

The Neuropsychiatric Inventory-Clinician rating scale (NPI-C): reliability and validity of a revised assessment of neuropsychiatric symptoms in dementia

K. de Medeiros,^{1,2} P. Robert,³ S. Gauthier,⁴ F. Stella,⁵ A. Politis,⁶ J. Leoutsakos,^{2,15} F. Taragano,⁷ J. Kremer,⁸ A. Brugnolo,⁹ A. P. Porsteinsson,¹⁰ Y. E. Geda,¹¹ H. Brodaty,¹² G. Gazdag,¹³ J. Cummings¹⁴ and C. Lyketsos^{2,15} for the NPI-C Research Group

¹The Copper Ridge Institute, Sykesville, Maryland, U.S.A.

²Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore, Maryland, U.S.A.

³Centre Mémoire de Ressources et de Recherche – CHU – Nice–Sophia Antipolis University, France

⁴McGill Center for Studies in Aging, Montreal, Canada

⁵UNESP – São Paulo State University, Biosciences Institute, Rio Claro, SP, Brazil

⁶Division of Geriatric Psychiatry, Egnition Hospital, Athens Hospital, Athens, Greece

⁷CEMIC, University Hospital Dementia Research Unit, Buenos Aires, Argentina

⁸Institute Privado Kremer, Cordoba, Argentina

⁹Department of Neurosciences, DINOG, Genoa, Italy

¹⁰AD-CARE, University of Rochester School of Medicine and Dentistry, Rochester, New York, U.S.A.

¹¹Mayo Clinic, Rochester, Minnesota, U.S.A.

¹²Dementia Collaborative Research Centre, University of South Wales, Sydney, Australia

¹³1st Department of Psychiatry and Psychiatric Rehabilitation, Jahn Ferenc Hospital, Budapest, Hungary

¹⁴Department of Neurology, University of California, Los Angeles, California, U.S.A.

¹⁵Department of Psychiatry, Johns Hopkins Bayview, Baltimore, Maryland, U.S.A.

ABSTRACT

Background: Neuropsychiatric symptoms (NPS) affect almost all patients with dementia and are a major focus of study and treatment. Accurate assessment of NPS through valid, sensitive and reliable measures is crucial. Although current NPS measures have many strengths, they also have some limitations (e.g. acquisition of data is limited to informants or caregivers as respondents, limited depth of items specific to moderate dementia). Therefore, we developed a revised version of the NPI, known as the NPI-C. The NPI-C includes expanded domains and items, and a clinician-rating methodology. This study evaluated the reliability and convergent validity of the NPI-C at ten international sites (seven languages).

Methods: Face validity for 78 new items was obtained through a Delphi panel. A total of 128 dyads (caregivers/patients) from three severity categories of dementia (mild = 58, moderate = 49, severe = 21) were interviewed separately by two trained raters using two rating methods: the original NPI interview and a clinician-rated method. Rater 1 also administered four additional, established measures: the Apathy Evaluation Scale, the Brief Psychiatric Rating Scale, the Cohen-Mansfield Agitation Index, and the Cornell Scale for Depression in Dementia. Intraclass correlations were used to determine inter-rater reliability. Pearson correlations between the four relevant NPI-C domains and their corresponding outside measures were used for convergent validity.

Results: Inter-rater reliability was strong for most items. Convergent validity was moderate (apathy and agitation) to strong (hallucinations and delusions; agitation and aberrant vocalization; and depression) for clinician ratings in NPI-C domains.

Conclusion: Overall, the NPI-C shows promise as a versatile tool which can accurately measure NPS and which uses a uniform scale system to facilitate data comparisons across studies.

Key words: dementia, Alzheimer's disease, neuropsychiatric symptoms, depression, apathy, agitation, neuropsychiatric inventory

Introduction

Neuropsychiatric symptoms (NPS) affect almost everyone with dementia (Lyketsos *et al.*, 1999; 2000; Schreiner *et al.*, 2005; Steinberg *et al.*, 2006). Although patients with dementia may have

Correspondence should be addressed to: Kate de Medeiros, The Copper Ridge Institute, 710 Obrecht Road, Sykesville, Maryland, 21784, U.S.A. Phone: +1 (410) 552-3201; Fax: +1 (410) 552-0344. Email: kdemed1@ummc.edu. Received 11 March 2010; revision requested 2 April 2010; revised version received 5 April 2010; accepted 8 April 2010. First published online 1 July 2010.

NPS at any disease stage, and the prevalence of NPS may increase with time (Lopez *et al.*, 2003), certain NPS are more common in mild dementia (e.g. depression, anxiety, irritability, and apathy) (Feldman *et al.*, 2004) while others are more frequent in later states (aberrant vocalizations, delusions, hallucinations and disinhibition) (Lopez *et al.*, 2003; Mayer *et al.*, 2006). Given their frequent occurrence, NPS are a major focus of study and treatment (Tariot *et al.*, 1995; Steinberg *et al.*, 2004; 2006; Sink *et al.*, 2005; Lyketsos, 2007; Gauthier *et al.*, 2010). NPS measurement is therefore a crucial aspect of dementia research, especially in clinical trials where change (or lack thereof) are key indicators of treatment effect (Lyketsos, 2007; Ballard *et al.*, 2008).

The accurate assessment of NPS through valid, sensitive, and reliable measures is key to interpreting results from individual investigations. The use of uniform measures is important to the field's ability to compare and subsequently interpret results across studies. Multiple NPS measures have been developed, each with strengths and weaknesses, with different measures being used in different studies. These differ in what they measure (i.e. symptoms within one particular NPS domain or several domains); scale properties such as the targeted informant (e.g. caregiver, clinician, patient); and rating approach (e.g. rating frequency alone, severity, other), all of which can complicate interpretation of findings, especially in clinical trials targeting specific NPS (Perrault *et al.*, 2000). For example, Sink and colleagues (2005), in their meta-analysis of NPS-related randomized controlled trials, reported conflicting results among several trials of cholinesterase inhibitors and memantine attributable to use of several different scales and subscales. In addition, such differences can also increase the chance of Type I errors. A uniform measure that can be used in a range of different studies involving NPS is needed.

Commonly used NPS measures include the Alzheimer's Disease Assessment Scale non-cognitive section (ADAS-noncog; Rosen *et al.*, 1984) and the Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD; Reisberg, *et al.*, 1987). The ADAS-noncog, one of the first measures developed for this purpose, consists of ten items rated for frequency by a knowledgeable informant. Although good reliability and validity have been reported (Weyer *et al.*, 1997), the ADAS-noncog does not rate some common NPS such as aggressiveness and anxiety and includes other items not related to emotional behavior (e.g. tremors). The BEHAVE-AD is a 25-item questionnaire developed for pharmaceutical trials (Reisberg *et al.*,

1996) in which a knowledgeable informant provides a severity rating for each item over the past two weeks. As with the ADAS-noncog, reliability and validity have been established. However, the scale's brevity provides only limited insight into NPS occurrence and may not be effective in monitoring behavior change (Perrault *et al.*, 2000).

For about a decade, the Neuropsychiatric Inventory (NPI; Cummings *et al.*, 1994) and its slightly modified version the Neuropsychiatric Inventory Questionnaire (NPIQ; Kaufer *et al.*, 2000) have been widely accepted as the standard measure of NPS in clinical trials for "anti-dementia therapies" and in several population-based studies in dementia and related disorders (Lyketsos *et al.*, 2002; Geda *et al.*, 2008). It is a brief and quick-to-administer questionnaire composed of a screening question and seven to nine items for each of 12 domains. A knowledgeable informant (usually a caregiver) indicates via the screening question whether the patient has experienced any domain-related NPS in the past month. If the screening question is validated, the caregiver is then asked whether each item within the domain has occurred in the past month and provides a global rating of frequency, severity and caregiver distress for all items in the domain at the same time (not item by item).

The NPI has established several important methodological approaches to NPS measurement. These include: (1) items that are behavior-based and observable, which facilitates report of frequency and severity of symptoms by a knowledgeable informant; (2) items that are grouped into domains with a screening question which enables quick completion and interpretation of results; (3) items that are specific to populations with dementia unlike other assessments that may apply to a nursing home population or to adults in general; and (4) standard ratings of domain frequency, severity and caregiver distress, unlike other assessments which may capture frequency or severity but not both.

Despite its many strengths, however, the NPI has certain weaknesses: (1) data are acquired from informants not directly from patients. As a result, neuropsychiatric data obtained from caregivers are susceptible to recall bias. Caregiver reports may be influenced by caregiver mood (e.g. he or she may be depressed), cultural beliefs (e.g. caregiver's views of how he/she should appropriately respond or what is "normal" for older people to experience), denial (caregiver's minimization of symptoms) and/or the caregiver's education; (2) unknown reliability of ratings for individual items versus global domain ratings; (3) few items are specific to severe or mild dementia; (4) limited depth of items in specific domains such that individual domains cannot be used as a stand-alone measure in studies targeting

individual NPS (e.g. depression, agitation); and (5) limited sensitivity to change when compared to measures that incorporate clinician judgment (Mayer *et al.*, 2006).

To address limitations of the NPI and other measures of NPS, we have developed a revised and expanded version of the NPI, known as the NPI-C (“clinician rated”), which includes additional domains and individual items within various domains, and a clinician rating methodology. Here, we report on the development of the NPI-C and the results from an international validation study in which we examined inter-rater reliability and convergent validity for the NPI-C domains of apathy, delusions and hallucinations, agitation, aberrant vocalization and depression. We hypothesized a strong correlation between raters (intraclass correlations: ICC > 0.50) and moderate to strong convergent validity ($r \geq 0.40$) between the NPI-C domain and a validated outside measure. We anticipate that the NPI-C will provide a single versatile method of rating for NPS that can be used in a range of clinical studies as either a “broad spectrum” measure of several types of NPS or as a domain-specific measure when a study calls for in-depth monitoring of a limited set of domains, such as apathy or agitation.

Methods

Development of the NPI-C

APPROACH

The NPI-C was developed and validated by an international collaborative group of researchers in this area, led out of The Johns Hopkins University by two of the authors (KdM and CGL), working closely with the original developer of the NPI (JC). The foundation of the new instrument is the original NPI, whose items and domains have been expanded as described below, and whose rating approach has been modified to include a clinician rating methodology. Additionally, several translations have been developed to make the new instrument available for international studies. Translated versions were developed using the language-specific validated NPI translation for original NPI items and translation/back translation for newly added NPI-C items. The NPI-C languages include English, French, Greek, Hungarian, Italian, Portuguese and Spanish.

EXPANSION OF DOMAINS AND INDIVIDUAL ITEMS

Expansion of the domains was based on comprehensive review of 19 existing NPS measures.

A master table of all symptoms was constructed and items were sorted based on existing NPI domains or on face similarities (e.g. vocalization) if an applicable domain did not exist. Items not corresponding to observable NPS and duplicate items were removed. A total of 78 new items were added to the nine domains of the NPI (see Table 1). No additional items were added to three NPI domains: hallucinations, delusions and elation. The face validity for the new structure and items was obtained through a Delphi panel of eight experts in dementia research (CL, FT, JK, PR, SG, HB, FS and Jacobo Mintzer, Professor of Neurosciences and Psychiatry at the Medical University of South Carolina) through an in-person meeting and email correspondence. (For the list of items, see Appendix A2 published as supplementary material online attached to the electronic version of this paper at <http://www.journals.cambridge.org/ipg>.) For the apathy domain, new items were chosen to fit as well as possible with the new diagnostic criteria for apathy reported by Robert and colleagues (2009).

We made two changes to the existing NPI domains based on prevalence of specific NPS. The reported prevalence rates for agitation, measured separately from aggression, range from 28% to 53% (Brodsky *et al.*, 2001). “Aggression” alone has been estimated to occur in 15%–20% of people with dementia (Lyketsos *et al.*, 1999). Consequently, the NPI-C separates the NPI domain of “Agitation/aggression” into two separate domains, with new items added to each.

The domain “Aberrant vocalization” was added to capture symptoms most evident in advanced dementia. Although the prevalence of disruptive vocalization in long-term care residents ranges from 11% to 31% (Dwyer and Byrne, 2000), only one symptom related to vocalization is present in the NPI. Eleven items representing various types of vocal behavior were included in this domain after review and approval by the Delphi panel.

RATING METHOD

There are two important rating changes in the NPI-C. The first is item-by-item scoring. In the NPI, a knowledgeable informant provides a global domain rating for frequency, severity, and level of caregiver distress related to the group of items within a given domain. The domain score is the product of the ratings for domain frequency and severity. In the NPI-C, ratings for frequency, severity and distress are provided for *each item* and summed to create a total domain score. Item scores are better suited to clinical trials where assessment of change with finer detail is needed.

Table 1. Comparison of number of items per domain for the NPI, the initial NPI-C, and NPI-C items with ICC>0.50 and the percentage of items indicated per domain (total domain score >0)

ORIGINAL NPI			NPI-C		
DOMAIN	TOTAL NUMBER OF ITEMS	% ITEMS WITH SCORE >0	DOMAIN	INITIAL NUMBER OF ITEMS	FINAL NUMBER OF ITEMS ¹ % FINAL ITEMS WITH SCORE >0
Delusions	8	28%	Delusions ²	8	8 28%
Hallucinations	7	13%	Hallucinations ²	7	7 13%
Agitation/aggression	6	66%	Agitation	16	13 68%
			Aggression	14	8 25%
Depression/dysphoria	7	49%	Dysphoria	13	13 70%
Anxiety	6	48%	Anxiety	15	14 75%
Elation/euphoria	6	12%	Elation/euphoria ²	6	6 14%
Apathy/Indifference	9	65%	Apathy	12	12 79%
Disinhibition	6	26%	Disinhibition	17	16 42%
Irritability/lability	6	41%	Irritability/lability	14	12 58%
Aberrant motor behavior	6	36%	Motor disturbances	9	9 43%
Sleep	7	27%	Sleep disorders*	8	8 34%
Appetite and eating disorders	7	38%	Appetite and eating disorders	9	8 43%
—	—	—	Aberrant vocalization	11	8 25%
TOTAL	81			159	142

¹Final NPI-C items comprise initial NPI-C items with ICC>0.50.

²No additional items were added to the original NPI items, although the clinician rating method was used.

CLINICIAN RATING METHOD

The second important rating change is the use of ratings based on expert clinical judgment using a “LEAD” standard (longitudinal data, expert rater, all data) to make severity ratings for individual items (Wilberg *et al.*, 2000). In the clinician rating approach, the knowledgeable informant is first asked to provide frequency, severity and distress ratings for items as described above. Separately, the rater also interviews the patient. The importance of involving the patient in ratings of NPS in individuals with mild or moderate dementia cannot be over-emphasized. Even if the patient lacks the ability or insight to describe experiences of NPS accurately, the interview gives the clinician rater an opportunity to compare the knowledgeable informant’s insights to the patient’s perceptions.

To rate various item responses, the clinician can ask for additional details during either interview and may consult other sources of information such as the patient’s chart or other caregivers familiar with the case in order to provide an overall severity rating for a given domain item. Such an approach has been successfully implemented in large multi-center clinical trials to rate depression in dementia using the Cornell Scale for Depression in Dementia (CSDD) where clinician rated CSDD demonstrated sensitivity to change in depressive symptoms that were superior to the NPI (Mayer

et al., 2006). Clinical judgment also becomes important in distinguishing domains which may seem similar to the non-clinician, such as apathy and depression, but which can be distinguished by a knowledgeable clinician (Starkstein, 2000). The clinician rating approach reduces bias from knowledgeable informant interviews in which their own experience with depression, cognitive decline, or other factors may affect their ability to report symptoms accurately in the patient.

TRAINING TO MAKE CLINICAL RATINGS

As part of the development of the NPI-C, the authors developed a training methodology for clinicians rating this measure. An instructional DVD which features complete NPI-C interviews with dyads and an accompanying workbook were used to illustrate how to obtain additional information from informants and patients effectively, and presented strategies to reconcile conflicting information in order to provide an overall severity rating for each item. A completed NPI-C worksheet based on the DVD interviews allowed the trainee to see the actual scores produced from the interview. A brief discussion with the training rater also provided detailed explanations regarding rating choices.

NPI-C validation study

Ten sites from eight countries participated in the cross-sectional validation study. Sites were located in Argentina, Brazil, Canada, France, Greece, Hungary, Italy and the U.S.A. (three sites). A total of 128 pairs of participants (knowledgeable informants/patients) in three categories of dementia severity (mild, moderate and severe) were recruited. The majority of patients (79.5%) lived at home with a spouse or child. The remaining resided in various care institutions (e.g. assisted living, nursing homes, residential hospitals.) In addition, most knowledgeable informants were family members (46.4% spouses, 35.7% children, and 8.9% "other" relative). Only 9% were professional caregivers. All patients met criteria for probable Alzheimer's disease (McKhann *et al.*, 1984). Inclusion criteria for knowledgeable informant participants were: ability to comment accurately on NPS in the patient over the past month, and verbal contact with the patient at least three times per week during the past three months. Inclusion criteria for patients were: presence of a knowledgeable informant and a medical diagnosis of probable Alzheimer's disease. All sites obtained approval from their respective ethics review boards prior to the start of the study. Written informed consent was obtained from all participants.

Dementia severity was determined by the Mini-mental State Examination (MMSE; Folstein *et al.*, 1975) and the Global Deterioration Scale (GDS; Reisberg *et al.*, 1982), both of which were used to rate all patients. Mild dementia was defined by an MMSE 18–26 and GDS 0–3, moderate dementia by MMSE 10–17, GDS 4–5, and severe dementia by MMSE < 10, GDS $\geq$ 6, which correspond to the commonly accepted MMSE range in clinical trials (Lopez *et al.*, 2003; Perneczky *et al.*, 2006). In cases where either MMSE or GDS score did not fall within the dementia severity range, the rating clinician determined the appropriate severity category.

RATER CHARACTERISTICS AND TRAINING

The NPI-C is to be administered by an experienced clinician. Of the 21 raters who participated in the study, there were 11 physicians, 2 research nurses, 7 researchers with a master's degree or higher, and 1 clinical social worker. Seven sites attended in-person training at Johns Hopkins in September 2008, and were provided with training notebooks and an NPI-C training DVD. The other three sites (Argentina, Hungary and Italy) were trained from a distance through the training notebooks and DVDs that were used at the in-person meeting, along with email support from the lead investigators.

PROCEDURES

To estimate the inter-rater reliability of the NPI-C, and to compare its ratings to those of the NPI, each informant/patient dyad was interviewed at different times (no later than one week apart) by two independent, trained raters. Both raters completed the NPI (Cummings *et al.*, 1994) and the NPI-C. All questions for the NPI and NPI-C were asked regardless of the caregiver's response to the screening question. We note that given the need to ask all NPI-C questions for the purpose of the validation study, administration times would not reflect "real world use" where entire domains could be omitted based on response to the screening question. Administration times were therefore not recorded. When administering the NPI-C, raters first asked caregivers to rate the frequency and severity of each item in each domain. Patients were also interviewed. Although caregivers also rated their distress level for each item indicated, data on distress were used to inform clinician rating and were not part of the study analysis. Clinicians then rated severity (0 to 3) for each item based on caregiver and patient interviews and any additional clinical information. Caregivers were also asked to rate their level of distress for each item indicated. This information was used to inform the overall clinician rating for each item, however, and was not used as a stand-alone rating for this study.

To estimate the convergent validity of specific domains of the NPI-C, one of the two raters also administered four additional measures: the Apathy Evaluation Scale (AES; Marin *et al.*, 1991); the Cohen-Mansfield Agitation Index (CMAI; Cohen-Mansfield *et al.*, 1989); the Cornell Scale for Depression in Dementia (CSDD; Alexopoulos *et al.*, 1988); and the Brief Psychiatric Rating Scale (BPRS; Ventura *et al.*, 1993).

ANALYSIS

Inter-rater reliabilities for clinician ratings per domain were determined by estimating intraclass correlations (ICC) (form (1,1); Shrout and Fleiss, 1979). New NPI-C items (but not original NPI items) with ICCs < 0.5 were removed prior to analyses of convergent validity. Pearson correlations between four NPI-C and NPI domains and their respective validated and established measures of the corresponding NPS were used to determine convergent validity.

Results

128 patient/caregiver dyads completed NPI-C interviews with two raters for a total of 256

Table 2. Mean (SD) and range for selected demographics and scores by dementia severity

	MILD (N = 58)	MODERATE (N = 49)	SEVERE (N = 21)	TOTAL (N = 128)
Age in years	74.3 (7.6) (58 to 90)	77.4 (10.0) (54 to 91)	75.7 (9.6) (55 to 94)	75.7 (9.0) (54 to 94)
% patients living at home	90%	79%	57%	80%
MMSE	23.2 (2.9) (0 to 30)	16 (3.1) (9 to 25)	5.7 (4.1) (0 to 12)	17.6 (7.0) (0 to 28)
GDS	3 (0.8) (1 to 7)	4.7 (0.6) (4 to 6)	5.9 (0.8) (4 to 7)	4.1 (1.3) (1 to 7)
AES*	29.8 (12.1) (0 to 54))	34.3 (8.7) (8 to 49); n = 44	39.9 (11.7) (4 to 51); n = 17	33.1 (11.3) (0 to 51); n = 113
BPRS	27.5 (15.2) (0 to 168)	38.7 (15.9) (7 to 90)	54.0 (22.0) (13 to 98)	36.1 (19.1) (3 to 98)
CMAI	11.8 (10.7) (0 to 174)	15.9 (12.3) (0 to 47)	20.0 (20.1) (0 to 70)	14.7 (13.5) (0 to 70)
CSDD	6.6 (4.6) (0 to 38)	6.7 (5.8) (0 to 28)	8.4 (6.1) (0 to 18)	6.9 (5.3) (0 to 28)

*Data from Brazil (n = 15) have been excluded due to inconsistencies in translation of the AES.

MMSE = Mini-mental State Examination; GDS = Global Depression Scale; AES = Apathy Evaluation Scale; BPRS = Brief Psychiatric Rating Scale; CMAI = Cohen-Mansfield Agitation Index; CSDD = Cornell Scale for Depression in Dementia.

observations. Table 2 includes descriptive statistics of the sample by dementia severity for selected demographic characteristics and scores on the MMSE, GDS, AES, BPRS, CMAI, CSDD. Total cases by dementia severity for the sample were: 58 mild, 49 moderate and 21 severe. Total dyads by language were: 37 English, 16 French, 27 Greek, 5 Italian, 10 Hungarian, 15 Portuguese and 18 Spanish. 83 patients were married, 18 widowed, 10 divorced and 18 never married.

Inter-rater reliabilities based on clinician ratings of each NPI-C item were generally moderate to strong (Table 3). Eight existing NPI and nine newly added NPI-C items (this includes three items in the new domain “aberrant vocalizations”) had ICC values <0.50. For seven items ICCs could not be calculated due to lack of variability. Regardless of ICC value, original NPI items were not removed prior to convergent validity analysis. A total of 17 new items with missing ICCs or ICC <0.50 were removed, leaving a total of 142 NPI-C (61 more items than in the original NPI). Table 1 includes a count of total items per domain for the NPI, the initial NPI-C and the NPI-C after low ICCs were removed. It also includes the total percentage of items per domain for the NPI and NPI-C that were indicated (i.e. had a non-zero rating).

Convergent validity

Table 4 includes Pearson correlations and confidence intervals for four NPI-C domains and

their corresponding outside measure. Correlations for the sum of clinician ratings of NPI-C domains and original NPI domain scores (global rating of frequency $\times$ severity) are also presented. Correlations for all NPI-C domains and corresponding assessments were moderate to strong, with “depression/CSDD” having the strongest correlation ($r = 0.61$). Except for the correlations between “delusions” and “hallucinations” and the psychosis items of BPRS, where there was little difference between the clinician rating (0.60) and the sum of the two global domain caregiver ratings (0.56), the clinician ratings had stronger correlations with the validated measures. This could be attributable to the addition of new items, the clinician rating method, or both.

Of special interest is the difference in strength of correlation between CMAI and the NPI-C agitation domain ($r = 0.40$) and the traditional NPI global rating of agitation items ($r = 0.19$). The correlation strengthened for the NPI and CMAI when the global ratings of “agitation” and “aggression”, which comprise one domain on the NPI, were summed ($r = 0.31$; CI: 0.14 to 0.48). The correlation strengthened considerably for the NPI-C when the global ratings of “agitation” and “aberrant vocalization” were summed ($r = 0.60$). The rationale for this is that the CMAI includes items about verbal aggression. “Aberrant vocalization” was added as a new domain to the NPI-C and is not found in the NPI.

Table 3. Inter-rater reliability ICCs, confidence intervals (95%) and frequencies of scores with > 0 rating for clinician ratings of all items by domain (shaded cells represent NPI items)

A. DELUSIONS			B. HALLUCINATIONS			C. AGITATION			D. AGGRESSION		
ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)
1	0.92	0.89–0.95	1	0.85	0.80–0.91	1	0.69	0.59–0.79	1	0.76	0.68–0.85
2	0.89	0.84–0.93	2	0.80	0.72–0.88	2	0.76	0.68–0.84	2	0.66	0.55–0.77
3	0.97	0.95–0.98	3	0.72	0.61–0.82	3	0.70	0.60–0.80	3	0.76	0.67–0.84
4	0.80	0.72–0.87	4**	0.	0.	4*	0.41	0.26–0.57	4**	0.	0.
5	0.85	0.79–0.91	5*	0.00	0.00–0.22	5	0.74	0.66–0.83	5	0.77	0.69–0.85
6	0.81	0.75–0.88	6*	0.00	0.00–0.22	6	0.65	0.54–0.76	6	0.58	0.44–0.71
7*	0.28	0.09–0.47	7*	0.00	0.00–0.22	7	0.72	0.62–0.81	7**	0.	0.
8	0.84	0.77–0.90				8	0.54	0.41–0.68	8**	0.	0.
						9*	0.33	0.16–0.50	9	0.50	0.35–0.66
						10	0.64	0.52–0.75	10**	0.	0.
						11	0.80	0.73–0.87	11	0.88	0.84–0.93
						12	0.63	0.51–0.74	12*	0.00	0.00–0.20
						13	0.64	0.52–0.75	13	0.90	0.87–0.94
						14*	0.40	0.25–0.57	14**	0.	0.
						15	0.58	0.45–0.71			
						16	0.89	0.85–0.93			
E. DYSPHORIA			F. ANXIETY			G. ELATION			H. APATHY		
ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)
1	0.85	0.80–0.90	1	0.84	0.79–0.90	1	0.81	0.74–0.88	1	0.83	0.78–0.89
2	0.66	0.56–0.76	2	0.87	0.83–0.91	2	0.90	0.86–0.94	2	0.80	0.74–0.86
3	0.83	0.77–0.88	3	0.87	0.82–0.91	3	0.81	0.74–0.88	3	0.78	0.71–0.85
4	0.72	0.63–0.80	4	0.77	0.70–0.84	4	0.87	0.82–0.92	4	0.87	0.82–0.91
5	0.71	0.62–0.80	5	0.85	0.80–0.90	5*	0.0	0.00–0.21	5	0.77	0.70–0.85
6	0.54	0.41–0.67	6	0.85	0.80–0.90	6	0.88	0.84–0.93	6	0.83	0.78–0.89
7	0.89	0.85–0.93	7	0.54	0.41–0.67				7	0.89	0.85–0.93
8	0.86	0.81–0.91	8	0.88	0.84–0.92				8	0.89	0.85–0.93
9	0.83	0.78–0.89	9	0.97	0.96–0.98				9	0.77	0.70–0.84
10	0.75	0.67–0.83	10	0.75	0.67–0.83				10	0.83	0.78–0.89
11	0.76	0.68–0.83	11	0.75	0.67–0.83				11	0.80	0.74–0.87
12	0.80	0.74–0.87	12	0.58	0.46–0.70				12	0.74	0.66–0.82
13	0.67	0.57–0.77	13	0.73	0.65–0.81						
			14*	0.27	0.10–0.44						
			15	0.81	0.74–0.88						
I. DISINHIBITION			J. IRRITABILITY			K. ABERRANT MOTOR			L. SLEEP DISTURBANCES		
ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)
1	0.81	0.74–0.88	1	0.83	0.77–0.90	1	0.91	0.88–0.95	1	0.91	0.88–0.95
2	0.79	0.71–0.86	2	0.78	0.71–0.85	2	0.85	0.79–0.90	2	0.85	0.79–0.90
3	0.80	0.73–0.88	3	0.56	0.43–0.69	3	0.74	0.65–0.83	3	0.74	0.65–0.83
4*	0.26	0.73–0.45	4	0.78	0.71–0.85	4	0.80	0.74–0.87	4	0.80	0.74–0.87
5	0.64	0.52–0.76	5	0.68	0.58–0.78	5	0.64	0.52–0.75	5	0.64	0.52–0.75
6	0.77	0.68–0.85	6	0.87	0.82–0.91	6*	0.50	0.35–0.64	6*	0.50	0.35–0.64
7	0.70	0.60–0.80	7	0.74	0.67–0.83	7	0.95	0.93–0.97	7	0.95	0.93–0.97
8	0.89	0.73–0.87	8	0.75	0.67–0.83	8	0.92	0.89–0.95	8	0.92	0.89–0.95
9	0.56	0.42–0.70	9*	0.41	0.26–0.57	9	0.66	0.54–0.77			
10	0.50	0.35–0.65	10	0.90	0.86–0.93						
11	0.66	0.54–0.77	11*	0.43	0.28–0.59						
12	0.77	0.69–0.85	12	0.57	0.45–0.70						
13**	0	0	13	0.75	0.66–0.83						
14	0.63	0.52–0.75	14	0.77	0.69–0.85				0.	0.	
15	0.69	0.58–0.79									
16	0.85	0.80–0.91									
17	0.90	0.86–0.94									

Table 3. Continued

M. APPETITE AND EATING			N. ABERRANT VOCALIZATION		
ITEM	ICC	CI (95%)	ITEM	ICC	CI (95%)
1	0.75	0.66–0.83	1	0.96	0.93–0.98
2	0.63	0.52–0.75	2	0.95	0.92–0.97
3	0.70	0.60–0.80	3	0.96	0.94–0.98
4	0.81	0.76–0.88	4	0.84	0.77–0.91
5	0.62	0.51–0.74	5	0.70	0.58–0.82
6	0.74	0.65–0.82	6	0.83	0.75–0.90
7	0.81	0.74–0.87	7	0.95	0.93–0.97
8	0.96	0.94–0.97	8*	0.23	0.01–0.45
9*	0.37	0.20–0.54	9	0.64	0.50–0.78
			10*	0.00	0.00–0.24
			11*	0.41	0.22–0.61

Table 4. Pearson correlations and confidence intervals (95% confidence intervals) for select domains and relevant measures, by NPI version and rating method

DOMAIN	NPI-C SUM OF CLINICIAN RATINGS	NPI CAREGIVER DOMAIN RATING OF FREQUENCY × SEVERITY
AES – Apathy	0.31 (0.14 to 0.42)	0.22 (0.02 to 0.42)
BPRS – Delusions plus hallucinations	0.60 (0.47 to 0.70)	0.56 (0.39 to 0.69)
CMAI – Agitation only	0.40 (0.25 to 0.54)	0.19 (0.00 to 0.37)
CMAI – Agitation plus aberrant vocalization*	0.60 (0.42 to 0.73)	*
CSDD – Depression	0.61 (0.49 to 0.71)	0.34 (0.18 to 0.49)

*Aberrant vocalization is not a domain on the NPI.

AES = Apathy Evaluation Scale; BPRS = Brief Psychiatric Rating Scale; CMAI = Cohen-Mansfield Agitation Index; CSDD = Cornell Scale for Depression in Dementia.

Discussion

In this paper, we report on the development of a new state-of-the art instrument to measure NPS in people with dementia and the results from an international validation study. The NPI-C capitalizes on the existing strength of NPI and has rectified the weaknesses of NPI. Overall, the NPI-C is responsive to shortcomings in NPS measures and has the potential to function as a single tool that could be used in clinical trials either as a “broad spectrum” assessment across many domains or for “in-depth” monitoring of a limited set of domains (Lyketsos, 2007).

We first note that nested within the NPI-C are the original NPI items. This provided us with the opportunity to examine ICCs and convergent validity using clinician ratings for this widely used measure. Twelve original items had ICC values <0.5, the majority of which ($n = 5$) were in the domain of hallucinations. One item (item 4) in hallucinations lacked variability; three (items 5–7) had an ICC value of 0.00. This discrepancy may point to conflicting responses by caregivers and/or

patients regarding the occurrence of various types of hallucination-related behaviors and/or relative infrequency of them, in which any small discrepancy would affect the ICC value. Other NPI domains with low ICC items were delusions (item 7); elation (item 5), disinhibition (item 4), aberrant motor (items 4 and 6), and sleep disturbances (item 6).

For the NPI-C items that were not part of the original NPI, the domain with highest number of items <0.50 was “aggression.” Five items lacked any variability, suggesting their occurrence may be too rare to be worthwhile. Three new items in “agitation” also had ICC values <0.50. Overall, a total of 17 items were removed (see Table 1). For the remaining items, ICCs were moderate to strong. Although three items were removed in the new domain of aberrant vocalizations (ICC = 0.23), the eight remaining items had values ranging from 0.70 to 0.96. This suggests that the new domain may provide useful and previously unreported information on this behavior, which is relevant for people in moderate to severe dementia stages.

Convergent validity was moderate to strong for the four domains and their corresponding measures.

“Apathy” had the weakest correlation ($r = 0.31$) for the NPI-C. We note that the correlation for the NPI global domain rating was much lower ($r = 0.22$). One factor possibly contributing to the weaker correlation is the absence of data from the Brazil site. Inconsistencies in the translation of the AES necessitated the exclusion of data from this site ($n = 15$). Although Camozzato and colleagues (2008) reported reliability data for the Brazilian Portuguese version of the NPI, they did not examine validity, specifically whether the cultural interpretation of apathy-related items were similar in Brazilian caregivers as in other languages. It is interesting to note that Camozzato *et al.* found higher scores of severity and distress in the apathy domain than in any other NPI domain.

The strength of correlation between “agitation” and the CMAI for the NPI-C was moderate ($r = 0.40$), but increased substantially when the domain of aberrant vocalization was added ($r = 0.60$). The correlation was very weak, however, for the domain rating of agitation in the NPI ($r = 0.19$). Although the correlation strengthened for the NPI with the addition of aggression ($r = 0.31$), it was still not as strong as the NPI-C agitation domain alone and the CMAI. This points to the potential usefulness of the NPI-C agitation domain as a “stand alone” measure in trials. Another domain that shows strength as a stand-alone measure is the depression/dysphoria domain of the NPI-C ($r = 0.61$), which showed significant improvement over the NPI for the same domain ($r = 0.31$).

At this stage of the scale development, the NPI-C has not yet been incorporated into any clinical trials and its sensitivity to change is unknown. The raters in the current study were experienced in dementia research and in the assessment of NPS, and came from different clinical and research backgrounds. They included nurses, physicians, gerontologists, social workers and others with research expertise. The performance of the NPI-C in multicenter trials including raters of varying levels of expertise requires further investigation. Due to small samples sizes at each of the site, we are not able to assess reliability across languages and sites at this time but will address this shortcoming in future studies.

Overall, the study results demonstrate the utility of the NPI-C as a measuring tool of NPS in clinical trials. There are several notable advantages to this measurement approach. The NPI-C allows the flexibility of simultaneously administering the NPI. Since original NPI items are included, researchers can record NPI-C data in addition to NPI scores, which will facilitate cross-trial and site comparisons. Several NPI-C domains also show promise as stand-alone measures, which will also facilitate study comparison and eliminate the need to include other

outside measures. This can improve uniformity in study design and reduce error and administration time. The availability of the NPI-C in several languages through this validation study is another added benefit.

Overall, the NPI-C is a universal tool that can accurately measure several NPS. It uses a uniform scale system, which will facilitate data comparisons across studies. The NPI-C may be extremely useful in several settings, including clinical trials, observational studies, and potentially in clinical practice as well.

Conflict of interest

None.

Description of authors' roles

K. de Medeiros and C. Lyketsos co-designed the study, supervised the data collection and wrote the paper. P. Robert, S. Gauthier, F. Taragano, J. Kremer, A. Porsteinsson, Y. Geda and F. Stella contributed to the study design, participated in the Delphi panel, supervised data collection at their sites and assisted with writing the paper. H. Brodaty contributed to the study design, participated in the Delphi panel and assisted with writing the paper. A. Politis, A. Brugnolo and G. Gazdag supervised data collection at their sites and assisted with writing the paper. J. Leoutsakos contributed to the statistical design and data analysis and to writing the paper. J. Cummings assisted with study content and design.

Acknowledgments

The research reported in this paper was partially funded by an educational grant from Forest Pharmaceuticals, Inc. (Dr. de Medeiros) and grant P50AG005 from the Johns Hopkins Alzheimer's Disease Research Center (Dr. Lyketsos). Dr. Geda was supported by NIMH (K01 MH068351) and Harold Amos Scholar (RWJ foundation) grants. Dr. Leoutsakos was supported by the National Institute of Aging (RAG031348Z NIH/NIA). Dr. Robert received support from Lundbeck France. Dr. Taragano was supported by Lina Esevich grant #310618 to the CEMIC University Hospital Dementia Research Unit.

References

- Alexopoulos, G. S., Abrams, R. C., Young, R. C. and Shamoian, C. A. (1988). Cornell Scale for Depression

- in Dementia. *Biological Psychiatry*, 23, 271–284.
- Ballard, C., Day, S., Sharp, S., Wing, G. and Sorensen, S.** (2008). Neuropsychiatric symptoms in dementia: Importance and treatment considerations. *International Review of Psychiatry*, 20, 396–404.
- Brodaty, H. et al.** (2001). Psychosis, depression and behavioral disturbances in Sydney nursing home residents: Prevalence and predictors. *International Journal of Geriatric Psychiatry*, 16, 504–512.
- Camozzato, A. L. et al.** (2008). Reliability of the Brazilian Portuguese version of the Neuropsychiatric Inventory (NPI) for patients with Alzheimer's disease and their caregivers. *International Psychogeriatrics*, 20, 383–393.
- Cohen-Mansfield, J., Marx, M. S. and Rosenthal, A. S.** (1989). A description of agitation in a nursing home. *Journal of Gerontology: Medical Sciences*, 44, M77–M84.
- Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thomson, S., Carusi, D. A. and Gornbein, J.** (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, 44, 2308–2314.
- Dwyer, M. and Byrne, G. J. A.** (2000). Disruptive vocalization and depression in older nursing home residents. *International Psychogeriatrics*, 12, 463–471.
- Feldman, H. et al.** (2004). Behavioral symptoms in mild cognitive impairment. *Neurology*, 62, 1199–1201.
- Folstein, M. F., Folstein, S. E. and McHugh, P. H.** (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Gauthier, S. et al.** (2010). Management of behavioral problems in Alzheimer's disease. *International Psychogeriatrics*, 22, 346–372. doi:10.1017/S1041610209991505.
- Geda, Y. E. et al.** (2008). Prevalence of neuropsychiatric symptoms in mild cognitive impairment and normal cognitive aging. *Archives of General Psychiatry*, 65, 1193–1198.
- Kaufer, D. I. et al.** (2000). Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. *Journal of Neuropsychiatry and Clinical Neurosciences*, 12, 233–239.
- Lopez, O. L. et al.** (2003). Psychiatry symptoms vary with the severity of dementia in probable Alzheimer's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, 15, 346–353.
- Lyketsos, C. G.** (2007). Neuropsychiatric symptoms (behavioral and psychological symptoms of dementia) and the development of dementia treatments. *International Psychogeriatrics*, 19, 409–420.
- Lyketsos, C. G. et al.** (1999). Physical aggression in dementia patients and its relationship to depression. *American Journal of Psychiatry*, 156, 66–71.
- Lyketsos, C. G., Steinberg, M., Tschanz, J. T., Norton, M. C., Steffens, D. C. and Breitner, J. C. S.** (2000). Mental and behavioral disturbances in dementia: findings from the Cache County Study on memory and aging. *American Journal of Psychiatry*, 157, 708–714.
- Lyketsos, C. G., Lopez, O., Jones, B., Fitzpatrick, A. L., Breitner, J. and DeKosky, S.** (2002). Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the Cardiovascular Health Study. *JAMA*, 288, 1475–1483.
- Marin, R. S., Biedrzycki, R. C. and Firinciogullari, S.** (1991). Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Research*, 38, 143–162.
- Mayer, L. S. et al.** (2006). Comparison of three rating scales as outcome measures for treatment trials of depression in Alzheimer disease: findings from DIADS. *International Journal of Geriatric Psychiatry*, 21, 930–936.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. and Stadlan, E. M.** (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34, 939–944.
- Perneczky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J. and Kurz, A.** (2006). Mapping scores onto states: Mini-mental State Examination and Clinical Dementia Rating. *American Journal of Geriatric Psychiatry*, 14, 139–144.
- Perrault, A., Oremus, M., Demers, L., Vida, S. and Wolfson, C.** (2000). Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of behavior and mood scales. *Journal of Geriatric Psychiatry and Neurology*, 13, 181–196.
- Reisberg, B., Ferris, S. H. and De Leon, M. J.** (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139, 1136–1139.
- Reisberg, B., Borenstein, J., Salob, S. P., Ferris, S. H., Franssen, E. and Georgotas, A.** (1987). Behavior symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, 48 (Suppl. 5), 9–15.
- Reisberg, B., Auer, S. F. and Monteiro, I.** (1996). Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. *International Psychogeriatrics*, 8, 301–308.
- Robert, P. H. et al.** (2009). Proposed diagnostic criteria for apathy in Alzheimer's disease and other neuropsychiatric disorders. *European Psychiatry*, 24, 98–104.
- Rosen, W. G., Mohs, R. C. and Davis, K.** (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, 141, 1356–1364.
- Schreiner, D. et al.** (2005). Components of behavioral pathology in dementia. *International Journal of Geriatric Psychiatry*, 20, 137–145.
- Sink, K. M., Holden, K. F. and Yaffe, K.** (2005). Pharmacological treatment of neuropsychiatric symptoms of dementia. *JAMA*, 293, 596–608.
- Starkstein, S. E.** (2000). Apathy and withdrawal. *International Psychogeriatrics*, 12 (Suppl.1), 135–137.
- Steinberg, M. et al.** (2004). The persistence of neuropsychiatric symptoms in depression: the Cache County Study. *International Journal of Geriatric Psychiatry*, 19, 19–26.
- Steinberg, M. et al.** (2006). Risk factors for neuropsychiatric symptoms in dementia: the Cache County Study. *International Journal of Geriatric Psychiatry*, 21, 824–830.
- Tariot, P. N., Schneider, L. S. and Katz, I. R.** (1995). Anticonvulsant and other non-neuroleptic treatment of

agitation in dementia. *Journal of Geriatric Psychiatry and Neurology*, 8, S28–S39.

Ventura, J. *et al.* (1993). Training and quality assurance with the brief psychiatric rating scale: “the drift buster.” *International Journal of Methods in Psychiatric Research*, 3, 221–224.

Weyer, G., Erzigkeit, H., Kanowski, S., Ihl, R. and Hadler, D. (1997). Alzheimer’s disease assessment scale: reliability and validity in a multicenter clinical trial. *International Psychogeriatrics*, 9, 123–138.

Wilberg, T., Dammen, T. and Friis, S. (2000). Comparing personality diagnostic questionnaire 4+ with longitudinal, expert, all data (LEAD) standard diagnoses in a sample with a high prevalence of axis I and axis II disorders. *Comprehensive Psychiatry*, 41, 295–302.

Appendix 1

Members of the NPI-C Research Group

Henry Brodaty
Andrea Brugnolo
Kelly M. Cosman

Jeffrey Cummings
Kate de Medeiros
Renaud David
Éva Frey
Gabor Gazdag
Serge Gauthier
Yonas E. Geda
Sebastiao Gobbi
Janus Kremer
Jeanne LaFountain
Jeannie-Marie Leoutsakos
Constantine Lyketsos
Emmanuel Mulin
Katalin Nagy
Maria Passa
Larissa Pires de Andrade
Antonis Politis
Anton P. Porsteinsson
Philippe Robert
Florindo Stella
Fernando Taragano