Validation of a 4-item Negative Symptom Assessment (NSA-4): a short, practical clinical tool for the assessment of negative symptoms in schizophrenia

LARRY ALPHS,¹ ROBERT MORLOCK,¹ CHERYL COON,² PILAR CAZORLA,³ ARMIN SZEGEDI³ & JOHN PANAGIDES³

1 Pfizer Global R&D, Ann Arbor, MI, USA

2 RTI Health Solutions, Research Triangle Park, NC, USA

3 Schering-Plough, Summit, NJ, USA

Key words

negative symptoms,
schizophrenia, NSA-4

Correspondence

Larry Alphs, Ortho-McNeil
Janssen, 1125 Trenton-
Harbourton Road, P.O. Box 200,
Titusville, NJ 08560, USA.
Telephone (+1) 609-730-3100
Email: larryd349@comcast.net

Received 3 April 2009;
revised 7 October 2009;
accepted 4 December 2009

Abstract

The 16-item Negative Symptom Assessment (NSA-16) scale is a validated tool for evaluating negative symptoms of schizophrenia. The psychometric properties and predictive power of a four-item version (NSA-4) were compared with the NSA-16. Baseline data from 561 patients with predominant negative symptoms of schizophrenia who participated in two identically designed clinical trials were evaluated. Ordered logistic regression analysis of ratings using NSA-4 and NSA-16 were compared with ratings using several other standard tools to determine predictive validity and construct validity. Internal consistency and test–retest reliability were also analyzed. NSA-16 and NSA-4 scores were both predictive of scores on the NSA global rating (odds ratio = 0.83–0.86) and the Clinical Global Impressions–Severity scale (odds ratio = 0.91–0.93). NSA-16 and NSA-4 showed high correlation with each other (Pearson $r = 0.85$), similar high correlation with other measures of negative symptoms (demonstrating convergent validity), and lesser correlations with measures of other forms of psychopathology (demonstrating divergent validity). NSA-16 and NSA-4 both showed acceptable internal consistency (Cronbach α , 0.85 and 0.64, respectively) and test–retest reliability (intraclass correlation coefficient, 0.87 and 0.82). This study demonstrates that NSA-4 offers accuracy comparable to the NSA-16 in rating negative symptoms in patients with schizophrenia. Copyright © 2011 John Wiley & Sons, Ltd.

Introduction

The schizophrenic symptoms of poverty of thought, impoverished emotional experience, diminished affect,

decreased motivation toward goal-directed behavior, and decreased social drive (collectively referred to as negative symptoms) account for a significant portion of the morbidity associated with schizophrenia. As Kraepelin

expressed, “The result ... is emotional dullness, failure of mental activities, loss of mastery over volition, of endeavor, and of ability for independent action. The essence of personality is thereby destroyed ...” [Kraepelin E. *Dementia Praecox and Paraphrenia* (1919), translated by R.M. Barclay, edited by G.M. Robertson; reprinted 1971, New York, Robert E. Krieger. Cited by Buchanan (2007)]. Given the impact of these symptoms on functioning in the lives of affected individuals, improved treatments may have significant beneficial effect on long-term outcome in persons with schizophrenia.

The 16-item Negative Symptom Assessment (NSA-16) is a validated instrument that can be used to track the effects of treatment over time (Axelrod *et al.*, 1993). In addition to the individual items, the NSA-16 also includes a global rating in which the clinician estimates the overall degree of impairment specifically related to the patient’s negative symptoms (in comparison with a normal healthy young adult).

The NSA-16 is a sensitive and reliable rating instrument; however, because it requires 15–30 minutes to complete, it may not be practical for use in a busy clinical setting. To meet the needs of the clinician, a shorter version of the NSA-16 has been developed. This four-item assessment tool (the NSA-4) is quick and easy to administer, and captures the essential features of the NSA-16. The NSA-4 comprises items 2, 5, 8, and 13 from the NSA-16 (Figure 1). These four items were selected based on their coverage of previously identified sub-domains of negative symptoms as determined by validation studies and expert opinion (Axelrod *et al.*, 1993). The NSA-4 items were chosen so that one item from each domain on the NSA-16 was selected for inclusion on the NSA-4 to keep from one domain being overrepresented in the short form.

Our hypothesis was that the NSA-4 is an acceptable alternative to the NSA-16. The objective of this study was to evaluate the psychometric properties and predictive power of the NSA-4 by comparison with assessments made using the NSA-16 and other standard rating instruments.

Methods

The NSA-16 was administered at randomization in two identically designed phase III randomized, double-blind, flexible-dose clinical trials completed in two distinct geographic regions. A total of 561 patients from these studies with predominant and persistent negative symptoms of schizophrenia were selected for the validation evaluations. Study A7501013 (clinicaltrials.gov registry number: NCT00145496) was conducted in North and

South America, while Study 25543 (clinicaltrials.gov registry number: NCT00212836) was conducted in Europe, South Africa, and Australia. Additional instruments used in the clinical trials included the Positive and Negative Syndrome Scale (PANSS) (Kay *et al.*, 1987); the Clinical Global Impression–Severity of Illness scale (CGI-S) (Guy, 1976); the Strauss–Carpenter Level of Functioning scale (LOF) (Strauss and Carpenter, 1977); the Calgary Depression Scale for Schizophrenia (CDSS) (Addington *et al.*, 1990); the Personal Evaluations of Transitions in Treatment (PETiT) (Voruganti and Awad, 2002); the Quality of Life, Enjoyment and Satisfaction Questionnaire (Q-LES-Q) (Endicott *et al.*, 1976); the Quality of Life Scale (QLS) (Heinrichs *et al.*, 1984). These instruments (Table 1) provided comparative data used in assessing the psychometric properties of the NSA-4.

Entry requirements for both studies at screening included PANSS negative symptoms Marder factor (Marder *et al.*, 1997) score ≥ 20 , negative symptoms Marder factor score higher than positive symptoms Marder factor score, symptoms stable for five months, and low scores on rating scales for extrapyramidal symptoms and depression. By design, the studies from which this validation sample was drawn required that all patients be clinically stable for at least five months by history and that this be demonstrated prospectively by a 30-day period of observation prior to randomization into the main study. Only those patients who were prospectively demonstrated to be clinically stable and completed both points of evaluation were included in this validation sample.

Predictive validity was assessed using the proportional odds model to determine the ability of the NSA-4 and NSA-16 to predict overall severity as assessed by the NSA global rating and CGI-S scores, both rated on seven-point scales. This analysis was considered the most important test of the NSA-4 validity. If the NSA-4 and the NSA-16 showed similar ability to predict the levels of the NSA global rating and CGI-S, then the NSA-4 could be considered sufficiently similar to be substituted for the NSA-16 in a setting where a more rapid assessment of negative symptoms is necessary.

Construct validity describes the relationships among multiple indicators of a construct and the degree to which they follow predictable patterns. The baseline assessment provides data for the evaluation of the construct validity of the NSA-4. Correlational analyses were conducted to examine the construct validity of the NSA-4.

Correlations between the NSA-4 and the other available measures were computed (all correlations were Pearson correlations except those with the NSA global

<div>1. Prolonged time to respond</div> <div>2. Restricted speech quantity</div> <div>1) normal speech quantity</div> <div>2) minimal reduction in quantity, may be extreme of normal</div> <div>3) speech quantity is reduced, but more obtained with minimal prodding</div> <div>4) flow of speech is maintained only by regularly prodding</div> <div>5) responses usually limited to a few words, and/or detail is only obtained by prodding or bribing</div> <div>6) responses usually non-verbal or limited to one or two words, despite efforts to elicit more</div> <div>9) not ratable</div> <div>3. Impoverished speech content</div> <div>4. Inarticulate speech</div> <div>5. Emotion: Reduced range (specify time frame for this assessment)</div> <div>1) normal range of emotion</div> <div>2) minimal reduction in range, may be extreme of normal</div> <div>3) range seems restricted relative to a normal person, but during the specified time period subject convincingly reports at least four emotions</div> <div>4) subject convincingly identifies two or three emotional experiences</div> <div>5) subject can convincingly identify only one emotional experience</div> <div>6) subject reports little or no emotional range</div> <div>9) not ratable</div> <div>6. Affect: Reduced modulation of intensity</div> <div>7. Affect: Reduced display on demand</div>	<div>8. Reduced social drive</div> <div>1) normal social drive</div> <div>2) minimal reduction in social drive, may be extreme of normal</div> <div>3) desire for social interactions seems somewhat reduced</div> <div>4) obvious reduction in desire to initiate social contacts, but a number of social contacts are initiated each week</div> <div>5) marked reduction in desire to initiate social contacts, but a few contacts are maintained at subject's initiation (as with family)</div> <div>6) no desire to initiate any social interactions</div> <div>9) not ratable</div> <div>9. Poor rapport with interviewer</div> <div>10. Sexual interest</div> <div>11. Poor grooming and hygiene</div> <div>12. Reduced sense of purpose</div> <div>13. Reduced interests</div> <div>1) normal interests</div> <div>2) minimal reduction in interests, may be extreme of normal</div> <div>3) range of interests and/or commitment to them seems diminished</div> <div>4) range of interests is clearly diminished and subject is not particularly committed to interests held</div> <div>5) only one or two interests reported, and these pursued superficially</div> <div>6) little or nothing stimulates interest</div> <div>9) not ratable</div> <div>14. Reduced daily activity</div> <div>15. Reduced expressive gestures</div> <div>16. Slowed movements</div> <div>GLOBAL NEGATIVE SYMPTOMS RATING</div> <div>1. No evidence of this symptom</div> <div>2. Minimal evidence of this symptom</div> <div>3. Mild evidence of this symptom</div> <div>4. Moderate evidence of this symptom, apparent to the casual observer</div> <div>5. Marked evidence of this symptom, readily apparent to casual observer</div> <div>6. Severe, not only obvious but has marked impact on functioning</div> <div>7. Extremely severe symptom, it is incapacitating for subject</div>
---	--

Figure 1 The NSA-16, highlighting items 2, 5, 8, and 13 (used in the NSA-4), and the NSA global rating.

rating, which were polyserial correlations that are appropriate when one variable is continuous and the other variable uses an ordinal scale to capture a continuous underlying variable). Convergent validity would be demonstrated by relatively high correlations ($r > 0.50$) between NSA-4 and closely related rating instruments (NSA global rating, the PANSS negative subscale and negative symptom Marder factor, and the LOF) (Cohen, 1992). Conversely, divergent validity would be demonstrated by lower correlations ($r < 0.30$) between NSA-4 and dissimilar instruments (CDSS, PANSS Marder factors for anxiety/depression, disorganized thought, hostility/

excitement, and positive symptoms, PETiT, Q-LES-Q subscales for leisure time activities and social relations, and QLS subscales for intrapsychic foundations, common objects and activities, interpersonal relations, and instrumental role) (Cohen, 1992). Inferences regarding convergent and divergent validity emphasize the patterns of correlations among the measures.

The internal consistency of the NSA was evaluated by computing Cronbach α (Cronbach, 1951). The optimal range for α is 0.70–0.90 (Streiner and Norman, 1995), indicating a set of items that is strongly related and capable of supporting a unidimensional scoring structure

Table 1 Assessment instruments used

Instrument	Description	Comments
PANSS (Positive and Negative Syndrome Scale)	30-item clinician-rated instrument. Subscales rate positive symptoms, negative symptoms, and general psychopathology. Total score = sum of subscale scores.	The PANSS rating scales are <i>not</i> true Likert scales, because each point on each item is described specifically.
PANSS Marder factors	Factor analysis classifying data from the 30 PANSS items into five symptom domains: positive symptoms, negative symptoms, disorganized thought, hostility/excitement, and anxiety/depression.	Seven items in the negative symptoms Marder factor include five from the PANSS negative subscale and two from the general psychopathology subscale.
CGI-S (Clinical Global Impression–Severity of Illness)	Single-item clinician-rated scale for assessing the global severity of schizophrenia at a particular time.	Clinical improvement is suggested by reduction in CGI-S score but formally rated on CGI-Improvement scale.
LOF (Level of Functioning)	Nine-item clinician-rated scale. Subscales assess social contacts, work, symptoms, and general function. Higher scores indicate better functioning.	
CDSS (Calgary Depression Scale for Schizophrenia)	Nine-item clinician-rated scale for assessing depressive symptoms in schizophrenia. Higher scores indicate more severe symptoms.	
PETiT (Personal Evaluations of Transitions in Treatment)	30-item self-report questionnaire assessing subjective experience with treatment. Higher scores indicate better treatment outcomes.	
Q-LES-Q (Quality of Life, Enjoyment and Satisfaction Questionnaire)	Patient-reported questionnaire that measures perceived satisfaction with various life domains within the past week. Higher scores indicate greater satisfaction.	For this study, only the Leisure Time Activities and Social Relations subscales were administered.
QLS (Quality of Life Scale)	21-item clinician-rated scale for assessing psychosocial functioning over the past four weeks. Higher scores indicate better functioning.	

without redundancy. However, given that each of the items of the NSA-4 represents a unique subdomain of the NSA-16 that had previously been demonstrated to load on distinct factors, and knowing that α typically decreases as items are removed, α was expected to be lower for the NSA-4 than it was for the NSA-16.

We assessed the test–retest reliability of the NSA by computing intraclass correlation coefficients (ICCs) using data from two time points – screening and randomization (visits one and two, respectively). The 30-day observation window between these visits was defined by the protocol from which the sample was derived and was a period during which the underlying condition was to be demonstrated to be stable. ICC is the ratio of true variance among scores to the sum of true variance plus random

error variance; it indicates the proportion of the observed variance that is due to true differences across the separate assessments (i.e. a change in rating upon repeat assessment). We used a two-way (subject $\times$ time) random effects analysis of variance to compute ICC (Schuck, 2004). It is generally recommended that ICC should be at least 0.70 for multiple-item scales (Nunnally and Bernstein, 1994).

Results

Table 2 presents the demographic and baseline clinical characteristics of the study sample. The only substantial difference between the two clinical trial samples is on the distribution of ethnicity, which is not surprising given that study A7501013 was conducted in North and South

Table 2 Demographics and baseline clinical evaluations

	Overall	Protocol 25543	Protocol A7501013
<i>N</i>	561	274	287
Mean age at baseline, years (standard deviation, SD)	42.0 (11.7)	40.6 (11.8)	43.4 (11.6)
Sex, <i>n</i> (%)			
Men	403 (71.8)	195 (71.2)	208 (72.5)
Women	158 (28.2)	79 (28.8)	79 (27.5)
Race, <i>n</i> (%)			
White	359 (64.0)	235 (85.8)	124 (43.2)
Black	146 (26.0)	26 (9.5)	120 (41.8)
Asian	4 (0.7)	0 (0.0)	4 (1.4)
Other or not recorded	52 (9.3)	13 (4.7)	39 (13.6)
<i>Mean (SD) clinical ratings at baseline</i>			
NSA-16 score	62.8 (9.8)	61.9 (9.0)	63.6 (10.5)
NSA-4 score	69.7 (10.8)	68.4 (10.3)	70.9 (11.2)
CDSS score	2.3 (2.5)	2.3 (2.8)	2.5 (2.7)
CGI-S score	4.3 (0.8)	4.4 (0.8)	4.2 (0.8)
LOF score	17.3 (5.1)	17.1 (5.1)	17.6 (5.0)
PANSS scores			
Negative subscale	26.9 (4.3)	26.9 (4.0)	26.9 (4.6)
Negative symptoms Marder factor	26.8 (4.2)	26.7 (4.2)	26.8 (4.2)
Anxiety/depression Marder factor	7.8 (3.1)	7.4 (3.0)	8.1 (3.1)
Disorganized thought Marder factor	17.3 (4.6)	17.3 (4.6)	17.4 (4.7)
Hostility/excitement Marder factor	5.5 (1.9)	5.3 (1.7)	5.7 (2.1)
Positive symptoms Marder factor	17.6 (4.4)	16.3 (4.3)	18.8 (4.2)
PETIT score	35.0 (9.8)	33.8 (9.8)	36.0 (9.8)
Q-LES-Q scores			
Leisure time activities	53.6 (21.9)	50.6 (22.3)	56.3 (21.2)
Social relations	47.9 (19.3)	47.6 (18.6)	48.2 (19.9)
QLS scores			
Intrapsychic foundations	17.1 (6.3)	17.6 (6.0)	16.7 (6.6)
Common objects and activities	5.7 (2.4)	5.9 (2.4)	5.4 (2.3)
Interpersonal relations	16.7 (7.1)	16.8 (6.8)	16.6 (7.5)
Instrumental role	9.0 (5.4)	12.4 (2.8)	5.7 (5.3)

Table 3 Predictive validity: odds ratios (99% confidence intervals) for NSA-4 and NSA-16 total score predicting NSA global rating and CGI-S

Scale	NSA-16	NSA-4
NSA global rating	0.83 (0.81, 0.86)	0.86 (0.83, 0.89)
CGI-S	0.91 (0.89, 0.93)	0.93 (0.91, 0.95)

America, whereas study 25543 was conducted in Europe, South Africa, and Australia.

Predictive validity

The NSA-4 and NSA-16 scores were both highly predictive of scores on the NSA global rating and CGI-S, and the

NSA-16 and NSA-4 are comparable in terms of their predictive ability, as seen in the overlapping confidence intervals (CIs) (Table 3).

Construct validity

The NSA-4 showed a high correlation with NSA-16 (Pearson $r = 0.85$; 99% CI, 0.82 to 0.88; $p < 0.0001$) and there was a high degree of overlap between the two scales (72% shared variance).

Convergent and divergent validity

Convergent validity was demonstrated for NSA-4 with high correlations (Pearson $r > 0.5$, $p < 0.0001$) seen between NSA-4 and related instruments (Table 4). Divergent validity was demonstrated for the NSA-4 and

Table 4 Convergent validity: correlations (99% confidence intervals) for NSA-4 and NSA-16 total score compared with closely related measures

Scale	NSA-16	NSA-4
NSA global rating	0.70 (0.64, 0.77)	0.68 (0.62, 0.75)
PANSS negative subscale	0.59 (0.51, 0.66)	0.52 (0.44, 0.60)
PANSS negative symptoms Marder factor	0.63 (0.56, 0.69)	0.57 (0.49, 0.64)
LOF	−0.51 (−0.59, −0.42)	−0.47 (−0.55, −0.38)

Table 5 Divergent validity: correlations (99% confidence intervals) for NSA-4 and NSA-16 total score compared with less related measures

Scale	NSA-16	NSA-4
CDSS	−0.10 (−0.21, 0.01)	−0.11 (−0.22, 0.00)
PANSS anxiety/depression Marder factor	−0.03 (−0.14, 0.08)	−0.06 (−0.16, 0.05)
PANSS disorganized thought Marder factor	0.42 (0.33, 0.51)	0.29 (0.18, 0.38)
PANSS hostility/excitement Marder factor	0.06 (−0.05, 0.17)	0.03 (−0.08, 0.13)
PANSS positive symptoms Marder factor	0.23 (0.13, 0.33)	0.13 (0.02, 0.23)
PETiT total	−0.15 (−0.26, −0.03)	−0.13 (−0.24, −0.02)
Q-LES-Q leisure time activities	−0.20 (−0.31, −0.09)	−0.19 (−0.29, −0.08)
Q-LES-Q social relations	−0.29 (−0.39, −0.18)	−0.32 (−0.41, −0.21)
QLS intrapsychic foundations	−0.48 (−0.56, −0.39)	−0.44 (−0.53, −0.35)
QLS common objects and activities	−0.34 (−0.44, −0.24)	−0.29 (−0.38, −0.18)
QLS interpersonal relations	−0.39 (−0.48, −0.29)	−0.41 (−0.50, −0.32)
QLS instrumental role	−0.24 (−0.34, −0.14)	−0.19 (−0.30, −0.09)

NSA-16 by less robust correlations (Pearson $r < 0.5$, $p = 0.05$ – 0.0001) with dissimilar instruments (Table 5).

Internal consistency

Internal consistency was demonstrated for both NSA instruments (Cronbach α : 0.85 for NSA-16 and 0.64 for NSA-4).

Reliability

Acceptable test–retest reliability was demonstrated for both NSA instruments (ICC: 0.87 for NSA-16 and 0.82 for NSA-4).

Discussion

The NSA-16 is a valid and reliable instrument, but widespread use by clinicians may be limited by the time required to complete the full 16-item assessment. A derivative four-item version of the NSA-16 was developed to produce a more concise rating scale without sacrificing the overall comprehensiveness and

accuracy of constituent domains provided by the original instrument.

This series of psychometric assessments described here demonstrate that the NSA-4 can gauge the presence and severity of negative symptoms in schizophrenia with accuracy and breadth of coverage comparable to that of the NSA-16, as seen in the significant proportional odds models predicting the NSA global rating and CGI-S rating. Therefore, the NSA-4 may be a useful substitute for the NSA-16 when rapid assessment of negative symptoms is needed. For both NSA-4 and NSA-16, correlations were generally more robust with other measures of negative symptoms (convergent validity) than with dissimilar instruments (divergent validity).

On the measure of internal consistency, the Cronbach α value for the NSA-4 was lower than usually recommended. This likely reflects the fact that each NSA-4 item represents a single item sample from the major factors that have been identified from the NSA-16, so that consistency among the set of four distinct items is unlikely to be as consistent as the set of overlapping 16 items. Thus, it is expected that α would be lower for

NSA-4, with a smaller number of items measuring multiple domains, than for NSA-16. For this reason, although the NSA-4 might be useful as a quick assessment of negative symptoms, it would not be recommended for specific research evaluations of negative symptoms.

Finally, there was a clear similarity of the test–retest reliability measures for the NSA-4 and NSA-16. These values suggest good replicability across assessments, especially considering that the test–retest window was substantial (25–30 days).

From these psychometric assessments, we conclude that the NSA-4 is an acceptable alternative to the NSA-16 for certain settings; the fact that it is quicker and easier to

use than the NSA-16 makes the NSA-4 especially attractive for use in the clinical setting.

Acknowledgments

Editorial support was provided by Complete Healthcare Communications, Inc., with funding from Schering-Plough.

Declaration of interest statement

Competing interests: Larry Alphs is the owner of the copyright to the NSA-4. The other authors have no competing interests.

References

- Addington D., Addington J., Schissel B. (1990) A depression rating scale for schizophrenics. *Schizophrenia Research*, **3**(4), 247–251.
- Axelrod B.N., Goldman R.S., Alphs L.D. (1993) Validation of the 16-item Negative Symptom Assessment. *Journal of Psychiatric Research*, **27**(3), 253–258.
- Buchanan R.W. (2007) Persistent negative symptoms in schizophrenia: An overview. *Schizophrenia Bulletin*, **33**(4), 1013–1022.
- Cohen J. (1992) A power primer. *Psychological Bulletin*, **112**(1), 155–159.
- Cronbach J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297–334.
- Endicott J., Spitzer R.L., Fleiss J.L., Cohen J. (1976) The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, **33**(6), 766–771.
- Guy W. (1976) Clinical global impressions. In Guy W. (ed.) *ECDEU Assessment Manual for Psychopharmacology*, pp. 217–222, Washington, DC, US Department of Health, Education, and Welfare.
- Heinrichs D.W., Hanlon T.E., Carpenter W.T. Jr (1984) The Quality of Life Scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin*, **10**(3), 388–398.
- Kay S.R., Fiszbein A., Opler L.A. (1987) The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, **13**(2), 261–276.
- Marder S.R., Davis J.M., Chouinard G. (1997) The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: Combined results of the North American trials. *Journal of Clinical Psychiatry*, **58**, 538–546.
- Nunnally J.C., Bernstein I.H. (1994) *Psychometric Theory*, New York, McGraw-Hill.
- Schuck P. (2004) Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Quality of Life Research*, **13**, 571–586.
- Strauss J.S., Carpenter W.T. Jr (1977) Prediction of outcome in schizophrenia. III. Five-year outcome and its predictors. *Archives of General Psychiatry*, **34**, 159–163.
- Streiner D.L., Norman G.R. (1995) *Health Measurement Scales: A Practical Guide to Their Development and Use*, New York, Oxford University Press.
- Voruganti L.N., Awad A.G. (2002) Personal evaluation of transitions in treatment (PETIT): A scale to measure subjective aspects of antipsychotic drug therapy in schizophrenia. *Schizophrenia Research*, **56**, 37–46.