognitive sequelae (eg, human immunodeficiency virus infection). It highlights the types of outcome measures that would be relevant in studies of treatments and also reflects moderator variables that may attenuate effects of cognitively beneficial treatments on certain outcomes.

The 3 types of outcomes depicted in figure 1 are changes in cognitive impairment, functional “capacity,”

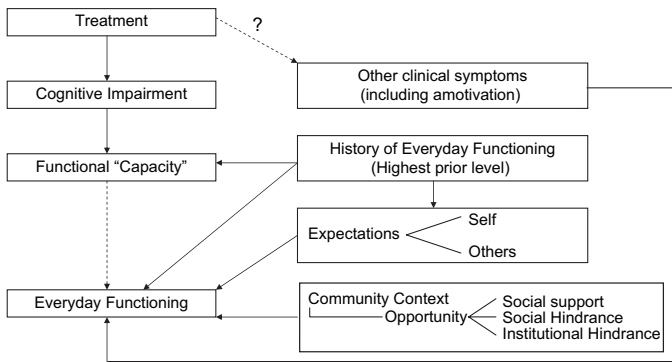


Fig. 1. Conceptual Model for Assessing Outcomes of Cognitively Enhancing Pharmacologic Treatments for Mental Disorders.

and cognitively related aspects of everyday functioning. Especially for pharmacologic treatments designed to enhance neurocognitive functioning, the most direct and straightforward effects should be seen on a neuropsychological test battery that covers domains of cognition affected by the disorder in question. Conceivably, a pharmacologic treatment also could have beneficial effects on more distal elements of the model (everyday functioning) due to influences on noncognitive aspects of the disorder (eg, amotivation, positive or negative symptoms of schizophrenia). However, especially in studies of drugs thought to operate on neural systems that support relevant aspects of cognition, the first level of outcome assessment will be an appropriately designed neuropsychological test battery.

Economic or life quality benefits from reducing disease-related cognitive impairment are likely to come from improved ability to engage in important everyday tasks and activities. The latter is called “functional capacity” and is most sensitively measured by standardized, performance-based assessments of everyday activities. These are examples of behavioral observations in simulated setting that were described in the previous section. Currently, these include measures of financial management, medication management, shopping, cooking, vocational abilities, driving, use of public transportation, planning recreational outings, and communication skills.^{74–80} These measures of functional capacity have much greater face validity for relevance to everyday functioning than neurocognitive measures. Also, measures of functional capacity have important advantages over observations of real-world functioning: performances on relevant everyday tasks can be efficiently measured, effects of some noncognitive symptoms (eg, amotivation) may be minimized, and constancy of task requirements and environmental context of the assessment provides increased reliability and greater comparability of scores across examinees.

As with real-world outcomes, the relevance of specific functional capacity measures will vary across individuals, partly as a result of differences in past experience and

level of functioning. Accordingly, in figure 1 the “History” box on the right is shown as directly influencing both functional capacity and real-world outcomes (everyday functioning) and possibly indirectly influencing the latter outcomes through history-related levels of self-expectations and expectations of others. This reasoning suggests that, regardless of whether cognition or functional capacity have changed as a result of treatment, the likelihood of any near-term change in everyday functioning will be small unless the person has a history of higher real-world functioning in the past.

Figure 1 also proposes that factors other than cognition/capacity and history will influence the real-world functioning of the patient. First, other (noncognitive) clinical symptoms may cause a “disconnect” between what a person is cognitively capable of doing (functional capacity) and what he/she actually does outside of the assessment laboratory or treatment facility. For example, positive symptoms of schizophrenia (hallucinations, delusions) have been shown to have very little relationship to cognition but can prevent gainful employment and independent living. Amotivational states associated with schizophrenia and other disorders may also affect whether a person independently performs tasks he/she clearly is capable of doing.

A final set of factors that may strongly influence a patient’s everyday functioning, independently of his/her ability to perform the relevant tasks, is subsumed within the figure box labeled “Community Context.” Patients’ real-world environments are variable and frequently are not conducive to optimal expression of their skills and abilities. For example, a nursing home or board and care facility may be organized to dispense medications and perform other IADLs (laundry, shopping, cooking, etc) for all patients. There may even be safety-related rules against independent performance of certain tasks (eg, medication management, cooking).

Clinical trials of cognitively enhancing treatments often may not identify and address the multiple noncognitive factors that this model predicts will affect everyday functioning of patients with mental disorders. Instead, such studies are unlikely to result in immediate, substantial changes in everyday functioning, particularly in patients with chronic, severe mental disorders, and long-standing disabilities. It would be reasonable for such studies to assess all 3 types of outcomes (cognition, capacity, everyday functioning) but that cognition and/or functional capacity be considered primary. Once initial studies have demonstrated beneficial effects of drugs on cognition, more elaborate (and expensive) studies of combined drug and psychosocial treatments should become a high priority for the NIMH and the mental health field.

*Model 2: Brekke et al.*⁸¹ Brekke et al have proposed a model of functional outcome in schizophrenia that

was tested with path analysis. The model has 5 main components: neurocognition, social cognition, social competence, social support, and functional outcome.

Neurocognition. The neurocognition component of the model included measures of episodic memory, immediate working verbal memory, executive functioning, and sustained attention. Each of these domains is considered to be related to functional outcome in schizophrenia.^{81–83} In this model, neurocognition was initially proposed to have both direct and indirect (mediated) effects on functional outcome, but in fact, the effects were almost entirely mediated by other variables.

Social Cognition. In this model, social cognition is differentiated from basic (nonsocial) cognition. Social cognition refers generally to the interface and interdependence of cognitive and social/interpersonal processes.⁸⁴ In this model, the measure of social cognition was affect perception (the ability to identify emotional expression in photos of faces and audiotapes). Affect perception had both direct and indirect influence on functional outcome in the model. It acted as a mediator between neurocognition and functional outcome. Its effects on functional outcome, in turn, were partially mediated by social competence and social support.

Social Competence. Social competence skills include sets of abilities, such as verbal and nonverbal behavior, that are needed for effective interpersonal performance.⁸⁵ The results suggested that competence in social skills (assessed during an interview) mediated the relationship between affect perception and functional outcome.

Social Support. Social support includes caring and sustenance provided by the social environment, as well as the emotional and material support that people obtain from their social relationships.⁸⁶ It was proposed that perceived social support would act as a mediator between affect perception and functional outcome: that is, if patients have problems in social cue perception, they may have difficulty in perceiving or accessing social support. This mediating relationship was supported in the data.

Functional Outcome. In this model, the focus was on the functional outcomes of work, independent living, and social interactions of individuals living in the community.^{87,88} The multiple domains of outcome in schizophrenia appear to be linked, but separate, and they could have distinct causal pathways. Using multiple domains of functional outcome provided a way to examine the differential or generalized impact of predictors in the model on specific outcome domains and on global outcome.

Despite some variations, the results of the path analyses were fairly consistent across the 3 outcome domains

(work, independent living, and social) and global outcome. Importantly, the results for the baseline outcome model (in which all the measures were collected at the same time) were very similar for the 1-year follow-up model (in which the outcome measures were collected 1 year later). The models had explanatory effect sizes in the medium range for both the cross-sectional and 12-month outcomes. Overall, there was consistent evidence that these models across domains of functional outcome in schizophrenia had a strong fit to the data. The results of this model strongly suggest the importance of assessing the key mediating steps (including social cognition and observed social competence) for intervention studies in which functional outcome is the goal. The rationale, as described in the section above, is that one does not know ahead of time whether an intervention will affect a distal variable (functional outcome) as opposed to a mediating outcome (eg, social cognition or social competence).

*Model 3: Harvey et al.*⁷⁴ Harvey and colleagues developed a model based on the competence/performance distinction in terms of functional skills: the separation of what one can do from what one actually does and discrimination of influences that may impact on either or both of these domains. Harvey characterizes the 4 main components of the model: (1) NP, (2) performance on measures of adaptive skills competence (functional capacity as described above), (3) symptom variables, and (4) additional influences. These additional influences include, but are not limited to, opportunities for real-world performance and various incentives and disincentives to perform real-world skills. According to this model (and shared with the model of Heaton et al⁷³), symptoms may impact at the level of NP performance, functional capacity, and real-world performance.

Preliminary tests of this model were conducted with confirmatory path modeling.⁷⁴ The results suggest that NP performance exerts its influence on real-world outcome in a manner completely mediated by functional capacity and that symptoms do not have a direct impact on either NP performance or functional capacity measures. In addition, both depression and negative, but not positive, symptoms were related to real-world outcomes but unrelated to NP and functional capacity measures.

The psychometric qualities (eg, difficulty level, reliability, and variability) of NP tests exert a substantial influence on the likelihood of finding group differences in performance (ie, discriminating power) and also for the reliable detection of correlational relationships between NP scores and other variables of interest, such as measures of symptomatic variables and functional skills competence and performance.⁸⁹ Similarly, measures of functional capacity are also performance-based tests, and the scores from these measures have distribution characteristics that could influence sensitivity to

differential deficits and differences in correlations. However, due to the substantial disability in everyday living skills shown by people with schizophrenia, distributions of scores on functional capacity measures may have different distribution characteristics across samples than those seen with NP performance. Many functional skills domains in which people with schizophrenia show impairments may have substantial ceiling effects when applied to healthy controls. For example, for the functional capacity instrument that was used to test the model in the study mentioned above (the UPSA,⁷⁸ the average healthy comparison subject received a score of 92.6 (SD = 5.5) on the 100-point overall scale and the average patient with schizophrenia received a mean score of 58.8 (SD = 27.1) on the same scale. Thus, there are substantial differences in both performance level and variance estimates across the 2 samples.

These results suggest that tests of functional skills that are sensitive to the functional impairments of people with schizophrenia (and show normal distributions within this population) may not be as amenable to comparisons using normative standards from the general population to the same extent as NP tests. It will be important to consider whether measures of functional capacity should be evaluated with comparisons to healthy controls or whether the development of normative standards for performance on these measures should be limited to comparisons with people with schizophrenia (or other neuropsychiatric conditions that induce disability in everyday living skills).

Model 4: Wykes et al.^{90,91} Wykes and colleagues in the United Kingdom have taken an approach to model building that emphasizes not only the cognitive targets themselves but also the way in which basic cognitive abilities might be transferred to real-world functioning. Data from intervention studies (as well as studies described above) have not been universally supportive of a direct link between cognition and action. Substantial evidence from a number of studies⁹⁰⁻⁹³ suggests that cognitive remediation acts as a mediator of the effects of cognitive change on outcome. The key components of this model include processes involved with the transfer to community functioning (including metacognition) and cognitive schema.

Independent factors that affect the transfer of cognitive performance improvement into action include opportunity, prior skill level, social support, and motivation. Although these are important variables that have been implicated in learning skills and performance in a number of different areas, there may also be specific theoretical links between cognition and action that might be essential in determining whether there is a transfer of cognitive improvements into functional domains.

For example, actions are governed by cognitive schemas which are generic knowledge structures or templates

that are stored within long-term memory and are the means by which mental representations are organized.^{94,95} When the schema is highly specified, it primes a complete set of actions so that additional internal selection is not required. The actions produced by these highly specified schemas could be called routine because they proceed in the same manner each time they are carried out and do not depend on specific circumstances. Controlled processes such as executive functioning or explicit memory retrieval may be required to implement the set of actions in full, but the selection of actions and their temporal sequence is specified by the cognitive schema. In this model, improvements in specific cognitive processes will improve the efficiency with which the action is carried. So, in situations in which functioning involves routine inflexible actions, the action would be improved if basic cognitive performance improved.

However, most important actions are not fully specified by an existing cognitive schema and rely upon the ability of the actor to select actions appropriately from a range of options. This selection ability is known as metacognition.⁹⁶ Metacognitive knowledge contains information about strategy variables (eg, cognitive variables), task variables (eg, nature of the task demands), and person variables (eg, strengths and limitations, motivation, etc). Metacognitive regulation or experience allows the person to generate and realize intentions, set goals, generate and implement strategies, problem-solve, etc.

Metacognition forms part of transfer ability, which derives its empirical base from work in education and training. Transfer depends on the presence of existing knowledge and skills but actually describes their implementation and application across different settings and situations. This transfer ability depends not only upon cognition, including metacognitive skills and cognitive resources, but also upon internal (eg, motivation, mood) and external (eg, the nature of the task, the environment) noncognitive factors. Actions in this model refer to behavioral performance on neuropsychological tasks, measures of functional capacity, as well as community functioning. This model emphasizes the cognitive schema in defining functional outcome rather than specific basic cognitive abilities and is therefore complementary to the models described above. According to this model, these transfer, schema, and metacognitive constructs are considered dynamic processes, whereas the basic cognitive abilities would be considered static processes. Thus, the model by Wykes et al suggests additional types of constructs that are likely to be relevant for assessment in studies of intervention, including measures of metacognition and cognitive schema.

Clinically Significant Change

A final issue that was addressed by the workgroup concerned “clinically significant change.” Outcome data

from clinical trials has been typically expressed in terms of descriptive statistics (means, standard deviations) and significance levels. More recently, emphasis has shifted away from a sole focus on statistical significance towards the added consideration of effect sizes and confidence intervals. While that transition makes considerable statistical sense and also facilitates interpretation of the results of trials, it still does not answer the practical question of whether or not the effect made a real difference in real-world outcomes to trial participants and their families. Consumers, family members, and governmental agencies want to know if the person can work, live independently, marry and raise children, and function at age-expected levels after treatment, not that there was an increase in X points on a scale. Unfortunately, no extant treatments can reliably produce such clear-cut life changes, and there are no such treatments on the immediate horizon. That being the case, the intermediate goal is to detect clinically significant change. However, as indicated above, there is no gold standard on which to make that judgment; consequently, any determination is subject to debate.

Several interrelated issues were discussed. One concerns whose perspective to take in judging outcome. As indicated above, functioning can be assessed from the perspective of the patient, from the perspective of significant others, from the perspective of trained independent raters, and from the perspective of clinicians. Which perspective is most valid? The answer depends (at least) on (1) the behavior to be assessed, (2) the reliability of the individual reporter, and (3) the purpose of the assessment. In many cases, the patient is the sole source of information about a behavior, but he or she may not be a reliable reporter. Significant others often know the patient best, but their judgments may be biased by historical experience. Trained raters are putatively the most objective, but their sources of information are generally limited by observations during interviews and the reports of others. Validity is only one factor to be considered in determining which sources take priority. Another perspective comes from the consumer movement, which places a premium on what the consumer thinks and feels about his or her functioning, regardless of the objective levels. A related question concerns the need for normative standards. Should that be the gold standard for comparison or does it present another set of problems: what is the normal cohort with which to compare performance? Is "normal" a statistical attribute (eg, within a standard deviation of a population mean) or a qualitative judgment? Is a more appropriate standard premorbid functioning? If so, how does one determine premorbid status? If the disease has an insidious onset that begins in childhood, premorbid adult functioning may be a nonsequitor and reflects a level of performance less than what was anticipated in childhood.

This issue is easily seen in the context of employment. Consider a person who had their first episode in medical

school and had an insidious course of illness such that she/he was unable to return to school or hold any job for the next 10 years. If a new treatment enabled the person to begin working in a part-time job in a mail room, would that be considered clinically significant? It certainly reflects meaningful change from pre-treatment status, but just as clearly, it does not reflect premorbid functioning. This type of change might be seen as highly meaningful from the perspective of family and clinician but might still be regarded as a failure by the person if his or her reference point was medical school. The workgroup did not reach any conclusions about this complex set of issues but underscored that resolving these issues must be considered in parallel to the development of valid measures.

Future Directions

Developing new assessment methods and testing models will require a considerable effort by the field. In addition to convening the workshop that generated this article, NIMH established a new Functional Assessment program area located within the Health and Behavior Branch of the Division of AIDS and Health and Behavior Research to increase effort in this area. Examples of research relevant to this new program area are listed below.

- Studies tying the evaluation of new drugs and psychosocial interventions for mental illness to functional effects and linking those effects to contextually defined disabilities.
- Studies examining functional assessment in mental disorders in relation to occupational deficits or problems in daily living.
- Studies examining the relationship between functional performance and functional capacity among people with various mental illnesses.
- Studies linking traditional assessment of psychiatric symptoms with their functional consequences.
- Studies to develop and validate surrogate or laboratory measures of functional capacity that can be used to predict real-world functioning cross-sectionally and longitudinally.
- Studies to create and validate new measures of functional capacity that can be used as endpoints or copriary endpoints in pharmacological and psychosocial intervention trials.
- Studies of various theory-based methods of functional assessment in various under-served or high-risk/special-need populations across the life course.
- Studies of how factors underlying functional assessment in one domain, eg, eligibility for Social Security benefits, might inform functional assessment aimed at mental disorders.
- Studies of factors underlying provider strategies and behaviors that enhance functional assessment.

- Studies clarifying the characteristics of study participants that may contribute to the need for varying measures of functioning, such as age, gender, ethnicity, level of education, occupation, marital status, parental status, socioeconomic status, sexual orientation, religion, personality traits, personal and/or cultural beliefs about health, the nature of their disorder or their treatments, and the presence of comorbid mental illness.
- Studies of the health provider/client alliance and communication patterns and how these processes affect functional assessment.

It should be noted that in addressing each of these issues, it will be important to consider cultural factors that are germane to assessment. People with schizophrenia come from diverse cultural backgrounds, and their response to assessment, the burden illness imposes on the person and family, and the aspirations of the person and family all reflect the local and family culture. Assessment strategies and interpretation of resultant data must be sensitive to these differences. Ultimately, better understanding of functioning in individuals with schizophrenia and better methods of measuring functioning will help NIMH achieve its mission to reduce the burden of mental illness.

Summary and Conclusions

This article summarized the discussion in a workshop on functional outcome sponsored by NIMH. The participants were selected to represent diverse perspectives and areas of expertise. The far-ranging discussion highlighted a variety of important issues that the field needs to address. A variety of measurement approaches have been employed, including self-report, informant ratings, clinician ratings, and behavioral observation. Each approach has advantages and disadvantages. The workgroup concluded that all 4 approaches should be employed and that more work needs to be done to determine the circumstances in which each approach will and will not provide valid and useful information. Questions were also raised about how to combine the measures and how priority among them should be determined when results are not consistent. Overall, the field has not paid sufficient attention to psychometric issues in developing measures or selecting measures for use in particular studies. Some key psychometric issues were reviewed, and it was emphasized that validation is a process and that validity is a property of person–test–situation interactions, not of tests per se. There was considerable agreement on the importance of theory for developing and selecting instruments, although there was not complete agreement on any particular theoretical model. Several examples were provided to illustrate different ways of thinking about the role of cognition in functioning. Paralleling the emphasis on psychometrics, there was also consideration of the

complexity of validation of models, and some useful statistical approaches were described. Finally, the group discussed the concept of clinically significant change. It was simultaneously agreed that this is a critically important issue and one that presents considerable challenges. This topic needs to be a priority in future work.

Acknowledgment

Preparation of the manuscript and the workshop on which it was based were supported by the NIMH.

References

1. Eisen SV, Normand S-L, Belanger AJ, Spiro A, Esch D. The Revised Behavior and Symptom Scale (BASIS-R): reliability and validity. *Med Care*. 2004;42:1230–1241.
2. Morosini PL, Magliano L, Brambilla L, Uglioni S, Pioli R. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatr Scand*. 2000;101:323–329.
3. Birchwood M, Smith J, Cochrane R, Wetton S, Copestake S. The Social Functioning Scale: the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenia patients. *Br J Psychiatry*. 1990;157:853–859.
4. Wykes T, Sturt E. The measurement of social behaviour in psychiatric patients: an assessment of the reliability and validity of the SBS schedule. *Br J Psychiatry*. 1986;148:1–11.
5. Jaeger JJ, Czobor P, Berns S. Multidimensional scale of independent functioning: a new instrument for measuring functional disability in psychiatric populations. *Schizophr Bull*. 2003;29:153–168.
6. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*. 2002;59:877–883.
7. Barton R. Psychosocial rehabilitation service in community support systems: a review of outcomes and policy recommendations. *Psychiatr Serv*. 1999;50:525–534.
8. Cook JA, Hoffschmidt S. Comprehensive models of psychosocial rehabilitation. In: Flexer RW, Solomon P, eds. *Psychiatric Rehabilitation in Practice*. Boston, Mass: Butterworth-Heinemann; 1993:81–97.
9. Cook JA, Jonikas JA. Outcomes of psychiatric rehabilitation service delivery. *New Dir Ment Health Serv*. 1996;71:33–48.
10. Lehman AF. A quality of life interview for the chronically mentally ill. *Eval Program Plann*. 1988;11:51–62.
11. Ferrans CE, Powers MJ. Psychometric assessment of the Quality of Life Index. *Res Nurs Health*. 1992;15:29–38.
12. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29:321–326.
13. Anthony WA, MacDonald-Wilson KL. Functional assessment in psychiatric rehabilitation. In: Bolton B, ed. *Handbook of Measurement and Evaluation in Rehabilitation*. 3rd ed. Baltimore, MD: Paul H. Brookes; 2001:423–448.
14. Bond GR. Vocational rehabilitation. In: Liberman RP, ed. *Handbook of Psychiatric Rehabilitation*. Needham Heights, Mass: Allyn and Bacon; 1992:244–275.
15. Newman FL, Ciarlo JA. Criteria for selecting psychological instruments for progress and outcome assessment. In:

- Maruish M, ed. *Use of Psychological Testing for Treatment Planning and Outcome Assessment*. Hillsdale, NJ: Lawrence Erlbaum and Associates; 1994:60–69.
16. Blankertz L, Cook JA. Choosing and using outcome measures. *Psychiatr Rehabil J*. 1998;22:167–174.
 17. Cook JA. Research on psychosocial rehabilitation services for persons with psychiatric disabilities. *Psychother Rehabil Res Bull*. 1995;4:5–11.
 18. Arns P, Rogers ES, Cook J, Mowbray C. The IAPSRs toolkit: development, utility, and relation to other performance systems. *Psychiatr Rehabil J*. 2001;25:43–52.
 19. Leon AC. Multiplicity-adjusted sample size requirements: a strategy to maintain statistical power when using the Bonferroni adjustment. *J Clin Psychiatry*. 2004;65:1511–1514.
 20. Laughren T, Levin R. Food and drug administration perspective on negative symptoms in schizophrenia as a target for a drug treatment claim. *Schizophr Bull*. 2006;32:220–222.
 21. Buchanan RW, Davis M, Goff D, et al. A summary of the FDA-NIMH-MATRICES workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophr Bull*. 2005;31:5–19.
 22. Kirkpatrick B, Fenton WS, Carpenter WT Jr, Marder SR. The NIMH-MATRICES consensus statement on negative symptoms. *Schizophr Bull*. 2006;32:214–219.
 23. Bellack AS. Scientific and consumer models of recovery in schizophrenia: concordance, contrasts, and implications. *Schizophr Bull*. 2006;32:432–442.
 24. Benson J. Developing a strong program of construct validation: a test anxiety example. *Educ Meas Issues Prac*. 1998;17:10–22.
 25. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281–302.
 26. Landy FJ. Stamp collecting versus science. *Am Psychol*. 1986;41:1183–1192.
 27. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: Macmillan Publishing Company; 1989:13–103.
 28. Cronbach LJ. *Essentials of Psychological Testing*. 3rd ed. New York, NY: Harper & Row; 1988.
 29. Ware JE Jr, Kosinski M, Bjorner JB, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res*. 2003;12:935–952.
 30. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: time to let the cat out of the bag? *Health Serv Res*. 2005;40(pt 2):1694–1711.
 31. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care*. 2000;38(suppl):1143–1159.
 32. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res*. 2003;12:485–501.
 33. Meltzer HY. Dimensions of outcome with clozapine. *Br J Psychiatry*. 1992;160:46–53.
 34. Awad AG, Voruganti LNP, Heslegrave RJ. A conceptual model of quality of life in schizophrenia: description and preliminary clinical validation. *Qual Life Res*. 1997;6:21–26.
 35. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–483.
 36. Kaplan RM, Bush JW, Berry CC. The reliability, stability, and generalizability of a health status index. In: *American Statistical Association, Proceedings of Social Status Section*. 1978;704–709.
 37. Wilkinson G, Hedson B, Wild D, et al. Self-report quality of life measure for people with schizophrenia: the SQLS. *Br J Psychiatry*. 2000;177:42–46.
 38. Atkinson M, Zibin S, Chuang H. Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. *Am J Psychiatry*. 1997;154:99–105.
 39. Arfken C. Self-reported life satisfaction (Letter). *Am J Psychiatry*. 1997;154:1478.
 40. Rohland BM, Langbehn DR. Self-reported life satisfaction. *Am J Psychiatry*. 1997;154:1478–1479.
 41. Loew F, Rapin H. The paradoxes of quality of life and its phenomenological approach. *J Palliat Care*. 1994;10:37–41.
 42. Barker DA, Shergill SS, Higginson I, Orrell MW. Patients' views towards care received from psychiatrists. *Br J Psychiatry*. 1996;168:641–646.
 43. Morgado A, Smith M, Lecrubier Y, Widlocher D. Depressed subjects unwittingly overreport poor social adjustment which they reappraise when recovered. *J Nerv Ment Dis*. 1991;179:614–619.
 44. Serban G, Gidynski CB. Relationship between cognitive defect, affect response and community adjustment in chronic schizophrenics. *Br J Psychiatry*. 1979;134:602–608.
 45. Jenkins CD. Assessment of outcomes of health intervention. *Soc Sci Med*. 1992;35:367–375.
 46. McKibbin C, Patterson TL, Jeste DV. Assessing disability in older patients with schizophrenia: results from the WHO-DAS-II. *J Nerv Ment Dis*. 2004;192:405–413.
 47. Keefe RSE, Poe M, Walker TM, Kang JW, Harvey PD. The Schizophrenia Cognition Rating Scale SCoRS: interview based assessment and its relationship to cognition, real world functioning and functional capacity. *Am J Psychiatry*. 2006;163:426–432.
 48. Harvey PD, Davidson M, Mueser KT, Parrella M, White L, Powchik P. Social-Adaptive Functioning Evaluation (SAFE): a rating scale for geriatric psychiatric patients. *Schizophr Bull*. 1997;23:131–145.
 49. Hindmarch I, Lehfeld H, de Jongh P, Erzigkeit H. The Bayer Activities of Daily Living Scale (B-ADL). *Dement Geriatr Cogn Disord*. 1998;9(suppl 2):20–26.
 50. Patterson TL, Semple SJ, Shaw WS, Grant I, Jeste DV. Researching the caregiver: family members who care for older psychotic patients. *Psychiatr Ann*. 1996;26:772–784.
 51. Beck C, Heacock P, Mercer S, Walton C, Shook J. Dressing for success: promoting independence among cognitively impaired elderly. *J Psychosoc Nurs Ment Health Serv*. 1991;29:30–35.
 52. Dickerson FB. Assessing clinical outcomes: the community functioning of persons with serious mental illness. *Psychiatr Serv*. 1997;48:897–902.
 53. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep*. 1962;10:799–812.
 54. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res*. 1983;17:37–49.
 55. Baker R, Hall JN. REHAB: a new assessment instrument for chronic psychiatric patients. *Schizophr Bull*. 1988;14:97–111.
 56. Harvey CA, Curson DA, Pantelis C, Taylor J, Barnes TRE. Four behavioural syndromes of schizophrenia. *Br J Psychiatry*. 1996;168:562–570.

57. Bellack AS, Hersen M, Turner SM. Role-play tests for assessing social skills: are they valid? *Behav Ther.* 1978;9:448–461.
58. Bellack AS, Sayers M, Mueser KT, Bennett M. Evaluation of social problem solving in schizophrenia. *J Abnorm Psychol.* 1994;103:371–378.
59. Hamera E, Brown CE. Developing a context-based performance measure for persons with schizophrenia: the test of grocery shopping skills. *Am J Occup Ther.* 2000;54:20–25.
60. DeHaes JC, deRuiter JH, Tempelaar R, Pennick BJ. The distinction between affect and cognition in the quality of life of cancer patients-sensitivity and stability. *Qual Life Res.* 1992;1:315–322.
61. Pavot W, Deiner E. The affective and cognitive context of self-reported measures of subjective well-being. *Soc Indic Res.* 1993;28:1–20.
62. Patterson TL, Moscona S, McKibbin CL, Hughs T, Jeste DV. UCSD performance-based skills assessment (UPSA): development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull.* 2001;27:235–245.
63. Bellack AS, Brown CH, Thomas-Lohrman S. Psychometric characteristics of role play assessments of social skill in schizophrenia. *Behav Ther.* In press.
64. Bellack AS. Recurrent problems in the behavioral assessment of social skills. *Behav Res Ther.* 1983;21:29–42.
65. Wright S. Correlation and causation. *J Agric Res.* 1921;20:557–585.
66. Wright S. The method of path coefficients. *Ann Math Stat.* 1934;5:161–121.
67. Joreskog K, Sorborn D. *LISREL 8.7*. Chicago, Ill: Scientific Software International; 2005.
68. Bentler PM. *EQS Structural Equations Program Manual*. Encino, Calif: Multivariate Software, Inc; 1995.
69. Bentler PM, Wu EJC. *EQS for Windows User's Guide*. Encino, Calif: Multivariate Software, Inc; 1995.
70. SPSS Inc. *AMOS*, Chicago, Ill: SPSS Inc; 2005.
71. SAS Institute. *CALIS*, NC: SAS Institute; 2005.
72. Bollen KA, Long JS. *Testing Structural Equation Models*. Newbury Park, Calif: Sage; 1993.
73. Heaton RK, Marcotte TD, Rivera-Mindt M, et al. The impact of HIV-associated neuropsychological impairment on everyday functioning. *J Int Neuropsychol Soc.* 2004;10:317–331.
74. Bowie DR, Reichenberg A, Patterson TL, Heaton RK, Harvey PD. Determinants of real world functional performance in schizophrenia: correlations with cognition, functional capacity, and symptoms. *Am J Psychiatry.* 2006;163:418–425.
75. Lee HC, Cameron D, Lee AH. Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accid Anal Prev.* 2003;35:797–803.
76. Lowenstein DA, Durara R, Rubert MP, Arguelles T. Neuropsychological test performance and prediction of functional capacities among Spanish-speakers and English-speaking patients with dementia. *Arch Clin Neuropsychol.* 1995;10:75–88.
77. Marcotte TD, Heaton RK, Wolfson T, et al. The impact of HIV-related neuropsychological dysfunction on driving behavior. *J Int Neuropsychol Soc.* 1999;7:579–592.
78. Patterson TL, Goldman S, McKibbin CL, Hughes T, Jeste DV. USCD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull.* 2001;27:235–245.
79. Patterson TL, Lacro J, McKibbin CL, Goldman S, Hughes T, Jeste DV. Medication Management Ability Assessment (MMAA): results from a performance-based measure in older outpatients with schizophrenia. *J Clin Psychopharmacol.* 2002;22:11–19.
80. Rizzo M, McGehee DV, Dawson JD, Anderson SN. Simulated car crashes at intersections in drivers with Alzheimer disease. *Alzheimer Dis Assoc Disord.* 2001;15:10–20.
81. Brekke JS, Kay DD, Kee KS, Green MF. Biosocial pathways to functional outcome in schizophrenia. *Schizophr Res.* 2005;80:213–225.
82. Green MF. What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry.* 1996;153:321–330.
83. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the “right stuff”? *Schizophr Bull.* 2000;26:119–136.
84. Green MF, Oliver B, Crawley JN, Penn DL, Silverstein S. Social cognition in schizophrenia: recommendations from the MATRICS New Approach Conference. *Schizophr Bull.* 2005;31:882–887.
85. Mueser KT, Bellack AS. Social skills and social functioning. In: Mueser KT, Tarrrier N, eds. *Handbook of Social Functioning in Schizophrenia*. Boston, Mass: Allyn and Bacon; 1998:79–96.
86. Ell K. Social networks, social support, and health status: a review. *Soc Serv Rev.* 1984;58:133–149.
87. Brekke JS, Long JD. Community-based psychosocial rehabilitation and prospective change in functional, clinical, and subjective experience variables in schizophrenia. *Schizophr Bull.* 2000;26:667–680.
88. Strauss JS, Carpenter WT Jr. The prediction of outcome in schizophrenia: I. Characteristics of outcome. *Arch Gen Psychiatry.* 1972;27:739–746.
89. Chapman LJ, Chapman JP. The measurement of differential deficits. *J Psychiatr Res.* 1978;14:303–311.
90. Wykes T, Reeder C. *Cognitive Remediation for Schizophrenia: Theory and Practice*. London, England: Routledge; 2005.
91. Reeder C, Newton E, Frangou S, Wykes T. Which executive skills should we target to affect social functioning and symptom change? A study of a cognitive remediation therapy program. *Schizophr Bull.* 2004;30:87–100.
92. Spaulding WD, Reed D, Sullivan M, Richardson C, Weiler M. Effects of cognitive treatment in psychiatric rehabilitation. *Schizophr Bull.* 1999;25:657–676.
93. Bell MD, Bryson G, Fiszdon JM, Greig T, Wexler BE. Neurocognitive enhancement therapy and work therapy in schizophrenia: work outcomes at 6 months and 12 month follow-up. *Biol Psychiatry.* 2004;55:335.
94. Schank RC. *Scripts, Plans, Goals, and Understanding*. Hillsdale, NY: Earlbaum Associates; 1997.
95. Johnson-Laird PN. *Mental Models—Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, Mass: Harvard University Press; 1983.
96. Flavell JH, Miller PH, Miller SA. *Cognitive Development*. 4th ed. Upper Saddle River, NJ: Prentice Hall; 2002.