Can Severely Mentally Ill Adults Reliably Report Their Needs?

Pатрик W. Cоrrиган, Фл. D., Бредт Buican, B.S., и Стэйнлей McCraeKen, Ph.D.¹

Попонета психиатрической реабилитации убеждена, что активное участие пациентов в плане реабилитации является ключевым. Однако, не все пациенты, страдающие от тяжелого психического здоровья, могут надежно отражать свои потребности. В связи с этим, в данной работе, были проведены два исследования: одно было проведено дважды, с интервалом в одну неделю. В результате, было установлено, что индивидуальные и стандартные показатели индикаторов в обоих случаях были более надежны (по сравнению с усредненными данными) на основе более высоко структурированных и стандартизованных показателей. Дальнейшие исследования показали, что данные о потребностях были значительно связаны с уровнями притязаний. Анализ методических средств должны быть обеспечены чтобы помочь пациентам с нарушенным мышлением и бедным здоровьем получать надежные данные.


Одна из целей психиатрической реабилитации – помочь пациентам со сложными состояниями значительно увеличить свои возможности. Эти цели достигаются при активном участии пациентов в плане реабилитации, особенно когда они могут общаться с их специалистами по идентификации и инструментальному обоснованию нужд (Антони и Liberman, 1992; Bachrach, 1992). В качестве активного участника в процессе, пациенты, в той или иной степени, могут представлять свои текущие потребности, т.е., качество их отчета не подвергается влиянию сложных состояний, которые могут искать их ответы. Несмотря на то, что есть общее мнение, что введение необходимо для успешного преодоления этих проблем, есть исследования, которые показывают, что введение может повышать уровень знаний об оценке и выполнении задач. С другой стороны, не все исследования подтверждают это утверждение. Надежность должна быть обеспечена, чтобы пациенты могли отвечать адекватно.

Изучение показывает, что введение может быть использовано для оценки уровня знаний об оценке и выполнении задач. С другой стороны, не все исследования подтверждают это утверждение. Надежность должна быть обеспечена, чтобы пациенты могли отвечать адекватно.

1 University of Chicago, Center for Psychiatric Rehabilitation, 7230 Arbor Drive, Tinley Park, Illinois 60477. Send reprint requests to Dr. Corrigan.

This study was made possible, in part, by a grant from the Home Health Care Research Committee of the University of Chicago.

523
affected by characteristics of the test. Test items written in standardized format (i.e., close-ended test items presented in a highly structured format) will affect the reliability of responding differently than an individualized format (i.e., open-ended questions without structured response categories; Ciarlo et al., 1986). As in the Patient Satisfaction Inventory study, Likert scales are frequently adopted for standardized items and are valued because they minimize confounds resulting from idiosyncratic response styles. Similar scales could be developed to represent the patients' perception of needs; e.g., patients rate their satisfaction with housing on a Likert scale. The idiosyncratic nature of a patient's needs, however, suggests that individualized assessment strategies are also essential. Items representing an individualized format (Ciarlo et al., 1986; Mintz and Kiesler, 1982) include open-ended questions that provide the subject an opportunity to more broadly respond; e.g., “What specific housing needs do you have?” Items representing an individualized format, however, are likely to yield less reliable responses because they are less structured.

Reliability of patients' responses to individualized and standardized items of a needs assessment will be assessed in this study with standardized items expected to yield more reliable results. Patient reliability on these two item formats were compared with that of control subjects and were expected to be significantly lower. The reliability of patients' responding to individualized items was expected to be diminished by their thought disturbance.

**Methods**

**Subjects**

Severely mentally ill outpatients attending a psychiatric rehabilitation program were invited to participate in this study within 1 week of intake into the program. Patients who provided informed consent were included in the study if they were between 18 and 65 years old, did not have a primary DSM-III-R chart diagnosis of organic mental disorder, had at least a third grade reading level on the Wide Range Achievement Test-Revised (Jastak and Wilkinson, 1984), and had 20/30 corrected vision as measured on the Snellen eye chart. The sample was intentionally not limited to patients with DSM-III-R diagnoses in the schizophrenia spectrum to avoid confounding thought disturbance in general with a specific deficit that might be unique to schizophrenia. Attempts were, however, made to include equal numbers of schizophrenic and nonschizophrenic subjects to compare reliability across the two diagnostic areas.

Ninety-four percent of the subjects meeting inclusion criteria agreed to participate (N = 35). The mean (± SD) age of the 35 patients participating in the study was 33.9 ± 8.9 years, and mean years of education were 12.2 ± 2.0. The sample was 45.7% female. In terms of marital status, 71.4% of the sample were single, 2.9% were married, and 25.7% were divorced or widowed. They were 57.1% African-American and 42.9% Caucasian. For their principle DSM-III-R diagnoses, 42.1% suffered schizophrenia or schizoaffective disorder, 39.4% affective disorder, and 18.1% axis II diagnoses (i.e., borderline, schizotypal, and paranoid personality disorders). Eighty-five percent of the sample was self-administering antipsychotic medications (mean chlorpromazine equivalents, 664.3 ± 1030 mg). Ninety-five percent of subjects receiving antipsychotic medication were also receiving benzotrpine (88 ± 1.29 mg).

The reliability of patients' responses was compared with that of 16 individuals who had no history of inpatient hospitalization or medication for psychiatric reasons. Subjects in this group were younger than patients, with a mean age of 29.1 ± 8.1 years (F[1,49]) = 3.29, p < .10). They were 75% female, a significant difference from the patient sample (χ² = 3.80, p < .05). The control group was significantly better educated (F[1,49]) = 50.95, p < .0001), having completed 16.8 ± 2.3 years of schooling. The marital status of control subjects was similar to the patient sample: 68.7% were single, 12.5% were married, and 18.8% were divorced or widowed. The two samples differed significantly in terms of race (χ² = 14.2, p < .001): control subjects were 81.2% Caucasian, 6.2% African-American, and 12.5% Asian. Relationships between test-retest reliability on the needs assessment and these demographic variables are examined in Results.

**Needs Assessment**

Subjects were administered the Needs and Resources Assessment (NARA) as an individualized and standardized means of assessing needs across 13 domains of functioning (Corrigan et al., 1995). The NARA included two types of individualized questions. First, subjects were instructed to list specific needs per domain in response to open-ended questions; e.g., “What housing needs do you have?” Examiners were provided a set of follow-up questions that might help patients elaborate or otherwise clarify a response (e.g., “Tell me more about that.”). Subjects were also asked to list specific resources they might avail for each need; e.g., “You said you needed a new apartment. What resources do you have to help you obtain this apartment?” The reli-
ability of responses to resource questions, however, depend on whether needs are identified for each NARA item. This lack of independence confounded analysis of resource reliabilities across samples, so reliability information about resources was not examined in this study.

Two sets of standardized questions were also included in the NARA. The first addressed overall satisfaction with each of the 13 domains. For example, the question, "How satisfied are you with your physical health?" was rated on a scale from 1 to 7 (7 = delighted). This construct was chosen for the standardized items because ratings of satisfaction with various functioning domains was thought to correspond closely to the number of needs reported per domain. Subjects who are dissatisfied with a specific domain (e.g., housing) were generally expected to report more needs in that domain (e.g., "I need a new roommate, a better apartment, more money for utilities, and better furniture."). Conversely, subjects who were satisfied with individual functioning domains were expected to report no needs in that area.

The second set of standardized questions was concerned with the subject’s perception of the importance of each need. Subjects were also asked to rate how important each need was on a 7-point scale (7 = most important). Like resources, reliability of importance ratings were dependent on responses to the needs question and so were not included in the study.

Previous research has shown NARA subscales for individualized and standardized items to have satisfactory internal consistency and test-retest reliability (as measured by intraclass correlation coefficients) for the overall needs and satisfaction scales (Corrigan et al., 1985). The concurrent validity of the measure has also been demonstrated. Subscales were shown to converge significantly with predicted measures of social support, quality of life, and depression.

The NARA was administered as an interview to all subjects, rather than as a self-administered pencil and paper measure, because previous research suggested that interview formats yield more reliable responses from severely mentally ill subjects (Corrigan, 1990; Corrigan and Jakus, 1994; Kalman, 1983; LeBow, 1982). Individual NARA test questions were repeated when subjects asked. Examiners were instructed not to define or otherwise elaborate on any of the terms in the NARA. Scales for the standardized items were presented to subjects on 8 x 11 inch cards. The NARA was administered twice to both samples, with 1 week intervening, to examine its test-retest reliability.

Test-retest reliability was assessed by determining the raw agreement (Hartmann, 1982) of responses to items comprising NARA subscales for each subject. Raw agreement provides an index of subject-specific reliability; i.e., the reliability of each subject on reporting NARA items. Raw agreement of responses to needs items equaled the number of specific needs reported at both the first and second administration divided by the total number of distinct needs generated at the two test administrations. For example, if the subject reported 10 needs at the first administration and 6 needs at the second, 4 of which were also mentioned at the first, 12 distinct needs were identified. Raw agreement equaled \( r_s = .33 \). Two independent raters determined whether needs listed at the first NARA administration corresponded with needs listed at the second administration. Raters were trained on the raw agreement task using pilot NARA data and demonstrated an intraclass correlation coefficient \((r_s, 1, 1, \text{Shrout and Fleiss, 1979})\) exceeding .75 before reviewing data generated from this study.

Agreement on ratings of satisfaction was defined as item ratings within 1 point of each other at times 1 and 2; agreement was coded 1 for each domain, whereas nonagreement was coded 0. Preliminary analyses suggested that more liberal criteria for agreement (e.g., item ratings within 2 points) or more conservative criteria (exact item ratings at time 1 and time 2) were not sufficiently sensitive to distinguish patients from control subjects. Therefore, the index of overall raw agreement equaled the number of agreements on satisfaction ratings divided by total ratings. Two of the 13 NARA item domains (drug-related problems and legal problems) were not included in this study because control subjects did not report many needs in response to these items, thereby inflating their raw agreement.

**Measures of Cognitive Deficit**

Verbal intelligence in patients was assessed using the vocabulary subtest of the WAIS-R. Research has shown that correlations of this subtest with verbal IQ range from .82 to .91, depending on the age of the subject and, with overall IQ, range from .82 to .87 (Wechsler, 1981). Psychiatric symptoms were assessed using the expanded version of the Brief Psychiatric Rating Scale (BPRS) comprising 24 symptoms rated on a 7-point severity scale (7 = extremely severe; Lukoff et al., 1986). Interviewers who administered the BPRS had been trained previously to a minimum intraclass correlation coefficient \((r_s, 1, 1, \text{Shrout and Fleiss, 1979})\) of .80 based on consensus ratings at our lab. A thinking disturbance factor equaled the mean of scores on BPRS items representing unusual thought content, hallucinations, and conceptual disorganization (Hedlund and Vieweg,
1980). The BPRS and vocabulary subscales of the WAIS-R were administered when the first NARA was given.

### Results

Analyses showed that raw agreement of responses to needs and satisfaction did not significantly interact with age, education, marital status, race, or gender of the patients or the controls. Moreover, raw agreement was not associated with patients’ diagnosis or medication level.

#### How Reliable Were the Responses of Patients and Controls to Individualized and Standardized Items?

The mean ± SD of raw agreements for responses to the 11 needs and satisfaction items of the NARA are summarized in Table 1 for patient and control samples. A 2 x 2 (group by item format) multivariate analysis of variance was completed with raw agreements from the 11 NARA items as dependent variables. Results showed a significant main effect for group ($F[1,49] = 12.36, p < .001$) and item format ($F[1,49] = 97.66, p < .0001$). Nonsignificant trends were found for the interaction ($F[1,49] = 2.96, p < .09$).

Although nonsignificant, this interaction is apparent in Figure 1, in which the mean of raw agreements representing responses to items assessing needs (individualized format) and satisfaction (standardized format) across the 11 items are summarized for patients and control subjects. Post hoc analyses showed that both controls and patients were more reliable in responding to satisfaction than needs items ($F[1,15] = 158.34, p < .0001$ and $F[1,34] = 40.64, p < .0001$, respectively). Patients earned lower reliabilities than control subjects on satisfaction items ($F[1,49] = 15.48, p < .0005$); the differences on need items were described by nonsignificant trends ($F[1,49] = 3.00, p < .10$).

#### How Poor Is the Patient’s Reliability of Responses to Needs Assessment?

A case by case examination of patients’ raw agreements further illustrates the poor reliability of patients’ responses to individualized (needs) versus standardized (satisfaction) items. Results showed that no more than 48.5% of patients were able to report less than half of their needs at both test administrations. A much larger group (88.6% of patients), however, was able to show ≥50% raw agreements on the satisfaction items.

#### Do Patients Respond More Reliably to Specific Content Areas of the Needs Assessment?

Additional post hoc analyses showed interesting differences among raw agreements to specific needs and satisfaction items on the NARA. A repeated measures analysis of variances was significant for patients across the 11 needs ($F[1,10] = 5.65, p < .0001$) and satisfaction items ($F[1,10] = 2.10, p < .05$). Further examination showed that the raw agreement of responses to mental health needs and mental health satisfaction was among the lowest. In particular, post hoc planned contrasts showed that raw agreement of responses to mental health-related needs was significantly below ($p < .05$) raw agreements of five other content areas. Similarly, raw agreement of responses to mental health-related satisfaction was significantly below ($p < .05$) five other satisfaction items. No such differences were found in the control sample. Repeated measures analyses of variance were not significant for control subjects across need ($F[1,10] = .92$) or satisfaction items ($F[1,10] = 1.32$).
How Do Cognitive Deficits Correlate with Raw Agreements?

In terms of cognitive deficit, patients earned a mean vocabulary score of only 32.7 ± 16.1 on the WAIS-R. They exhibited fairly benign scores on the BPRS thinking disturbance factor (2.4 ± 1.5). Pearson product-moment correlations among overall raw agreements for the needs and satisfaction items and the two measures of cognitive functioning were determined. We found the WAIS-R vocabulary scores not to be significantly associated with either of the NARA raw agreements (needs, r = .25; satisfaction, r = .14). BPRS thinking disturbance scores of the patient sample were, however, significantly associated with raw agreements for needs (r = -.50, p < .01). Nonsignificant trends described the relationship among thinking disturbance and the raw agreement of responses to satisfaction (r = -.29, p < .10). The relationship of raw agreement with thinking disturbance seems to be caused by the deleterious effects of the cognitive deficit rather than the effects of psychiatric symptoms in general. Correlations among overall raw agreement scores and the sum of all BPRS items were low (p < .20).

The effect of thinking disturbance is apparent in Figure 2, in which the raw agreements of control subjects were compared with the scores of 15 patients who showed an absence of thinking disturbance (i.e., BPRS factor score = 3). Results of a 2 × 2 analysis of variance failed to show a significant difference between groups on raw agreement score (F[1,29] = 1.81, NS). The repeated measures factor, however, continued to be significant (F[1,29] = 91.63, p < .0001), as was the interaction (F[1,29] = 6.09, p < .05).

Discussion

The reliability of the responses of severely mentally ill outpatients and control subjects to needs assessments varies with the type of assessment strategy used. Responses to individualized assessments were significantly less reliable over the test-retest interval when compared with standardized assessments. Lack of reliability on the individualized tests was substantial; about half of the subjects were unable to provide agreement scores of at least 50% in response to individualized items assessing needs. This low test-retest reliability probably does not represent natural change in the measured constructs over the brief interim between test administrations. It is unlikely that specific needs would vary significantly over 1 week's time. Rather, these findings seem to represent the effects of relatively unstructured interview items. Contrast these findings to mean raw agreements of responses for the standardized NARA items, which, for the most part, exceeded 75%.

The nonsignificant trend for the interaction tentatively suggested that differences in raw agreements between responses to individualized items
and standardized items were greater for the control subjects than the patients. This finding, if supported by future research, implies that the greater structure of standardized items enhances the reliability of the control subjects responding to a greater degree than patients. These results may suggest that the cognitive deficits of severe mental illness interfere with their ability to understand and respond to individualized items. Alternatively, this difference might be attributed to the significant differences in several demographic variables across samples. These variables were not, however, found to be significantly associated with raw agreements of individualized or standardized NARA items. Still, future research needs to probe the reliability of responding between individuals who suffer severe mental illness and control samples who are matched on demographic variables.

The difference between individualized and standardized items may be a result of differences in the constructs measured by the two assessment approaches. Individualized items represent needs in each domain, whereas standardized items represent satisfaction with the domains. In Methods, however, we argued that the two constructs overlapped greatly; satisfaction (or lack thereof) with any domain would directly lead to perceived needs in that domain. Moreover, findings in this study seemed to correspond with the individualized-standardized distinction in assessment approach. Raw agreements for the individualized items were significantly lower than standardized items for all 11 domains of the control group and for 10 of 11 domains in the patient sample. This question needs to be revisited in other research that assesses the reliability of responding to individualized and standardized items on different constructs.

Further analyses were conducted to identify patient variables that were associated with reliable responding. Correlational analyses failed to show a significant relationship between raw agreements and verbal intelligence of patients. The test-retest reliability of patients' responses was, however, shown to be diminished by their thinking disturbances. Raw agreement of responses to NARA needs items was significantly associated with thinking disturbance as assessed on the BPRS; nonsignificant trends were found between thinking disturbance and raw agreement for response to satisfaction items. The importance of this relationship was even more apparent when comparing differences in the reliabilities of patients who exhibited little or no thinking disturbance with controls. No significant differences were found in the raw agreements of the control sample and patients in this subgroup.

The relationship between thinking disturbance and the reliability of responding is even more pronounced because of the relatively low mean and standard deviation of the thinking disturbance factor for the sample as a whole. This finding seemed to be specific to cognitive deficits related to psychosis and not psychiatric symptoms in general; e.g., no significant relationships were found between overall symptoms and reliability. The combined results suggest that psychotic patients are likely to report their needs less reliably. Findings from this study were, however, based on cross-sectional methods. Future research needs to determine whether reliable responding to needs assessments improves as thinking disturbance remits.

Differences in raw agreements between responses to individualized items and responses to standardized items may be confounded by differences in the quality of items sampling needs and satisfaction rather than representing the effects of less structure in individualized items. The individualized format asks for specific listing of needs, whereas the standardized format samples more global impressions. Future research needs to account for the effects of these qualitative differences on patient and staff differences in raw agreements. This confound should not, however, distract the reader from the major focus of the study; namely, deficits related to thinking disturbance seem to further diminish the reliability of patients’ responding to needs assessments.

Results of the study also suggested that raw agreement varied with the content of the needs assessment item. In particular, the reliability of reports about needs related to, and satisfaction with, mental health were among the lowest. This finding suggests that patients may be somewhat ambivalent about their personal goals related to their mental health problems. Clinicians should be sensitized to this difference when trying to comprehend patients' needs in this area.

Findings in this paper suggest that clinicians should cautiously consider individualized information about needs and resources, especially details provided by patients disturbed by cognitive deficits related to psychotic symptoms. Given this limitation, how can a profile representing the idiosyncratic needs of agitated or acutely ill patients be obtained? If additional research suggests that reliability of responding improves as psychosis remits, clinicians might wait until the patient is better organized cognitively. Alternatively, strategies that highly structure individualized assessment may improve responding. One such approach is being tested at our center. Patients are provided a menu
of response options to NARA test items; e.g., which of the following option—find a new roommate, move into a new apartment, move back in with your parents—represents one of your housing needs? Given the complexity of various options, a decision tree is used to guide patients through the multiple levels within one needs domain. Needs relevant for independent living yield a separate menu from those related to continued family living. Limits to patient participation in rehabilitation planning can be diminished by increasing the structure of the needs assessment task.

References


