REVIEW

The merits and problems of Neuropsychiatric Inventory as an assessment tool in people with dementia and other neurological disorders

Claudia KY Lai

School of Nursing, The Hong Kong Polytechnic University, Special Administrative Region of the People's Republic of China

Objective: The Neuropsychiatric Inventory (NPI) is one of the most commonly used assessment scales for assessing symptoms in people with dementia and other neurological disorders. This paper analyzes its conceptual framework, measurement mode, psychometric properties, and merits and problems.

Method: All articles discussing the psychometric properties and factor structure of the NPI were searched for in Medline via Ovid. The abstracts of these papers were read to determine their relevance to the purpose of this paper. If deemed appropriate, a full paper was then obtained and read.

Results: The NPI has reasonably good content validity and internal consistency, and good test–retest and interrater reliability. There is limited information about its sensitivity, specificity, positive and negative predictive values, and, in particular, responsiveness. Merits of the NPI include being comprehensive, avoiding symptom overlap, ease of use, and flexibility. It has problems in scoring (no multiples of 5, 7, and 11) and, therefore, analysis using parametric tests may not be appropriate. The use of individual subscales also warrants further investigation.

Conclusion: In terms of its content and concurrent validity, intra- and interrater reliability, test–retest reliability, and internal consistency, the NPI can be considered as valid and reliable, and can be used across different ethnic groups. The tool is most likely unable to deliver as good a performance in terms of discriminating between different disorders. More studies are required to further evaluate its psychometric properties, particularly in the areas of factor structure and responsiveness. The clinical utility of the NPI also needs to be further explored.

Keywords: measurement, neuropsychiatric symptoms, outcome assessment

Background

Behavioral disturbances are deemed the most problematic in the management and care of people with Alzheimer's disease (AD). Various instruments have been used to assess behavioral disturbances in dementia for treatment-evaluation purposes. Amongst them, the Neuropsychiatric Inventory (NPI) is deemed one of the most useful outcome measures for behavior and mood symptoms in people with dementia.[1] The NPI was developed by Cummings et al[2] in 1994. Although initially designed to target demented populations, it has been used to evaluate patients with psychotic, affective,[3] and other neurological disorders, such as Parkinson's disease and epilepsy.[4,5] Over the years, the NPI has gained in popularity and been translated into many different languages, including Chinese, Danish, Dutch, French, German, Greek, Hebrew, Italian, Japanese, Norwegian, Portuguese, Spanish, Swedish, and Thai. Although a widely used and important instrument, its properties are not entirely problem-free. This review examines the merits and concerns regarding its use, so that researchers and clinicians can be better informed of the proper use of the NPI.

Correspondence: Claudia KY Lai
School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Special Administrative Region of the People's Republic of China
Tel +852 2766 6544
Fax +852 2364 9663
Email claudia.lai@polyu.edu.hk

Methods

All articles discussing the psychometric properties of NPI from 1995 to 2013 were searched in Medline via Ovid using "Neuropsychiatric Inventory" or "NPI" and "psychometric properties" as keywords. Twenty-one papers were found after removing duplicates. "Neuropsychiatric Inventory-Questionnaire" or "NPI-Questionnaire" or "NPI-Q" and "psychometric properties" were then searched using the same strategy. Thirteen articles were found after removing duplicates. Last, a search using "Neuropsychiatric Inventory-Nursing Home" or "NPI-NH" and "psychometric properties" as keywords for the same period found 14 papers after removing duplicates. The abstracts of these papers were read to see if they were relevant to the purpose of this paper. If deemed appropriate, a full paper was then obtained. Appropriateness was defined as those papers that discussed the tool itself, not merely mentioning it briefly as part of a battery of assessment tools. Because this paper is not a systematic review, the search strategies were only conducted to ensure that the author had read as much as possible about the topic before conducting a critical review of the tool. The reference list of relevant papers was also examined in order not to miss any paper on the topic. In the process of writing up the manuscript, the author also searched for more papers using "factor structure" as the keyword search for NPI-related publications in order to better understand how studies reported the NPI's factor structure. One hundred and one papers were found after removing duplicates. Again, the abstracts were read to determine whether they were useful to the discussion before obtaining the full paper. All relevant papers obtained about NPI's factor structure were carefully read in full and are included in Table 1.

NPI the instrument

The NPI is a condition-specific measure designed to assess neuropsychiatric disturbances in people with AD, as well as other related dementing disorders. When first developed, it assessed 10 behavioral disturbances, namely delusions, hallucinations, dysphoria, anxiety, agitation/aggression, euphoria, disinhibition, irritability/lability, apathy, and aberrant motor activity. Subsequently, the tool was refined and expanded to 12 domains, adding night-time behavior disturbances as well as appetite and eating abnormalities to the scale.[6,7] The NPI assesses not only the presence, but also the frequency and severity of each behavior in the previous month. It also assesses the level of caregiver distress as a result of each of the neuropsychiatric problems.

Table I Studies on the factor structure of the NPI

Journal	Patient group	Setting	Sample size	MMSE score	Version of NPI (ten or 12 domains)	Factor structure	% of variance explained
Aalten et al[40]	Dementia	Twelve centers from the European Alzheimer's Disease Consortium	2,808	17.6±6.1	12	Hyperactivity/ psychosis/ affective symptoms/ apathy	23.27% 10.61% 9.39% 9.07%
Frisoni et al[23]	AD	Hospital	162	13.3±6.7	Ten	Mood/psychotic/frontal	56.30%
Hirono et al[41]	AD	Hospital	228	18.5±5.1	12	Psychosis	NA
Lange et al[3]	Sample aged 41–95 years, mean age =73.4 years	Hospital	204	NA	12	Agitation/mood/psychosis/sleep/motor activity/ elevated behavior	63.20%
Matsui et al[42]	Mild to moderate AD	Hospital	140	20.3±4.2	Ten	Mood/psychosis/euphoria	NA
Zuidema et al[19]	Mild to severe dementia GDS stage 4–7	Nursing home	Stage 4/5: 341 Stage 6: 728 Stage 7: 368	NA	12	Various factor structure for different GDS stage	15%–61%
Lam et al[43]	Dementia	Nursing home	125	8.62±6.04	12	Hyperactivity/hypoactivity behaviors/verbally aggressive/aberrant behaviors	NA

Study	Population	Setting	N	MMSE	No. of factors	Factors/symptoms	Prevalence
Dechamps et al[18]	Total	Nursing home	163	NA	12	Hyperactivity/affective/psychosis/hallucinations/frontal lobe symptoms/sundowning/psychomotor agitation	NA
	Dementia		109	14.2±7.1		Hyperactivity/affective/psychosis/hallucinations	63.90%
	Psychosis		54	17.4±7		Frontal lobe symptoms/sundowning/psychomotor agitation	61.30%
Selbaek et al[44]	Dementia	Nursing home	41	14.3±9.1	12	All 12 factors	NA
Zuidema et al[45]	Dementia	Nursing home	1,319	NA	12	Agitation/disinhibition/irritability/aberrant motor behavior/delusions/hallucinations/depression/anxiety/apathy	NA
Lam et al[46]	Dementia	Social center and nursing home	291		12	Apathy/depression	16.4%–45.8%
	(Subgroup QD)		195	23.64±3.14			
	(Subgroup AD)		96	17.96±3.30			
Kaufer et al[8]	AD	Outpatient clinic	60	18.4±5.6	12	Hallucinations/dysphoria/aberrant motor behaviors/appetite/delusions	NA
Fuh et al[47]	AD	Outpatient clinic	95	12.7±7.2	12	Mood-psychosis/psychomotor regulation/social engagement	59.00%
Leung et al[14]	Dementia	Outpatient clinic	62	NA	12	12 factors tested individually	2.70%–11.70%
Chow et al[35]	Possible AD	Outpatient clinic	430	NA	12	All 12 factors	NA
Aalten et al[48]	Dementia	Outpatient clinic	199	18.1±4.7	12	Mood/apathy	14.49%
						Psychosis/	11.08%
						hyperactivity	29.54%
Mirakhur et al[49]	AD	Outpatient clinic	435	13±9.2	12	Affect/physical/psychosis/hypomania	57%
Spalletta et al[50]	AD	Outpatient clinic	244	≤26	Ten	Mood-excitement/mood-depression-apathy/psychosis/hyperactivity/anxiety	73%
Fuh et al[36]	Dementia (with different subgroups)	Outpatient clinic	536	11.6–17.1 ±6.8–7.5	12	Apathy/sleep disturbance	NA
Matsumoto et al[51]	Dementia (with different subgroups)	Outpatient clinic	110	18.6±6.9	NA	NA	NA
Fuh et al[34]	AD	Outpatient clinic	81	20.1±4.5	12	Distress	6%
Kilmer et al[52]	Mild to moderate TBI	Outpatient clinic	51	NA	12	All 12 factors	NA
Blair et al[33]	Dementia (with different subgroups)	Outpatient clinic	177	20.8±5.6 –28.8±1.2	12	Aggression/apathy/disinhibition/irritability/aberrant motor behavior	NA
Gallo et al[37]	Dementia	Outpatient clinic	124	22.6±3.5	12		52%–83%

Abbreviations: AD, Alzheimer's disease; GDS, Global Deterioration Scale; MMSE, Mini-mental State Examination; NA, not available; NPI, Neuropsychiatric Inventory; QD, questionable dementia; TBI, traumatic brain injury.

The NPI-Questionnaire validated by Kaufer et al[8] is a shortened version of the NPI for use by clinicians. Limited discussion can be found about the NPI-Questionnaire. There is also a version developed for nursing home use known as the NPI-Nursing Home (NPI-NH),[9] also with limited discussion in the literature. There are no clear explanations about the differences between the NPI-NH and the NPI, except that the family distress score is renamed as the occupational disruptiveness score in the nursing home version. This paper focuses on the NPI (full version). A systematic approach to critically analyzing clinical outcome measures, put forward by Kane and Radosevich,[10] is adopted in the following critique of the NPI.

Measurement mode

The NPI is a quantitative measure employing caregiver rating. Cummings et al regarded caregivers as the most appropriate people to report behaviors based on the rationale that patients with dementia are often unable to recall or describe their symptoms and, therefore, are not optimal informants.[2] Also, patients may not exhibit behavioral abnormalities during the course of a clinical visit. Changes would be underestimated if the ratings were based on the clinician's observation during an interview.

Administration and scoring

A screening question is asked first in each of the domains. After the caregiver indicates that there is a behavioral disturbance with the screening question, she/he will then answer the seven or eight subquestions related to that particular behavior. After administering the subquestions, the researcher will ask the caregiver to rate the frequency and severity of each abnormality, and will then rate the associated caregiver distress. The frequency rating is from 1 (occasionally or less than once a week) to 4 (very frequently, more than once a day or continuously), and the rating of the symptom severity is 1, 2, or 3 (mild, moderate, or severe, respectively). The stress to the caregiver is rated from 0 (no distress) to 5 (extreme distress). The domain score is obtained by multiplying the frequency and severity scores. The total NPI score is the sum total of all of the individual domain scores (0–144). The caregiver distress level is not part of the total NPI score. The amount of time required to complete the NPI is around 20–30 minutes.

Development of the NPI

Cummings did not provide a direct account of the conceptual framework guiding the design of the NPI.[2,7] His team exhaustively examined the literature to come up with a list of neuropsychiatric behaviors that commonly occurred in people with AD and related dementing disorders, then grouped them into domains with sets of subquestions. The design and conceptualization of the NPI as a tool can be considered as a traditional medical model: the disease leads to symptoms; therefore, a measurement of the symptoms' response to treatment is needed. The meaning of the behaviors is not considered as important in the NPI. It only attempts to quantify the symptoms (behaviors). There is no attempt to distinguish between behaviors that are possibly triggered by the physical environment (eg, new place or new routines made the patient become disorientated and wander) or the psychosocial environment (eg, made to have a shower and therefore becomes resistive or agitated). This approach of the NPI has its merits and disadvantages. It can be difficult to determine the cause of disturbing behaviors in people with dementia. Avoiding identification of the meaning underlying the behavior renders it easier to administer the instrument. On the other hand, it introduces detection biases because of the indiscriminant attribution of behaviors as neuropsychiatric symptoms.

Second, as an evaluation tool, the NPI does not seek to know the patient's view in assessing outcomes. McKinlay et al[4] used the NPI as a measure to compare caregiver and self-reports in neuropsychiatric problems. The researchers observed that, although similar rates of symptoms were reported by both patients and caregivers, the level of agreement between the dyads was low. They postulated that the lack of agreement may be the result of caregivers being asked to report on problems that were not readily identifiable based on observed behavior, and concluded that the reports of caregivers and patients cannot be regarded as interchangeable. The NPI can be used as a caregiver rating as long as we are aware that it is an assessment coming from a third-party perspective. The patients and the observer, be they the family or formal caregivers, may have different perceptions of the problems with which they are dealing.

To reduce the administration time in using the NPI, Kang et al[11] studied the use of a caregiver-administered version. Sixty-one caregivers of people with dementia were asked to complete the written form of the worksheet with supervision. Kang et al found that the frequency, severity, and caregivers' distress scores of the caregiver-administered NPI correlated significantly with the results of the NPI rated by professionals ($r>0.6$, $P<0.001$), and the total caregiver-administered-NPI scores also correlated with total NPI scores ($R=0.86$, $P<0.001$). They suggested that the caregiver-administered

version could be substituted for the NPI by professionals to save administration time. However, the suggestion did not seem to be readily embraced by the field. Wood et al[9] compared the responses of certified nurses' aides and licensed vocational nurses with research observations and cautioned that the NPI may not be an appropriate instrument for tracking behavioral changes when used by non-research staff. In assessing psychopathologies in epilepsy patients, Krishnamoorthy and Trimble[5] also noted that caregivers reported fewer behavioral abnormalities in the NPI interview as compared to the results assessed by research personnel using the Brief Behavior Rating Scale. Although relatively easy to use, it is not yet confirmed that the NPI can be a caregiver-administered tool.

Psychometric properties
Content validity
Cummings et al[2] reported that, because there is no gold standard for comparison for the domains of disinhibition, euphoria, apathy, and irritability, they submitted the NPI to a panel of ten experts in neuropsychology, geriatric psychiatry, and behavioral neurology, and obtained the face validity of the instrument using the Delphi process. Each panel member rated each screening and subquestion in each domain from 1 (well assessed) to 4 (poorly assessed). The result was that each group of questions scored less than 2, except for the category of questions under "troublesome behavior", which was subsequently reformulated as "aberrant motor behavior" according to the recommendations of the panel. Based on this, the face validity of the NPI can be said to be good.

The behavior categories of dysphoria, aggression, aberrant motor behavior, anxiety, delusion, and hallucinations were compared with the affective disturbance, aggressiveness, activity disturbances, anxiety and phobia, delusion, and hallucinations items of the Behavioral Pathology of Alzheimer's Disease Rating Scale (BEHAVE-AD).[12] The NPI domain of dysphoria was compared with the Hamilton Rating Scale for Depression (HAM-D).[13] All the above correlations reached the 0.05 significance level in Cummings et al's[2] study involving 40 subjects and 40 caregivers. Concurrent validity in a study by Leung et al[14] in Hong Kong reported that the NPI demonstrated an acceptable level of concurrent validity with commonly used instruments for most of the domains.

In a recent report,[15] the correlations among corresponding subscales of the BEHAVE-AD and the NPI were found to be relatively weaker. They were between 0.54 and 0.78 for frequency of symptoms and 0.47 to 0.80 for severity of symptoms. The concurrent validity of the NPI probably needs further testing before a substantial claim can be made that it has reached an acceptable level of concurrent validity against standard instruments. The domains of "night-time behavior" and "appetite/eating change" have not been examined for concurrent validity, likely because the development of assessment tools in these two areas has been limited. The item of "sleep disturbed behaviors" was subsequently expanded to become the Sleep Disorders Inventory and is used for testing sleep disturbances in persons with Alzheimer's disease.[16] The NPI, however, has been used as a concurrent validity measure against the revised Cambridge Behavioral Inventory to establish the revised Cambridge Behavioral Inventory's validity for assessing behavioral symptoms in persons with dementia in general practice settings.[17]

Internal consistency and factor structure
Cummings et al[2] reported a high level of internal consistency for the overall score (α=0.88), and for severity and frequency ratings in 40 AD patients. Cummings et al[2] also noted that 78% of the scale's items showed no significant relationship with each other, indicating that these items were assessing different behaviors, rendering its internal consistency level somewhat intriguing. Subsequent reports of the internal consistency of the NPI were mainly conducted using the newer 12-domain version NPI and also the NPI-NH. Studies reported an α-range of 0.67–0.8 in terms of the NPI's internal consistency.[3,18] Overall, the NPI can be said to have reasonable to good internal consistency.

Zuidema et al[19] reported that the factor structure of the NPI is fairly stable. However, NPI's factor structure actually varies with different populations, as shown in Table 1. This is hardly surprising because of the intervening factors, which might include: version of the NPI used (ten or 12 domains); target patients; inclusion and exclusion criteria; different cut-off points for factor loading; and other demographic and clinical variables. For example, having five factors in the NPI-NH is, of course, quite different from having 12 factors in the 12-domains NPI. It can be considered as a trade-off between grouping behaviors together into the same factor or perceiving these behaviors as reflecting different domains (eg, both euphoria and dysphoria are related to mood but are appraised as different domains in the NPI). The finding that the NPI behavior items have little correlation with one another suggests that the information provided in the item scores may be more relevant than the overall total score.[20] Also, it needs to be mentioned that patients with dementia or AD are heterogeneous, with diverse behavioral profiles.

To say that differentiating different factor structures would help diagnose various dementing conditions is probably too high an expectation for the tool. A lot needs to be done if the NPI is to be able to make such a claim.

Sensitivity, specificity, and positive and negative predictive values

Information about the overall positive and negative predictive rates was not reported.[2,6] Reportedly, the NPI was tested in two groups of elderly subjects – one group with dementia and one group without dementia – and was able to distinguish between the two groups. According to the NPI authors,[2] the screening questions were found to have a false negative of less than 5%. Leung et al[14] validated the Chinese version of the NPI in a sample of 62 dementia outpatients. The false negative rates of most stem questions were found to be low, while those of dysphoria, sleep, and appetite were slightly above 10% in their study. Only one study compared the efficacy of the Empirical Behavioral Rating Scale (E-BEHAVE-AD), Neurobehavioral Rating Scale, and NPI in detecting behavioral and psychotic symptoms in dementia using receiver operating characteristic analysis.[21] The authors found that the instruments were equally likely to detect agitation. While the Neurobehavioral Rating Scale was most likely to detect psychosis, the NPI was best at detecting improvements in agitation. Discussion of this dimension of the NPI's psychometric properties has been limited.

Test–retest reliability

Twenty participants took part in Cummings's test–retest reliability testing with a 3-week interval.[7] Half of the second interviews conducted were via telephone. The authors reported that, overall, all measures of the NPI were significantly correlated, and that the test–retest reliability reached an acceptable level of 0.79 for frequency (P=0.0001) and a fairly good level of 0.86 for severity (P=0.0001). Moreover, the results of the telephone interviews did not differ significantly from the face-to-face interviews. Good test–retest reliability was again confirmed in other studies by Cummings et al and Frisoni et al.[22,23]

Interrater reliability

As an instrument, the NPI has been found to have good interrater reliability. Cummings reported having two blinded raters paired up to evaluate the same subject (who was interviewed by only one of the raters), and this was tested on 45 subjects.[6] Excellent interrater reliability levels in different domains were achieved (93.6%–100%). Interrater

reliability was reconfirmed by subsequent studies.[22] Leung et al[14] reported a range of kappa and intraclass correlation coefficients for all but one item (appetite severity) between 0.7 and 1.00.

Responsiveness

The NPI is reportedly sensitive to drug-induced behavioral changes.[9,22] Kaufer found a significant reduction in total NPI scores across all 40 subjects in their sample treated with antidementia drugs.[24] Mega et al[25] investigated the range of behavioral abnormalities in patients with AD compared with normal age-matched control subjects and demonstrated stage-specific trends in neuropsychiatric symptoms in AD patients.

Other researchers, however, queried the evidence supporting the NPI's responsiveness to change. In a clinical trial on metrifonate,[26] the NPI mean score was increased by 3.9 points in the placebo group and by 1.2 points in the treatment group (P=0.02). These data were used by Mega et al[25] to define cutoffs for improvement (decrease ≥4 points), no change (±3 points), and worsening (increase ≥4 points) on the scale. These cutoffs that were based on statistical significance provided no useful information about clinical significance.[20] In addition, Perrault et al[20] argued that behavior and mood improvement observed in clinical drug trials that were not double-blinded should not be considered as confirmation of the scale's responsiveness to change. Because many of the studies using the NPI as the primary outcome measure did not provide information about the spread of score (in quartiles) of the NPI,[27] the presence or extent of the floor or ceiling effects in the instrument cannot be ascertained.

Two subscales of the NPI (depression and apathy) were used as one of the measures to evaluate the effect of depression and apathy on functional recovery in post-stroke Japanese patients.[28] Posttest subscale scoring in a subsample of 59 patients with depression and 13 residents with apathy had a fairly narrow spread of scores. Graphical information was provided by the authors instead of numerics, rendering it difficult to interpret the actual responsiveness of these two NPI subscales.

Many drug trials have used NPI scores as the primary outcome measure.[29,30] However, there is limited discussion of the responsiveness of the instrument in their reports. Behavior and mood improvement observed in open label studies of cholinesterase inhibitors should not be used as evidence of the NPI's responsiveness to change. Perrault et al[20] argued that, in the absence of blinding, the results of many of these studies could be explained by regression to the mean, and

should not be considered as definitive confirmation of the scale's responsiveness to change. In addition, the relationship between symptom and intensity may be nonlinear; therefore, the various constructs measured by the NPI may have differential sensitivity to treatment.[15]

Profiling neuropsychiatric features among different neurological disorders

Cummings suggested that the NPI provides a profile of behavioral changes that helps to distinguish AD from other types of dementia.[6] A variety of conditions has been studied, including AD, frontotemporal dementias, progressive supranuclear palsy, and traumatic brain injury. In Cummings' report, significant differences on NPI profiles emerged. For example, patients with frontotemporal dementia exhibited significantly more apathy, disinhibition, euphoria, and aberrant motor behavior than those with AD, and patients with progressive supranuclear palsy had significantly more apathy and less agitation and anxiety than those with AD. Patients with vascular dementia were more likely to have depression and less likely to have delusions, and patients with dementia with Lewy bodies more often exhibited delusions and hallucinations than patients with AD.[18] In a report from the European Alzheimer Disease Consortium,[31] cross-sectional data of 2,354 patients with AD using the NPI for assessment of neuropsychiatric symptoms were collected from 12 centers. The authors reported the presence of four neuropsychiatric subsyndromes: hyperactivity; psychosis; affective symptoms; and apathy. The authors claimed that the data provided robust evidence for the existence of neuropsychiatric subsyndromes in AD.

Cummings considered that establishing behavior profiles that characterize different disorders may help to reduce diagnostic error when patients are recruited to participate in clinical trials.[6] Yet when Litvan et al[32] attempted to characterize the neuropsychiatric symptoms of patients with corticobasal degeneration (n=15) and patients with progressive supranuclear palsy (n=35) with normal controls (n=25), they found that the patients with corticobasal degeneration and progressive supranuclear palsy had overlapping symptomatic presentations as well as distinctive symptom profiles. The Frontal Behavioural Inventory developed by Blair et al[33] distinguished a higher percentage of frontotemporal dementia patients (>75% correct classification) from AD and other groups compared to the NPI (54.2%). Some antipsychotic treatment studies have commented that the NPI might not be as sensitive to change in the Parkinson's disease population

relative to the Brief Psychiatric Rating Scale.[15] More studies will be needed to determine whether the NPI can adequately differentiate different pathological conditions.

Biological correlates of the NPI

Also reported by Cummings was the NPI's capability to investigate the biological correlates of dementing disorders.[6] Frisoni et al reported a collection of studies examining the neurobiological correlates with neuropsychiatric symptoms as measured on the subscales of the NPI. These studies include using autopsy, imaging, electroencephalography, genetic, and biochemical examinations to correlate with subscales such as agitation, dysphoria/depression, psychosis, aggression, and other items on the NPI scale.[23] However, these studies have been too few to enable any definitive claims to be made. Many intervening variables could have confounded the outcomes and more substantial evidence will be required to further test the postulations.

Cross-cultural studies using the NPI

Transcultural studies have reported that neuropsychiatric symptom complexes are similar in US and European cultural groups.[22] Chow et al[35] compared the neuropsychiatric symptoms of Chinese subjects with AD at tertiary care centers in Taiwan and Hong Kong against those of Caucasian subjects in Los Angeles, California. The authors found that all items on the NPI were represented at each of the centers although not all the subjects had all the symptoms.

In Hong Kong, Leung et al[14] tested the psychometric properties of the Chinese version of the NPI in a sample of 62 dementia outpatients. The concurrent validity was tested by measuring the Spearman correlation between the Chinese version NPI subscales with the appropriate subscales of BEHAVE-AD and the Chinese HAM-D. Most Chinese version NPI behavioral domains achieved significant correlation with the corresponding BEHAVE-AD and Chinese HAM-D subscales. The Cronbach's alpha for the overall reliability was 0.84. The false negative rates of the screening question were found to be acceptable except for the dysphoria, sleep, and appetite domains. The interrater reliability was satisfactory, with the intraclass correlation coefficient of all subscales above 0.9. The authors concluded that the Chinese version NPI was applicable in assessing the neuropsychiatric symptoms of dementia in Chinese communities.

Fuh et al[36] validated a Chinese version of the NPI in Taiwan and reported their results together with researchers from Japan, Thailand, and Hong Kong. Fuh et al[34] argue that the NPI's reliability and validity have been shown in multiple

Asian studies, although few of these reports can be located from various databases. The focus of Fuh et al's report[34] was more about the ability of the NPI to capture the range of neuropsychiatric symptoms across different countries and ethnic groups. They noted that many similarities and some contrasts have emerged when comparing the results of their investigations with those from Western clinical centers, but gave few details on the statistical tests being done. From these two studies, it can be said that the performance of the NPI is fairly stable, and that it is noted to be a fairly valid and reliable instrument across countries.

Merits of the NPI
Comprehensiveness as an assessment tool for people with dementia
Cummings argued that many dementia rating scales used in research do not include alterations in personality, such as the commonly seen behaviors of apathy and irritability, which are items in the NPI.[6] Because other rating scales assess only behavioral presentations, the NPI helps to distinguish between different symptoms that are known to be rare in AD but that are common in other types of dementia, such as euphoria and disinhibition in frontotemporal dementias. These included items that help to enhance the comprehensiveness and utility of the NPI.

Avoiding symptom overlap in assessment
Differentiating between symptoms of depression and dementia can be challenging due to symptom overlap. It was the authors' intention to construct the NPI to address this problem. According to the authors,[22] the NPI depression/dysphoria scale contains only the central emotional aspects of depression (sadness, tearfulness, etc). Thus, a high score in this subscale establishes the presence of mood disturbance. Similarly, apathy and depression (possible confounders in many scales) are assessed independently on the NPI. The authors stated that the NPI allows identification of an apathy syndrome with or without a corresponding mood disorder with anhedonia.

Ease of understanding frequency scoring
Whether behavior and mood scales should be scored on frequency, severity, or both has been controversial. Reisberg et al[12] suggested that, because the time spent by caregivers with patients might vary greatly, frequency may be insensitive compared to the magnitude of the disturbance. They also considered magnitude to have greater clinical relevance. On the other hand, because the severity of an illness is difficult

to assess, Perrault et al[20] suggested that scoring by frequency of behavioral occurrence was preferable. Scoring by summation of items is logical, provided the items being summed are reflective of a single dimension of interest. Severity ratings are based on caregivers' subjective interpretation of how problematic symptoms appear to be for the patient, whereas frequency ratings could be more objectively and directly measured by the caregiver.[37] Although ease of use is one of its merits, the NPI has also been criticized as lumping together various behavioral presentations into neuropsychiatric symptoms.

Flexibility and ease of administration
There are various qualities of the NPI that render it flexible and easy to administer. The structured interview questions enable administration of the NPI by less clinically experienced professionals without affecting scale validity or reliability.[15] It is caregiver-based and, therefore, does not require the patient's cooperation, and can be used in agitated or advanced-disease patients.[7] The screening question strategy minimizes administration time. There are no restrictions on the intervals of assessment using the NPI. Cummings suggested that the measurement time interval could be adjusted according to the purpose of the evaluation, eg, since the onset of certain behaviors of interest or in the last month or last dose in a drug trial.[6] The assessment of frequency and severity of different behavioral categories in the NPI are two separate entities. It becomes easier to understand whether the symptoms occurred with the same frequency but less severely, or less frequently but with the same severity. As suggested by the authors, its utility in drug efficacy trials and other intervention studies is therefore increased. The NPI allows the rater to capture mild but very frequent phenomena or moderate but less frequent phenomena through separating symptom frequency from symptom severity to track the onset, frequency, and prevalence of various psychiatric syndromes over time.[15]

Problems with the NPI as an assessment tool
Problems with scoring
The NPI has a multiplicative scoring metric, which results in noncontinuous scores as symptom frequency and severity increase (ie, there are no multiples of 5, 7, and 11); it is also expected to depart from normality in its score distribution.[15,20] Noncontinuous scores may lead to problems in accurately evaluating the problem. Researchers have cautioned against the use of parametric methods in the analysis of NPI

scores.[20,25] Further work is needed in this area to identify the most suitable statistical method of analysis. Because the NPI is a retrospective (up to 1 month) caregiver-informant rating, the problem of recall bias in scoring cannot be disregarded.

Problems in its psychometric properties

The validity of the NPI is supported by comparison with existing scales. However, variations in the time period of recall can affect the reliability and validity of the scoring system, which is based on the product of frequency and severity. The resulting scores may therefore be difficult to compare across studies. There is also little information to substantiate the reported low false-negative rate (less than 5%) of the screening questions.[20] The inclusion of certain symptoms uncommon in AD (eg, euphoria, disinhibition, and compulsive and repetitive behaviors) may increase the NPI's diagnostic utility in dementia, but does not necessarily increase its responsiveness to changes in assessments. Strong evidence for the responsiveness to change of the NPI is not yet available. Perrault et al[20] noted that the reliability of the NPI seemed to be satisfactory based on available data. However, the reliability of the NPI was incompletely assessed and may have been overestimated because the reported studies had small sample sizes and employed suboptimal statistics (ie, correlation coefficients and percent agreement). More work will be needed to fully confirm the NPI's reliability.

NPI scores and caregiver distress

Behavioral problems and psychiatric symptoms are major sources of caregiver distress. The NPI quantifies the distress associated with each type of behavioral abnormality exhibited by the patient. The total NPI score is also found to be significantly associated with total caregiver distress scores. According to Cummings, the correlation between caregiver distress and patient behaviors has treatment implications. There would be potentials to reduce caregiver distress if individual behaviors responded to treatment.[6] However, this is an oversimplification of the concept of caregiver burden. Whether a caregiver would consider looking after a patient with AD as burdensome or distressing is not merely related to the types, frequency, and severity of the behavioral symptoms, but also to the relationship between the patient and the caregiver, whether the caregiver has good social support, whether they are financially secure, and so on. Caregiver distress and burden is a much more complex issue than simply tying it to the occurrence or reduction of behavioral concerns. In any case, this part of the NPI has not been adequately tested. More often than not, the distress severity rating was not used in studies and, if used, it was not reported.

Problems with subscale use

Use of the NPI subscales has been popular. For instance, the depression subscale was used by Leontjevas et al.[38] When studies use individual subscales, their validity warrants further attention. Even as a single item assessment in its subscales, researchers commented that there were associated problems. The ratings of the NPI produced one score per behavioral domain. Although this score is assumed to reflect the degree of disturbance of a particular domain/particular domains, raters are required to endorse a single frequency and a single severity score for each domain, which may include a number of symptoms.[2] This does not always provide specific information concerning the unique clinical picture of the patient being rated.[37] The NPI's subscale use therefore needs to be further tested for its clinical utility.

Summary

The NPI was introduced in 1994 and has since become widely popular as a standard instrument for clinical trials and other types of behavioral research in dementing disorders. Cummings et al[2,6] reported that studies examining the properties of the NPI in terms of its content and concurrent validity, intra- and interrater reliability, test–retest reliability, and internal consistency concluded that the instrument is both valid and reliable. The similarities of findings across cultures indicate that some neuropsychiatric abnormalities are more biologically and less culturally determined.[34] Thus, the NPI is probably relevant for patient populations across different ethnic groups. Some of the studies on the NPI provide support for a number of the researchers' claims, but more studies are required to further evaluate its psychometric properties, particularly in the areas of internal consistency, factor structure, and responsiveness. The clinical utility of the NPI also needs to be further explored. The tool is most likely unable to deliver as good a performance in terms of discriminating between different disorders.

We need to be aware that the majority of the studies discussed so far were surveys. Even if they had a temporal dimension, the time span was limited. Behavioral and mood disturbances vary in all patients with dementia and do not necessarily progress uniformly.[39] Heterogeneity of presentation and variability of progression make it challenging to track changes, therefore limiting the assessment of responsiveness to change for behavioral scales in longitudinal

studies.[20] Considering the limitations in the assessment of its psychometric properties, more work is needed to confirm the use of the NPI in clinical trials and as a tool for longitudinal studies.

Disclosure

The author reports no conflicts of interest in this work.

References

1. Conn D, Thorpe L. Assessment of behavioral and psychological symptoms associated with dementia. *Can J Neurol Sci.* 2007;34 Suppl 1: S67–S71.
2. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology.* 1994;44(12): 2308–2314.
3. Lange RT, Hopp GA, Kang N. Psychometric properties and factor structure of the Neuropsychiatric Inventory Nursing Home version in an elderly neuropsychiatric population. *Int J Geriatr Psychiatry.* 2004;19(5):440–448.
4. McKinlay A, Grace RC, Dalrymple-Alford JC, Anderson TJ, Fink J, Roger D. Neuropsychiatric problems in Parkinson's disease: comparisons between self and caregiver report. *Aging Ment Health.* 2008;12(5): 647–653.
5. Krishnamoorthy ES, Trimble MR. Prevalence, patterns, service needs, and assessment of neuropsychiatric disorders among people with epilepsy in residential care: validation of the Neuropsychiatric Inventory as a caregiver-rated measure of neuropsychiatric functioning in epilepsy. *Epilepsy Behav.* 2008;13(1):223–228.
6. Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology.* 1997;48(5 Suppl 6): S10–S16.
7. Connor DJ, Sabbagh MN, Cummings JL. Comment on administration and scoring of the Neuropsychiatric Inventory in clinical trials. *Alzheimers Dement.* 2008;4(6):390–394.
8. Kaufer DI, Cummings JL, Ketchel P, et al. Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. *J Neuropsychiatry Clin Neurosci.* 2000;12(2):233–239.
9. Wood S, Cummings JL, Hsu MA, et al. The use of the neuropsychiatric inventory in nursing home residents. Characterization and measurement. *Am J Geriatr Psychiatry.* 2000;8(1):75–83.
10. Kane RL, Radosevich DM. *Conducting Health Outcomes Research.* Sudbury, MN: Jones & Bartlett Learning; 2011.
11. Kang SJ, Choi SH, Lee BH, et al. Caregiver-Administered Neuropsychiatric Inventory (CGA-NPI). *J Geriatric Psychiatry Neurol.* 2004;17(1): 32–35.
12. Reisberg B, Auer SR, Monteiro IM. Behavioral Pathology in Alzheimer's Disease (BEHAVE-AD) Rating Scale. *Int Psychogeriatr.* 1996; 8 Suppl 3:301–308.
13. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56–62.
14. Leung VP, Lam LC, Chiu HF, Cummings JL, Chen QL. Validation study of the Chinese version of neuropsychiatric inventory (CNPI). *Int J Geriatr Psychiatry.* 2001;16:789–793.
15. Fernandez HH, Aarsland D, Fénelon G, et al. Scales to assess psychosis in Parkinson's disease: Critique and recommendations. *Mov Disord.* 2008;23(4):484–500.
16. Tractenberg RE, Singer CM, Cummings JL, Thal LJ. The Sleep Disorders Inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. *J Sleep Res.* 2003;12:331–337.
17. Nagahama Y, Okina T, Suzuki N, Matsuda M. The Cambridge Behavioral Inventory: validation and application in a memory clinic. *J Geriatry Psychiatry Neurol.* 2006;19:220–225.
18. Dechamps A, Jutand MA, Onifade C, Richard-Harston, Bourdel-Marchasson I. Co-occurrence of neuropsychiatric syndromes in demented and psychotic institutionalized elderly. *Int J Geriatr Psychiatry.* 2008; 23(11):1182–1190.
19. Zuidema SU, de Jonghe JF, Verhey FR, Koopmans RT. Neuropsychiatric symptoms in nursing home patients: factor structure invariance of the Dutch Nursing Home version of the Neuropsychiatric Inventory in different stages of dementia. *Dement Geriatr Cogn Disord.* 2007;24: 169–176.
20. Perrault A, Oremus M, Demers L, Vida S, Woflson C. Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of behavior and mood scales. *J Geriatr Psychiatry Neurol.* 2000;13(4):181–196.
21. Ismail Z, Emeremni MA, Houck PR, et al. A comparison of the E-BEHAVE-AD, NBRS, and NPI in quantifying clinical improvement in the treatment of agitation and psychosis associated with dementia. *Am J Geriatr Psychiatry.* 2013;21(1):78–87.
22. Cummings JL, McPherson S. Neuropsychiatric assessment of Alzheimer's disease and related dementias. *Aging (Milano).* 2001;13(3): 240–246.
23. Frisoni GB, Rozzini L, Gozzetti A, et al. Behavioral syndromes in Alzheimer's disease: description and correlates. *Dement Geriatr Cogn Disord.* 1999;10(2):130–138.
24. Kaufer D. Beyond the cholinergic hypothesis: the effect of metrifonate and other cholinesterase inhibitors on neuropsychiatric symptoms in Alzheimer's disease. *Dement Geriatr Cogn Disord.* 1998;9 Suppl 2: 8–14.
25. Mega MS, Cummings JL, Fiorello T, Gornbein J. The spectrum of behavioral changes in Alzheimer's disease. *Neurology.* 1996;46:130–135.
26. Morris JC, Cyrus PA, Orazem J, et al. Metrifonate benefits cognitive, behavioral, and global function in patients with Alzheimer's disease. *Neurology.* 1998;50(5):1222–1230.
27. Rocha FL, Hara C, Ramos MG, et al. An exploratory open-label trial of ziprasidone for the treatment of behavioral and psychological symptoms of dementia. *Dement Geriatr Cogn Disord.* 2006;22(5–6): 445–448.
28. Hama S, Yamashita H, Shigenobu M, et al. Depression or apathy and functional recovery after stroke. *Int J Geriatr Psychiatry.* 2007;22(10): 1046–1051.
29. Holmes C, Wilkinson D, Dean C, et al. The efficacy of donepezil in the treatment of neuropsychiatric symptoms in Alzheimer disease. *Neurology.* 2004;63(2):214–219.
30. Scripnikov A, Khomenko A, Napryeyenko O. Effects of Ginkgo biloba extract EGb 761 on neuropsychiatric symptoms of dementia: findings from a randomised controlled trial. *Wien Med Wochenschr.* 2007; 157(13–14):295–300.
31. Aalten P, Verhey FR, Boziki M, et al. Neuropsychiatric syndromes in dementia. Results from the European Alzheimer Disease Consortium: part I. *Dement Geriatr Cogn Disord.* 2007;24(6):457–463.
32. Litvan I, Paulsen JS, Mega MS, Cummings JL. Neuropsychiatric assessment of patients with hyperkinetic and hypokinetic movement disorders. *Arch Neurol.* 1998;55(10):1313–1319.
33. Blair M, Kertesz A, Davis-Faroque N, et al. Behavioural measures in frontotemporal lobar dementia and other dementias: the utility of the Frontal Behavioural Inventory and the Neuropsychiatric Inventory in a national cohort study. *Dement Geriatr Cogn Disord.* 2007;23: 406–415.
34. Fuh JL, Lam L, Hirono N, Senanarong V, Cummings JL. Neuropsychiatric Inventory workshop: behavioral and psychologic symptoms of dementia in Asia. *Alzheimer Dis Assoc Disord.* 2006;20(4): 314–317.
35. Chow TW, Liu CK, Fuh JL, et al. Neuropsychiatric symptoms of Alzheimer's disease differ in Chinese and American patients. *Int J Geriatr Psychiatry.* 2002;17(1):22–28.
36. Fuh JL, Wang SJ, Cummings JL. Neuropsychiatric profiles in patients with Alzheimer's disease and vascular dementia. *J Neurol Neurosurg Psychiatry.* 2005;76(10):1337–1341.

37. Gallo JL, Schmidt KS, Libon DJ. An itemized approach to assessing behavioral and psychological symptoms in dementia. *Am J Alzheimers Dis Other Demen.* 2009;24(2):163–168.

38. Leontjevas R, Van Hooren S, Mulders A. The Montgomery-Asberg Depression Rating Scale and the Cornell Scale for Depression in Dementia: a validation study with patients exhibiting early-onset dementia. *Am J Geriatr Psychiatry.* 2009;17(1):56–64.

39. Cohen-Mansfield J. Heterogeneity in dementia: Challenges and opportunities. *Alzheimer Dis Assoc Disord.* 1999;14(2):60–63.

40. Aalten P, Verhey FR, Boziki M, et al. Consistency of neuropsychiatric syndromes across dementias: results from the European Alzheimer Disease Consortium. Part II. *Dement Geriatr Cogn Disord.* 2008;25(1): 1–8.

41. Hirono N, Mori E, Yasuda M, et al. Factors associated with psychotic symptoms in Alzheimer's disease. *J Neurol Neurosurg and Psychiatry.* 1998;64(5):648–652.

42. Matsui T, Nakaaki S, Murata Y, et al. Determinants of the quality of life in Alzheimer's disease patients as assessed by the Japanese version of the Quality of Life-Alzheimer's disease scale. *Dement Geriatr Cogn Disord.* 2006;21(3):182–191.

43. Lam CL, Chan WC, Mok CC, Li S, Lam LC. Validation of the Chinese Challenging Behaviour Scale: clinical correlates of challenging behaviours in nursing home residents with dementia. *Int J Geriatr Psychiatr.* 2006;21(8):792–799.

44. Selbaek G, Kirkevold O, Sommer OH, Engedal K. The reliability and validity of the Norwegian version of the Neuropsychiatric Inventory, nursing home version (NPI-NH). *Int Psychogeriatr.* 2008;20(2): 375–382.

45. Zuidema SU, de Jonghe JF, Verhey FR, Koopmans RT. Predictors of neuropsychiatric symptoms in nursing home patients: influence of gender and dementia severity. *Int J Geriatr Psychiatry.* 2009;24(10): 1079–1086.

46. Lam LC, Tam CW, Lui VW, et al. Prevalence of very mild and mild dementia in community-dwelling older Chinese people in Hong Kong. *Int Psychogeriatr.* 2008;20(1):135–148.

47. Fuh JL, Liu CK, Mega MS, Wang SJ, Cummings JL. Behavioral disorders and caregivers' reaction in Taiwanese patients with Alzheimer's disease. *Int Psychogeriatr.* 2001;13(1):121–128.

48. Aalten P, De Vugt ME, Lousberg R, et al. Behavioral problems in dementia: a factor analysis of the neuropsychiatric inventory. *Dement Geriatr Cogn Disord.* 2003;15(2):99–105.

49. Mirakhur A, Craig D, Hart DJ, McLlroy SP, Passmore AP. Behavioural and psychological syndromes in Alzheimer's disease. *Int J Geriatr Psychiatry.* 2004;19(11):1035–1039.

50. Spalletta G, Baldinetti F, Buccione I, et al. Cognition and behaviour are independent and heterogeneous dimensions in Alzheimer's disease. *J Neurol.* 2004;251(6):688–695.

51. Matsumoto N, Ikeda M, Fukuhara R, et al. Neuropsychiatric Inventory Caregiver Distress Scale (NPI-D) and Japanese Version of the Neuropsychiatric Inventory-Brief Questionnaire Form (NPI-Q). Abstracts of the 20th Annual Meeting of the Japanese Psychogeriatric Society. *Psychogeriatrics.* 2005;5(4):A79–A80.

52. Kilmer RP, Demakis GJ, Hammond FM, Grattan KE, Cook JR, Kornev AA. Use of the neuropsychiatric inventory in traumatic brain injury: a pilot investigation. *Rehabil Psychol.* 2006;51(3):232–238.