Detecting Improvement in Quality of Life and Symptomatology in Schizophrenia

by Joyce Cramer, Robert Rosenheck, Weichun Xu, William Henderson, Jonathan Thomas, and Dennis Charney for the Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia

Abstract

Instrument-based scores are often used as outcome measures. However, little is known about what changes in scores mean in terms of a clinical assessment of improvement or deterioration. The purpose of this report was to determine how much change in standard instrument scores represents a clinically detectable improvement or deterioration. The Veterans Affairs (VA) Cooperative Study of Clozapine in Refractory Schizophrenia evaluated 423 patients on clozapine or haloperidol. Symptoms and quality of life scales were completed at baseline; 6 weeks; and 3, 6, and 12 months. Among patients judged as “improved” by clinicians, the average percentage changes were a 21 percent decrease in Positive and Negative Syndrome Scale (PANSS) scores and a 26 percent increase in Quality of Life Scale (QLS) scores across all followup periods. The change in mean seven-point item scores were −0.46 (PANSS) and 0.23 (QLS). A major gain in clinically assessed improvement to “much better” was associated with a 45 percent decline in PANSS scores and 50 percent increase in QLS scores (change in mean seven-point item scores −0.88 and 0.92, respectively). Thus, modest changes in psychometric scales assessing symptoms and quality of life reflect clinically detectable improvement.

Keywords: Schizophrenia, Quality of life, Clozapine, Haloperidol, Clinical trial


The development of standardized instruments to assess health-related quality of life (HRQOL) and symptomatology has helped clinicians and researchers better assess patients' health and function (Cramer and Spilker 1998). However, the lack of a basis (gold standard) with which to compare instrument-based scores limits their usefulness. Although interrater reliability rarely is evaluated, the judgment of an experienced physician who can integrate clinical input from multiple domains into a global rating of improvement or deterioration is generally accepted as standard.

Evaluation of people with chronic schizophrenia is complicated by the difficulty of obtaining reliable and valid responses on direct questioning (Cramer et al., 2000), as well as by the apparent lack of awareness of deficits and difficulty in perceiving change in some patients (Sen 1992; Atkinson et al. 1997). In a survey of three HRQOL instruments, we reported that instruments based on structured interviews provided the most consistent data for a clinical trial (Cramer et al., 2000). The structured interview format of the QLS (Heinrichs et al. 1984) and the Level of Function scale (Strauss and Carpenter 1977) were sensitive to change based on a number of standards. Self-reports from a Quality of Life Interview (Lehman 1988, 1996), in contrast, were less sensitive. The symptomatic features of schizophrenia can be assessed with the PANSS (Kay et al. 1987), which is an expansion of the widely used Brief Psychiatric Rating Scale (Overall and Gorham 1962). A structured clinical interview for the PANSS (SCI-PANSS) was developed by Opler and Ramirez (1992) to enhance the reliable categorization of symptoms by trained raters.

Scores from these instruments can be analyzed to determine statistically significant differences between treatment groups. However, given a large sample size in a clinical trial, many comparisons may reach statistical significance. The problem faced by researchers then is a lack of information about the association between the scores and clinically meaningful improvement. Using data from a VA Cooperative Study of Clozapine in Refractory Schizophrenia (Rosenheck et al. 1997), we proposed to determine the average magnitude of change in PANSS and QLS scores that is associated with clinician-detected improvement or deterioration.

Send reprint requests to J. Cramer, VA Connecticut Healthcare System (G7E), 950 Campbell Avenue, West Haven, CT 06516-2770; e-mail: joyce.cramer@yale.edu.
Methods

VA Cooperative Study #17 enrolled 423 patients with treatment-resistant schizophrenia in a 1-year, prospective, randomized, double-blind clinical trial comparing clozapine and haloperidol for efficacy and cost-effectiveness (Rosenheck et al. 1997). A central Human Rights Committee at the Hines VA Cooperative Studies Program Coordinating Center and committees at each of the 15 participating hospitals approved the study. Eligible patients were those who were diagnosed as having schizophrenia refractory to adequate trials of at least two standard antipsychotic medications, who were hospitalized for psychosis 30-365 days out of the preceding year, and who consented to participate. Patients were randomized to clozapine or haloperidol, with dose adjustments as clinically needed. Baseline assessments, performed while the patient was stable on prestudy medications, included measures of HRQOL, symptoms, and adverse effects of current treatment. Followup assessments were performed at 6 weeks, 3 months, 6 months, and 12 months. Symptom and HRQOL scales were completed by research coordinators who used a structured interview format to elicit responses from patients (see Instruments section below). Clinical status was rated by a physician-investigator experienced with the schizophrenia population who did not see symptom or HRQOL scores. The clinical status rating compared current status with prestudy status on a five-level scale: 1 = much worse, 2 = worse, 3 = unchanged, 4 = improved, and 5 = much better. Interrater reliability assessments performed quarterly throughout the clinical trial demonstrated excellent levels of agreement and lack of rater drift among and between research coordinators and physician-investigators for symptom and HRQOL scales (Cicchetti et al. 1997, 1999). Interrater reliability testing was not performed for clinical status ratings.

Instruments. Symptom assessment included the PANSS (Kay et al. 1987), a widely used rating scale for symptoms of psychosis that contains 30 seven-point items in three subscales, requiring approximately 30 minutes for completion. A series of issues related to positive, negative, and general attributes of schizophrenia are probed with specified probes. Anchors are provided to categorize responses into scores on a seven-point scale. Lower scores reflect fewer symptoms, and higher scores reflect more problems. A total score sums the positive, negative, and general subscale scores, reflecting overall clinical symptoms. To provide consistency across scales in this trial, we used a range of possible total scores of 0 to 180 points, with items scored from 0 to 6 points. This differs from the approach of other investigations, which used a total score range of 30 to 210 points, with items scored from 1 to 7 points.

HRQOL assessment was based on the QLS (Heinrichs et al. 1984), which contains 21 seven-point items in four subscales (interpersonal relations, instrumental role function, intrapsychic foundations, and common objects and activities). Completion requires approximately 45 minutes, during which various topics are explored using specified probes. Anchors are provided to categorize responses into scores on a seven-point scale. Higher scores reflect normal or unimpaired function, and low scores reflect severely impaired function. The range of possible total scores is 0 to 126 points.

Improvement or decline in overall clinical condition from baseline to each followup assessment was categorized by clinicians in a five-level global rating.

Analyses. The direction and proportion of change in PANSS and QLS scores between baseline and 6 weeks; 3, 6, and 12 months of followup were calculated. Data are presented as change in percentage of total scores and change in mean item scores for the QLS and PANSS from baseline to each followup period. Mean item scores are an average of the response items (e.g., a seven-point range of possible responses might average 4.0 at baseline, declining to 3.5 at followup, for a mean item change of -0.5). Analyses of variance were used to test for significant differences in psychometric scales across clinical change ratings, followed by t tests.

Results

The patients were 98 percent male, mean age 44 ± 8 years, 65 percent white, 30 percent black, and 4 percent Hispanic, with a mean of 12 ± 2 years of education. Age at onset of schizophrenia averaged 22 ± 5 years. Data were collected from patients who completed PANSS and QLS interviews at 6 weeks (n = 378 and 376, respectively), 3 months (n = 356 and 353), 6 months (n = 325 and 323), and 12 months (n = 308 and 317) for comparison with clinical global ratings at that time period. Mean scores at baseline were PANSS, 61 ± 15 (scale of 0–180 points), and QLS, 38 ± 17 (scale of 0–126 points). Two approaches were taken to compare clinical ratings with instrument score: average percentage changes in scores and mean item scores.

Percentage Change in Scores. The percentage change in scores from the baseline rating was determined for each rating level of clinical improvement at each assessment (figures 1 and 2). The relationship between change in PANSS or QLS scores and clinical improvement or worsening was similar at all followup time points except for the "much worse" clinical category, in which sample sizes were small. Because sample sizes for the "much worse"...
Figure 1. Positive and Negative Syndrome Scale scores: Mean percentage change scores at each global level

Overall mean percentage change scores: worse = +5%, unchanged = -5%, improved = -21%, much better = -45%.

Too few patients at Level 1 for comparison.

Figure 2. Quality of Life Scale scores: Mean percentage change scores at each global level

Overall mean percentage change scores: worse = +3%, unchanged = +6%, improved = +26%, much better = +50%.

Too few patients at Level 1 for comparison.
clinical category were small, this category was deleted from figures 1 and 2. The direction of change in scores was in keeping with clinical ratings, with the magnitude of change increasing in the appropriate direction.

Averaging all the PANSS data (figure 1), in which lower scores represent reduced symptoms, overall mean percentage change in scores from baseline showed some increase in symptomatology (5%) at Level 2 “worse,” small improvement (−5%) at Level 3 “unchanged,” greater reduction in symptoms (−21%) at Level 4 “improved,” and more than twice as much change at Level 5 “much better” (−45%). For the QLS (figure 2), in which higher scores represent better HRQOL, overall mean percentage change scores averaged 3 percent at Level 2 “worse,” 6 percent at Level 3 “unchanged,” 26 percent at Level 4 “improved,” and 50 percent higher at Level 5 “much better.”

Mean Item Scores. We also assessed how much change on average in the seven-point response scales was associated with clinician ratings of improvement or decline. PANSS mean item scores were based on 30 items, and QLS mean item scores were based on 21 items. Tables 1 and 2 list the mean change in baseline in PANSS and QLS mean item scores.

### Table 1. PANSS mean change in item scores

<table>
<thead>
<tr>
<th>Time points</th>
<th>Much worse</th>
<th>Worse</th>
<th>Unchanged</th>
<th>Improved</th>
<th>Much better</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 wks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 378</td>
<td>n = 6</td>
<td>n = 26</td>
<td>n = 148</td>
<td>n = 192</td>
<td>n = 6</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.47 (0.89)</td>
<td>0.36a (0.49)</td>
<td>−0.11ab (0.49)</td>
<td>−0.40b (0.49)</td>
<td>−0.79 (0.42)</td>
</tr>
<tr>
<td>3 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 356</td>
<td>n = 2</td>
<td>n = 35</td>
<td>n = 115</td>
<td>n = 191</td>
<td>n = 13</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.45 (0.83)</td>
<td>0.03 (0.71)</td>
<td>−0.16a (0.50)</td>
<td>−0.44ab (0.49)</td>
<td>−0.85b (0.39)</td>
</tr>
<tr>
<td>6 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 327</td>
<td>n = 2</td>
<td>n = 20</td>
<td>n = 98</td>
<td>n = 192</td>
<td>n = 15</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>−0.47 (0.33)</td>
<td>−0.13 (0.60)</td>
<td>−0.13a (0.48)</td>
<td>−0.48b (0.49)</td>
<td>−0.73 (0.48)</td>
</tr>
<tr>
<td>12 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 308</td>
<td>n = 4</td>
<td>n = 12</td>
<td>n = 82</td>
<td>n = 178</td>
<td>n = 32</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.02 (0.43)</td>
<td>−0.03 (0.46)</td>
<td>−0.22a (0.53)</td>
<td>−0.50ab (0.55)</td>
<td>−0.97b (0.52)</td>
</tr>
<tr>
<td>Mean item PANSS (SD)</td>
<td>0.20 (0.64)</td>
<td>0.08 (0.60)</td>
<td>−0.15 (0.50)</td>
<td>−0.46 (0.51)</td>
<td>−0.88 (0.47)</td>
</tr>
</tbody>
</table>

Note.—PANSS = Positive and Negative Syndrome Scale; SD = standard deviation.

1 Superscripts that match indicate that mean item scores are significantly different from each other.

2 Mean change from baseline to all follow-up assessments.

### Table 2. QLS mean change item scores

<table>
<thead>
<tr>
<th>Time points</th>
<th>Much worse</th>
<th>Worse</th>
<th>Unchanged</th>
<th>Improved</th>
<th>Much better</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 wks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 376</td>
<td>n = 6</td>
<td>n = 26</td>
<td>n = 148</td>
<td>n = 190</td>
<td>n = 6</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>−0.65 (0.99)</td>
<td>−0.44a (0.63)</td>
<td>0.00ab (0.54)</td>
<td>0.16bc (0.55)</td>
<td>0.69c (0.47)</td>
</tr>
<tr>
<td>3 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 353</td>
<td>n = 2</td>
<td>n = 34</td>
<td>n = 114</td>
<td>n = 190</td>
<td>n = 13</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>−0.19 (1.35)</td>
<td>−0.25a (0.58)</td>
<td>0.01ab (0.62)</td>
<td>0.19bc (0.65)</td>
<td>0.81c (0.84)</td>
</tr>
<tr>
<td>6 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 325</td>
<td>n = 2</td>
<td>n = 20</td>
<td>n = 96</td>
<td>n = 192</td>
<td>n = 15</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>−0.74 (0.51)</td>
<td>−0.04 (0.72)</td>
<td>0.03a (0.56)</td>
<td>0.26abc (0.71)</td>
<td>0.73b (0.71)</td>
</tr>
<tr>
<td>12 mos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 307</td>
<td>n = 4</td>
<td>n = 13</td>
<td>n = 81</td>
<td>n = 177</td>
<td>n = 32</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.27 (0.68)</td>
<td>−0.13 (0.69)</td>
<td>−0.02a (0.59)</td>
<td>0.32ab (0.74)</td>
<td>1.10b (0.88)</td>
</tr>
<tr>
<td>Mean item QLS (SD)</td>
<td>−0.33 (0.80)</td>
<td>−0.24 (0.63)</td>
<td>0.00 (0.58)</td>
<td>0.23 (0.67)</td>
<td>0.92 (0.56)</td>
</tr>
</tbody>
</table>

Note.—QLS = Quality of Life Scale; SD = standard deviation.

1 Superscripts that match indicate that mean item scores are significantly different from each other.

2 Mean change from baseline to all follow-up assessments.
scores at 6 weeks and 3, 6, and 12 months (see bottom line in tables 1 and 2). As with percentage change in scores, PANSS mean item average changes were monotonic, but were not balanced between improvement and decline: Level 1 “much worse” was 0.20, Level 2 “worse” was 0.08, Level 3 “unchanged” was -0.15, Level 4 “improved” was -0.46, and Level 5 “much better” was -0.88. QLS scores averaged across all assessment times were centered at zero for clinicians’ ratings of “unchanged,” and symmetrical between “worse” and “improved” ratings: Level 1 “much worse” -0.33, Level 2 “worse” -0.24, Level 3 “unchanged” 0.00, Level 4 “improved” 0.23, and Level 5 “much better” 0.92 points. The range of scores across all times was narrow (e.g., at Level 3 “unchanged,” the PANSS range was -0.11 to -0.22 and the QLS range was -0.02 to 0.03).

Analyses of variance of PANSS scores followed by t tests comparing pairs revealed statistically significant differences between Level 2 “worse” and Level 3 “unchanged” at 6 weeks ($p = 0.0001$), between Level 3 “unchanged” and Level 4 “improved” at all times ($p = 0.0001$, all), and between Level 4 “improved” and Level 5 “much better” at 3 and 12 months ($p = 0.007$ and $p = 0.0001$, respectively). Analyses of variance of QLS change scores followed by t tests comparing pairs revealed statistically significant differences between Level 2 “worse” and Level 3 “unchanged” at 6 weeks ($p = 0.0003$) and 3 months ($p = 0.04$), between Level 3 “unchanged” and Level 4 “improved” at all times ($p = 0.02-0.0002$), and between Level 4 “improved” and Level 5 “much better” at all times ($p = 0.02-0.0001$).

Discussion

Using structured assessments of HRQOL, symptoms, and a global clinician rating of improvement at multiple times during 12 months of followup, we are able to answer a critical question in HRQOL research: How much change in instrument scores represents a detectable change in patient status? We used clinicians’ ratings of clinical improvement or deterioration among people with schizophrenia to identify proportional changes in instrument-based HRQOL and symptom scores that correspond to ratings of “unchanged” to “improved” or “worse,” or to “much better” or “much worse.” The terminology implies that ratings other than “unchanged” would be clinically significant. These data demonstrate that among patients judged as “improved” by clinicians, the average percentage changes were a 21 percent decrease in PANSS total scores and a 26 percent increase in QLS total scores, and average change in mean item scores of -0.46 and 0.23, respectively.

A limitation of this study is the absence of change ratings by patients themselves. Unlike Osoba et al. (1988), who asked two groups of cancer patients to rate their change in health status, we based estimates of change on clinician judgment alone. This was because of the difficulty ascertaining a perception of change among patients with schizophrenia (Cramer et al., 2000). People with schizophrenia appear to be less able to provide psychometrically sensitive subjective self-ratings than other populations, because of cognitive deficits, delusions, and other features of the disorder (Cramer et al., 2000). Atkinson et al. (1997) and Sen (1992) described situations in which severely disadvantaged people tend to minimize the adversity of their condition in ways that could affect their judgment of small changes. Although both subjective and objective assessments are important, an external rater may provide the most reliable assessment of HRQOL for people with chronic, treatment-refractory schizophrenia. In this trial, current status ratings were made by clinicians who saw the patients (at least) weekly in the hospital or clinic throughout followup.

Another limitation is the lack of interrater reliability studies of clinician judgment. In addition, objective raters (e.g., clinician or hospital staff members) may miss features of relevance to other parties (e.g., family members). The use of raters other than the patient usually is limited to young children and people with mental handicaps. Hays et al. (1995) demonstrated that when family members and epilepsy patients simultaneously complete questionnaires, responses differed in selected domains. Family members thought that patients had many problems with physical domains, while patients reported problems with emotional domains. Additional research is needed to compare clinician and family ratings with patient self-ratings of global status. Our study is limited by the small number of patients at the lowest improvement level at all time points, and at the highest improvement level at 6 weeks. Some differences among scores might be attributable to drift upward in scoring by the raters or a practice effect in patient responses.

We also demonstrated the stability of the relationship between change in clinical status and instrument score changes over multiple assessments with a large sample of patients assessed at multiple time points over 12 months. Clinicians’ perceptions of patient status are reflected in similar percentage changes from baseline for both symptoms and HRQOL. The amount of change reflecting movement from one level to the next was in a narrow range over the year, especially among those who improved. This suggests that clinicians’ evaluation of improvement consistently correlated with standardized psychometric scales across all time points. However, the incremental amount of change in scores had to be especially large before clinicians rated patients at the highest levels. The small number of patients rated as “worse” or
Much better was due to the fact that few patients either deteriorated or greatly improved throughout followup. Nonetheless, mean change scores and percentage changes were proportional to scores for ratings with larger numbers of patients. The scope of change might differ if comparisons were made to a "washout" or crisis state instead of steady state baseline, as was done in this clinical trial.

Few other reports have determined the minimal amount of change detectable by patients or physicians in an HRQOL questionnaire. In a study of asthma, Juniper et al. (1994) found that a 0.5-point change in a seven-point scale matched minimal improvement or decline, 1.0 points matched moderate change, and 1.5 points matched large change. In a similar study of respiratory disease and heart failure, Jaeschke et al. (1989) found the same minimally important differences in a seven-point scale. Redelmeier et al. (1996) used a slightly different approach, asking patients to judge themselves relative to others with the same condition. They also found that 0.5 points was the minimally important difference using the Chronic Respiratory Questionnaire. We found that small changes in the range of 0.23-0.46 points were associated with detectable improvement, similar to the asthma change of 0.5 points. We also found that the mean QLS item change of 0.92 points (table 2) matched a much larger improvement, similar to the asthma change of 1.0 points. These data suggest some generalizability across illnesses of the amount of clinically detectable change on a seven-point scale.

However, many instruments use other scoring systems. Osoba et al. (1988) used a subjective significance questionnaire to ask patients with breast or lung cancer their perception of change. They were able to categorize HRQOL instrument score changes as small, moderate, or large. Weibe et al. (1997) compared total scores with several objective and subjective global ratings of change. They found that for two epilepsy-specific HRQOL instruments with standardized ranges of 0-100 points (mean 70 points), 12-15 points (approximately 20%) was a reliable minimum clinically significant change. These data are similar to our findings for symptoms (PANSS 21%) and HRQOL (QLS 26%), also suggesting some generalizability for the amount of change that matches clinically detectable change.

Barber et al. (1996) found that the type of global question affected patient response. Among patients reporting a minimal improvement, average domain scores were 0.20 to 0.52 when patients were asked "How well is your asthma controlled?" and 0.06 to 0.13 when patients were asked "Overall has there been any change in your asthma since the beginning of the study?" These findings demonstrate the importance of defining the context of the global question.

Observant physicians who match clinical scenarios with outcomes (e.g., drug serum concentrations above or below specific levels correlate with toxicity or lack of efficacy) usually base the concept of clinically meaningful change on extensive experience. Expanding the usefulness of instrument-based scores to a larger population of physicians is possible when categories of outcomes or conditions can be defined. When differences in scale scores are defined, they can be used as outcome measures (e.g., Glasgow Coma Scale, Teasdale and Jennett 1974). In mental health, laboratory tests are rarely available to measure clinical status. Instead, clinicians often rely on patient-completed questionnaires (e.g., Beck Depression Inventory, Beck et al. 1961) or assessments by an interviewer (e.g., Hamilton Depression Scale [Hamilton 1980] or the Brief Psychiatric Rating Scale [Overall and Gorham 1962]). Little work has been done to define how much change in scores for mental health assessment instruments represents improvement or deterioration in patient status. Although data for both instruments are presented, the PANSS and QLS assess different domains (i.e., symptomatology and quality of life). Selection of the appropriate instrument(s) coupled with definition of clinically important changes can be established as an outcome parameter for treatment programs.

Guyatt et al. (1998) discussed interpretation of score differences, recognizing that individual patients may respond differently from the mean of a population. They encouraged establishment of ranges of change scores for HRQOL and symptom questionnaires that represent meaningful differences for patients. Detecting these changes would allow clinicians to alter patient management. Although there are published population averages for some HRQOL instruments for people with mental illnesses (Lehman 1988, 1996), these averages are not an indication of how much change represents a clinically meaningful difference. In this report, we have explored the definition of clinically meaningful changes in HRQOL and symptoms for people with schizophrenia. These definitions could be useful in planning and interpreting data from clinical trials (e.g., estimation of sample size and identification of clinically significant responses) and in assessing treatment effects (e.g., case management and mental health carve-out managed care). In an area in which psychiatrists rely on standardized ratings, definition of how much change is meaningful provides a basis for documentation of treatment response and stage of condition (e.g., disease management programs). In lieu of the current system of documentation of clinical status, reductions of approximately 20 percent in PANSS total scores or a half-point mean items score can be considered a mark of detectable improvement. For example, Tollefson et al. (1997) found a 10 percent difference in PANSS scores among 1,196 patients treated with olanzapine compared to
Detecting Improvement in Quality of Life

those taking haloperidol. According to our interpretation, these finding are statistically significant but may not be clinically meaningful. Similarly, a structured HRQOL assessment can be added to the patient evaluation plan. Noting an approximately 20 percent increase in QLS total scores or a quarter-point mean item score can be considered a measure of effectiveness of a treatment program. These data also can be used for sample size calculations in future research.

If the evaluation of schizophrenia treatment moves toward widespread use of standardized ratings in clinical practice, the amount of change in scores could become a surrogate marker for significant treatment success in clinical practice, as it is in clinical trials. PANSS or QLS ratings can thus be used to evaluate treatment response, based on clinically meaningful differences in scores.

References


Barber, B.L.; Santangelo, N.C.; and Epstein, R.S. Impact of the global on patient perceivable change in an asthma specific QOL questionnaire. Quality of Life Research, 5:117–122, 1996.


**Acknowledgments**

This work was supported by the Department of Veterans Affairs Health Services Research and Development Service and the Cooperative Studies Program. We thank Jennifer Cahill for programming support. The Clozapine Cooperative Study #17 participants include E. Allen (Montrose, NY); S. Chang (Brockton, MA); J. Crayton (Hines, IL); R. Douyon (Miami, FL); L. Dunn (Durham, NC); J. Erdos (West Haven, CT); D. Evans (Augusta, GA); J. Grabowski, M.D. (Allen Park, MI); L. Herz (Bedford, MA); G. Jurgus (Brecksville, OH); J. Lauriello (Palo Alto, CA); W. Lawson (Little Rock, AK); M. Peszek (Perry Point, MD); J. Peters (Pittsburgh, PA); and J. Tekell (San Antonio, TX). Pharmacy Coordinating Center: C. Fye (Albuquerque, NM).

**The Authors**

Joyce Cramer, B.S., is Project Director; Robert Rosenheck, M.D., is Director, NorthEast Program Evaluation Center; and Dennis Charney, M.D., was Chief, Psychiatry Service (at the time this article was written), for the VA Connecticut Healthcare System, West Haven, CT, and Yale University School of Medicine, Department of Psychiatry, New Haven, CT. Weichun Xu, Ph.D., was Biostatistician (at the time this article was written); Jonathan Thomas, M.S., was Health Scientist (at the time this article was written); and William Henderson, Ph.D., is Chief, Cooperative Studies Program Coordinating Center at the Department of Veterans Affairs Cooperative Studies Program Coordinating Center, Hines, IL.