Am J of Geriatric Psychiatry 33:5 (2025) 524–534

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ajgponline.org

Regular Research Article

Practical Assessment of Neuropsychiatric Symptoms: Updated Reliability, Validity, and Cutoffs for the Neuropsychiatric Inventory Questionnaire

David Andrés González, PhD, John-Christopher A. Finley, MS, Samantha Evy Schoeneman Patel, MD, Jason R. Soble, PhD

ARTICLE INFO

Article history:
Received August, 11 2024
Revised October, 23 2024
Accepted October, 23 2024

Key Words:
Behavioral and psychological symptoms of dementia
neuropsychiatric inventory questionnaire
neuropsychiatric symptoms
psychometrics
reliability
validation

ABSTRACT

Objectives: *To improve assessment of neuropsychiatric symptoms (NPS) by expanding the measurement properties of the Neuropsychiatric Inventory Questionnaire (NPI-Q).* ***Design:*** *Multicenter, longitudinal observational study.* ***Setting:*** *Several Alzheimer's Disease Research Centers (ADRCs).* ***Participants:*** *Individuals (*n = *45,274) who presented to an ADRC with a collateral and completed the NPI-Q.* ***Measurements:*** *The NPI-Q total severity score, four NPI-Q subscales, dementia stage, expert NPS rating, consensus rating of dementia syndrome, global cognitive screening, collateral rating of daily functioning, and self-rating of depression.* ***Results:*** *There was strong evidence of criterion validity with both dementia stage and expert NPS rating for the NPI-Q total severity index, which informed cutoffs and interpretive ranges. Furthermore, subscales had adequate classification of dementia syndromes and appropriate convergent relationships with cognition, daily functioning, and mood. There was good-to-excellent evidence of reliability for the NPI-Q total severity index over several years, and subscales had adequate-to-good reliability.* ***Conclusions:*** *This is the first study to provide empirically established cutoffs, interpretive ranges, and evidence of reliability over a period longer than a month on the NPI-Q and its subscales. This will improve assessment of NPS in clinical and research contexts.* ***Article Summary:*** *Neuropsychiatric symptoms of neurodegeneration are increasingly understood as early disease markers*

From the Department of Neurological Sciences (DAG, SESP), Rush University Medical Center, Chicago, IL, USA; Department of Psychiatry & Behavioral Sciences (JCAF), Northwestern University Feinberg School of Medicine, Chicago, IL, USA; and the Departments of Psychiatry & Neurology (JRS), University of Illinois College of Medicine, Chicago, IL, USA. Send correspondence and reprint requests to David A. Gonzalez, Parkinson's Disease & Movement Disorders Care Program, 1725 W. Harrison St., Ste. 755, Chicago, IL, 60612. e-mail: david_a_gonzalez@rush.edu

https://doi.org/10.1016/j.jagp.2024.10.014

with tremendous functional impact later in disease, but are often missed or misdiagnosed. The most common measure of these symptoms, the Neuropsychiatric Inventory Questionnaire (NPI-Q), does not have clinically actionable guidance, which this article provided. We established cutscores for several conditions and test-retest reliability over longer periods for the total score and subscales using a multicenter database. (Am J Geriatr Psychiatry 2025; 33:524–534)

Highlights

- **What is the primary question addressed by this study?**
  How can clinicians and researchers fully leverage the Neuropsychiatric Inventory Questionnaire (NPI-Q) to reliably and validly assess neuropsychiatric symptoms?
- **What is the main finding of this study?**
  The NPI-Q has adequate reliability, even over periods of several years and validly corresponds with criterion ratings of neuropsychiatric symptoms and all-cause dementia staging. Furthermore, certain subscales can help distinguish particular dementia syndromes (e.g., Lewy body dementia), but not others (e.g., Alzheimer's); these subscales (as opposed to the total score or individual items) demonstrated incremental validity when added to machine learning algorithms to classify from a multitude of dementia syndromes.
- **What is the meaning of the finding?**
  These findings—the first to establish reliability, interpretive guidance, and empirically-derived cutoffs for the NPI-Q—improve surveillance of neuropsychiatric symptoms by clinicians and researchers; furthermore, the subscale validation lays groundwork for incorporation into digital workflows to improve accuracy of diagnoses.

OBJECTIVE

The impact of dementia is clear: 6.9 million Americans are living with Alzheimer's dementia (AD) alone, costing our healthcare system $360 billion. This impact is expected to balloon to an estimated 13.8 million cases by 2060.[1] Neuropsychiatric symptoms (NPS) are common in all dementia syndromes, considered hallmark to the diagnosis of some dementia syndromes[2–4] (e.g., frontotemporal dementia [FTD], dementia with Lewy Bodies]), and may present before the development of any dementia, leading to development of new criteria emphasizing NPS as prodromal symptoms for dementia.[5,6] However, up to 71% of individuals initially presenting with NPS will be misdiagnosed as having a primary psychiatric condition.[7] Further, NPS result in greater functional impairment and worse quality of life among those with dementia, leading to earlier institutional placement.[8–17] Given its importance in prodromal stratification, diagnosis, course, and outcome, NPS assessment is increasingly recognized as a critical aspect of dementia research and clinical services.[18] However, unlike the plethora of well-validated tests that exist for cognitive assessment within the context of dementia, comparatively fewer parallel instruments exist for objective assessment of NPS.

Among NPS measures, the Neuropsychiatric Inventory Questionnaire (NPI-Q) is most commonly used.[19] It is a brief, informant-rated version of the NPI interview designed to assess various types of NPS, including hallucinations, delusions, depression, anxiety, disinhibition, agitation, elation, apathy, irritability, aberrant motor behavior, and sleep and appetite disturbance. Surprisingly, despite its widespread use, there is minimal guidance on how to interpret the NPI-Q. Most concerningly, no study has identified interpretive ranges for NPS severity or established cutoffs indicative of dementia stages or syndromes.[19–21] Furthermore, interpretation of the NPI-Q subscales remains unclear. Although there has been little agreement on subscale use,[20,22–30] a recent study comparing different proposed subscales found that psychosis,

mood disturbance, behavioral activation, and behavioral suppression/somatic disturbance had the strongest evidence across various dementia populations.[24] Lastly, it is unknown whether the NPI-Q can reliably assess NPS over longer timeframes,[21] such as 6−12 months, since test-retest reliability for the NPI-Q has only been assessed over periods up to one month.[20,23,31−37]

Without addressing these gaps, clinicians and researchers using the NPI-Q may inaccurately assess and characterize NPS among individuals with neurodegenerative conditions. Inaccurate or inconsistent assessment can compromise patient care and hinder the advancement of therapeutic strategies which, in turn, may have downstream effects on the financial and emotional costs of dementia. Inaccurate or inconsistent assessment may also pose numerous methodological problems when the NPI-Q is used for research purposes. Virtually all healthcare practitioners and scientists who work with patients with dementia would benefit from empirical guidelines on how to interpret the NPI-Q. A comprehensive analysis of the NPI-Q could also facilitate data harmonization as well as the incorporation of the measure into electronic medical records and automated algorithms that are being developed to inform diagnoses, prognoses, and interventions. As such, the purpose of this study was to establish the reliability and validity of the NPI-Q and provide practical interpretation of its cutoffs, subscales, and ability to assess NPS over time.

METHOD

Study Design, Setting, and Participants

Individuals who had a knowledgeable collateral (e.g., family member) present to complete an NPI-Q rating during an initial Alzheimer's Disease Research Center (ADRC) visit between June 2005 and June 2023 were selected from the National Alzheimer's Coordinating Center (NACC). No other exclusion criteria were applied, which resulted in 45,274 individuals from 46 ADRCs. Of these, a subset of 29,679 returned with a collateral completing a second NPI-Q rating, which were used to evaluate reliability. These individuals returned, on average, after an interval of 4.84 years (SD = 3.71, range = 0.14−17.68). Although NACC requires certain data from ADRCs for the Uniform Data Set (UDS), reasons for missing data may not be fully documented. Furthermore, there are differences in recruitment strategies (e.g., clinic referral, community-based) and each ADRC has its own IRB approval.

MEASURES

Neuropsychiatric Symptoms

NPS was assessed via the NPI-Q. In brief, the NPI-Q is a 12-item, collateral-rated measure of NPS, with good internal consistency and test-retest reliability over short periods.[20,21]

A criterion rating of behavioral disturbance and NPS was completed with the newer Clinical Dementia Rating (CDR®) Dementia Staging Instrument plus NACC FTD Behavior and Language Domains.[38,39] This rating of behavioral disturbance (e.g., reduced awareness, disinhibition, apathy, interpersonal disengagement, affective lability, poor empathy, changed eating habits) and the extent to which these impact social relationships, is scored independently of other UDS data, such as NPI-Q. The rating produces five levels of behavioral disturbance and NPS: unimpaired, questionable, mild, moderate, and severe.

Self-reported depression symptoms were assessed via the Geriatric Depression Scale-Short Form (GDS-SF), which has been well-validated in older adult and neurodegenerative samples.[40,41] Higher scores indicate more depression symptoms.

Cognitive and Functional Assessment

Overall cognitive-dementia stage was determined using the gold standard CDR® Dementia Staging Instrument. This scale is based on evaluations from the patient and collateral sources, and includes ratings across six domains: memory, orientation, judgment/problem-solving, community functioning, home/hobby functioning, and personal care functioning.[42] This study used established CDR interpretive guidelines to create five cognitive-functional stages: cognitively unimpaired, mild cognitive impairment, mild dementia, moderate dementia, and severe dementia.[43] The CDR is rated independently of other UDS data elements, which precludes criterion

contamination with any of the other measures used in this study.

Cognition was measured using a brief global cognitive screener, the Mini Mental Status Examination (MMSE) or the Montreal Cognitive Assessment (MoCA). The UDS transitioned from including the MMSE to the MoCA in Version 3 (UDS3), and a published equipercentile linking study was used to convert MMSE scores into estimated MoCA scores.[44,45] Higher scores indicate more intact cognition.

Difficulties with instrumental activities of daily living (IADLs) were assessed via the Functional Activities Questionnaire (FAQ),[46] which has been empirically established in older adult and neurodegenerative samples.[47] Higher scores indicate greater functional difficulties.

Dementia Syndromes

For individuals with cognitive impairment, clinicians at each ADRC used established UDS3[48] updated diagnostic criteria and all available information to rate presumptive AD.[49] The same approach and UDS3 update was used to rate Lewy Body dementia (LBD)[3,50] and behavioral variant FTD (bvFTD).[4,51] Other presumed causes of impairment include primary progressive aphasia (PPA), motor FTD, other neurologic, other medical, vascular, or psychiatric and the criteria for these presumed etiologies are listed in NACC documentation (naccdata.org).

Statistical Analyses

Following empirical consensus recommendations, measurement properties were determined by criterion validity, structural validity, content validity, internal consistency, and test-retest reliability.[52]

Validity. Criterion validity (the extent to which scores correspond with gold standard measures) was evaluated in comparison with the CDR global dementia staging and behavioral disturbance ratings. First, a nonparametric Kruskal-Wallis test (χ^2_{KW}) determined group differences among these stages. If the effect size (rank epsilon squared, ε^2) was large (see guidelines below), receiver operating characteristic curve area under the (ROC-AUC) analysis determined optimal cutoffs for accurately classifying criterion diagnoses.

Structural validity (the extent to which scores correspond with the construct's latent dimensions) was previously suggested with a study that compared various factor structures and found a recommended four-factor structure.24 However, the extent to which these factors or subscales correspond with related constructs has not yet been evaluated. Thus, nonparametric Spearman correlations (r_s) assessed one facet of construct validity (the extent to which scores converge with related constructs, diverge from disparate ones, and correspond with hypotheses about group differences) of the NPI-Q subscales. We used general cognition, IADL functioning, and depression symptoms as related and disparate constructs in these construct analyses. To further evaluate other facets of construct validity, we tested the presence and extent of difference on the NPI-Q and subscales amongst dementia syndromes using nonparametric Mann-Whitney *U* tests. If subscales differences amongst diagnostic groups were large (based on rank-biserial correlation, r_{rb}, effect size; see guideline reference below), we calculated optimal cutoffs using ROC-AUC analysis.

To incrementally integrate criterion, structural, and construct validity evaluation, we implemented random forest classification algorithms determining whether NPI-Q subscales offer unique predictive value above demographic prediction. For random forests, we used a nine-category dementia syndrome classification as the outcome (cognitively unimpaired, AD, LBD, bvFTD, PPA, motor FTD, other neurologic, other medical, vascular, psychiatric). We evaluated the performance of four predictor sets: 1) demographics, 2), demographics + total NPI-Q, 3) demographics + four NPI-Q subscales, 4), demographics + 12 NPI-Q items.

Reliability. Internal consistency (degree of item interrelatedness) for total severity and subscales was evaluated using McDonald's omega (ω) for the NPI-Q administered at the first visit. Test-retest reliability (the extent to which scores are consistent over time) for total severity and subscales was evaluated using intraclass correlation coefficients (*ICC*) across the first and second visit. Multilevel modeling was used within a generalizability theory framework[53] to determine the extent to which score error/variability in the total severity score was attributable to different response across participants (inter-person variance), particular items within the scale (inter-item variance),

time points (inter-session variance), or interactions among these three components.

Software and Interpretive Rules of Thumb. Analyses were mostly conducted in the R environment (version 4.3.3). The *tidyverse* (version 2.0.0) and *Hmisc* (version 5.1-2) packages were used to prepare and summarize data. Validity analyses were supplemented with *correlation* (version 0.8.4), *effectsize* (version 0.8.8), *pROC* (version 1.18.5) packages, as well as JASP (version 0.8.3) for random forest modeling. Reliability analyses were supplemented with *psych* (version 2.4.3) package. Reliability and effect size coefficients were interpreted according to recommendations by Ben-Shachar and colleagues in the *effectsize* reference manual,[54] and interpretation is integrated into results text (i.e., negligible, small, medium, large). All tests were two-tailed unless otherwise specified. With a large sample size, we expected many tests to be statistically significant and gave weight to interpreting effect sizes and their confidence intervals.

RESULTS

The inclusion criteria (i.e., having a collateral complete an NPI-Q during an initial ADRC visit) resulted in 45,274 individuals. Of these, 56.7% were women, 37.1% were in their 70s (12.2% 50s or younger, 29.2% 60s, and 21.5% 80s or older; M_{age} = 71.19, SD_{age} = 10.37), 56.8% had a bachelor's degree or higher (7.2% had some high school or less, 36% had high school diploma to some college; $M_{years\ of\ education}$ = 15.19, $SD_{years\ of\ education}$ = 3.41), and 80.4% self-classified as white (13% as Black American, 8.3% as Hispanic, 2.8% as Asian). Descriptives for the retest subsample are presented in Supplementary Tables.

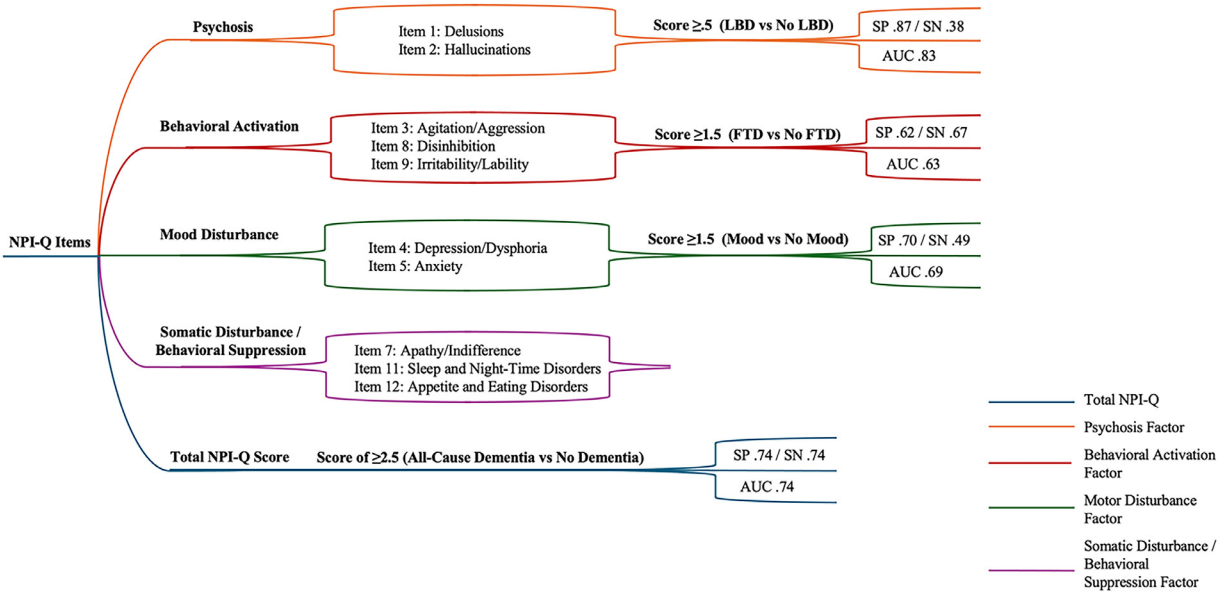
The most common cognitive stage was mild cognitive impairment (39.6%), most common cognitive syndrome was presumed Alzheimer's (38.4%), and most participants had no behavioral disturbance on the criterion rating (78.6%). This general lack of disturbance was observed in the distribution of NPI-Q total severity score (M = 3.07, SD = 4.35, range = 0−36) as well as psychosis (M = 0.18, SD = 0.67, range = 0−6), mood (M = 0.79, SD = 1.24, range = 0−6), behavioral activation (M = 0.87, SD = 1.58, range = 0−9), and somatic disturbance/behavioral suppression (M = 0.99, SD = 1.60, range = 0−9) subscales.

Validity

Criterion validity. With regard to cognitive stage criterion, all comparisons were statistically significant and *large* differences were observed on the NPI-Q total severity (χ^2_{KW} = 15534, *df* = 4, p < 0.001, ε^2 = 0.34, 95% CI [0.34, 0.35]) and subscale scores (all comparisons statistically significant; ε^2 range = 0.15−0.24) across CDR stages. Regarding criterion behavioral disturbance ratings, a similar pattern emerged where all comparisons were statistically significant and *large* for NPI-Q total severity (χ^2_{KW} = 8234.4, *df* = 4, p < 0.001, ε^2 = 0.27, 95% CI [0.26, 0.28]) and medium-to-large for subscales (all comparisons statistically significant; ε^2 range = 0.10−0.22). Given the largest differences with staging of cognitive impairment and behavioral disturbance were both found in the NPI-Q total severity score, we included the total score in ROC-AUC analyses, which were favorable for cognitive stage (demented vs. not; *AUC* = 0.80, 95% CI [0.80, 0.81]) and behavioral disturbance (clearly disturbed vs. not; *AUC* = 0.87, 95% CI [0.86, 0.87]). Youden's J analysis revealed an optimal cutoff of ≥3 for dementia (sensitivity = 0.74, specificity = 0.74) and ≥4 for clear behavioral disturbance (sensitivity = 0.76, specificity = 0.82). These cutoffs were integrated with distributional properties of the NPI-Q at different levels of disturbance (e.g., standard deviation of those with no disturbance, median of those with moderate-to-severe disturbance) to create interpretive ranges: 0−2 = normal, 3−7 = mild global disturbance, ≥8 = significant global disturbance (See Figure 1). Descriptives for the NPI-Q and its subscales among stages of impairment and disturbance are presented in Supplemental Tables 1 and 2.

Structural and Construct Validity. Subscales had positive, *medium*-strength correlations with each other, suggesting they capture related but nonredundant information. The NPI-Q total severity had a negative, *medium*-strength relationship with cognition (r_s = −0.44, 95% CI [−0.45, −0.43], p < 0.001), as did all subscales to a similar degree (r_ss = −0.30 to 0.36). There was a positive, *large*-strength correlation with functional/IADL dependence (r_s = 0.66, 95% CI [0.65,.66], p < 0.001) that varied widely amongst subscales (r_ss = 0.40 - 0.56), with psychosis having the weakest relationship and somatic disturbance/behavioral suppression having the strongest relationships. There was also a positive, *medium*-strength

FIGURE 1. Syndrome-specific cutoffs for the NPI-Q total score and subscales. ***Note.*** **NPI-Q = Neuropsychiatric Inventory Questionnaire; LBD = Lewy Body Dementia; FTD = Behavioral Variant Frontotemporal Dementia.**



relationship with depression symptoms (r_s = 0.38, 95% CI [0.37,.39], p < 0.001), which most strongly manifested on the mood subscale (r_s = 0.37) and to a lesser degree with others (r_ss = 0.14−0.31).

Compared to other syndromes, there did not appear to be a unique pattern of relationships with AD. The psychosis subscale was uniquely associated with LBD, although the somatic disturbance/behavioral suppression (including sleep) subscale was nonuniquely related. The behavioral activation subscale was uniquely associated with bvFTD, although the somatic disturbance/behavioral suppression subscale was also related. And the mood subscale was uniquely associated with psychiatric disturbance as primary cause of impairment. Subscale cutoffs by syndrome are presented in Figure 2. NPI-Q and subscale values are presented in Supplemental Table 3.

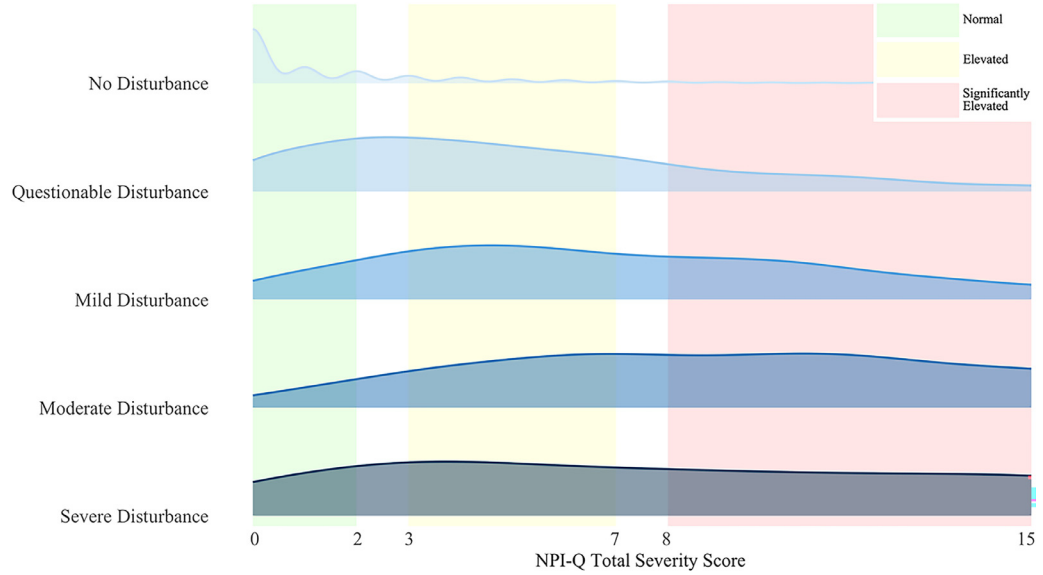
Integrated Random Forrest Performance. When all four subscales were integrated into the random forest classifier, comparing all nine dementia syndromes at once, there was adequate classification accuracy (AUC = 0.67), which significantly reduced when only the total severity was entered (AUC = 0.63), and further reduced when only demographics were included (AUC = 0.53). In contrast, adding all items into the classifier did not result in a significant improvement (AUC = 0.68).

Reliability

The NPI-Q total severity had *good* internal consistency (total ω = 0.86) at a single time point, and subscales had mostly adequate internal consistency except for a questionable psychosis value (ω range = 0.59−0.74). The temporal consistency was *good* for the total severity score (*ICC* = 0.79, 95% CI [0.79, 0.80], F = 4.8, df_1 = 29,678, df_2 = 29,679, p < 0.001) and adequate for all subscales (*ICC* range = 0.63−0.73). Across time points, the generalizability coefficient was *excellent* for the total severity score (g = 0.91), and multilevel generalizability analyses revealed that most error variance was attributable to a participant × item interaction (i.e., individual differences on particular items), closely followed by a main participant effect (i.e., individual differences in global severity responding), and then a participant × time interaction (i.e., individual temporal trajectories).

Given that the retest interval was quite large, and that clinical encounters may often occur annually or sooner, we reran test-retest analyses on a smaller subset that returned for their second visit within a year

FIGURE 2. Denisty plot of NPI-Q interpretive ranges and scores by CDR behavioral disturbance ratings. ***Note.*** **NPI-Q = Neuropsychiatric Inventory Questionnaire.**



(n = 5,615), and found similar values for the total severity score (ICC = 0.78, 95% CI [0.77, 0.79], F = 4.5, df_1 = 5,614, df_2 = 5,615, p < 0.001) and all subscales (ICC range = 0.63−0.71).

CONCLUSIONS

The NPI-Q was originally designed to help quantify the extent of NPS in individuals with dementia in a time-efficient fashion. However, given the clear relationship of NPS with dementia presence, severity, and syndrome, it is surprising that NPI-Q cutoffs and interpretive guidance have not been offered before. As such, the aim and major contribution of this study was to consolidate, update, and expand the analysis of the NPI-Q's measurement properties to improve NPS assessment. To this end, we found good evidence of criterion validity using both cognitive-dementia staging and independent expert ratings of NPS, and were able to generate cutoffs and interpretive suggestions. Notably, these cutoffs should not supersede clinical judgement, and there may be critical item endorsements, such as mild endorsement of hallucinations, that are worth exploring, even if a score is beneath cutoffs. Further, we found evidence of construct validity in the subscales with theoretically consistent convergent and divergent relationships with external ratings of cognition, daily functioning, and depression. Finally, in using multiple metrics of reliability over two time points spread over many years, we found good evidence for both the total severity index and subscales. Although the NPI-Q should not be used on its own to diagnose dementia or a particular dementia syndrome, our results suggest that it may be a useful part of the previsit workflow to empirically inform clinicians about differentials and symptom areas that warrant further assessment. This may, in turn, reduce the likelihood of misdiagnosis which can be high in psychiatric settings.[7]

Another important contribution of this study is further elaboration of the NPI-Q's structural validity. Although a recent study determined the best subscale structure, the recommended subscales have not been subject to further reliability analysis and validation.[24] The current study further supports use of these suggested factor structure (psychosis, mood, behavioral activation, somatic disturbance/suppression) as subscales. Not only was there appropriate reliability (save for a borderline result for psychosis), but there

was also theoretically consistent convergence and divergence with related constructs. Of the subscales, it appears that somatic disturbance/behavioral suppression has the largest relationship with daily functioning dependence. As this subscale includes apathy, one could hypothesize that decline in goal-directed behavior might require increasing structure and support. In addition to a lack of initiation, apathy correlates with cognitive-dementia severity quite strongly, such that this subscale could also be a marker for broader neurobehavioral dysfunction.[55] Overall, these associations support the use of this subscale structure in the literature.

As further evidence of structural validity, each subscale on its own had some strong criterion relationships with a syndrome (e.g., psychosis subscale with LBD), and when all subscales and syndromes were concurrently entered into a random forest classifier, there was adequate classification accuracy. This suggests that future efforts to create informed differential diagnoses from medical records could integrate NPS with other independent data points (e.g., IADL assistance, cognitive concerns, medications) to create increasingly accurate predictions and appropriate referrals. Since classification algorithms use feature reduction that is idiosyncratic to the sample and since adding items individually did not improve performance, future efforts should incorporate the proposed subscales as features instead of just items or a total severity score.

A significant contribution of our study is that it informs clinicians and researchers in monitoring NPS longitudinally and in response to treatment. Despite the importance of tracking NPS over long periods, the NPI-Q had not been subject to test-retest analyses longer than one month. The current findings suggest that the NPI-Q is a robust measure of NPS over the span of multiple years. These findings may be integrated with literature showing relations of NPI-Q to biomarkers and pathology.[56,57]

Alongside these strengths there are limitations to be considered. One lies in the measure itself including multipronged and "double-barreled" questions (e.g., the sleep item conflates prompts related to insomnia, hypersomnia, and parasomnia), and its design to focus on those with suspected dementia. Newer scales, such as the Mild Behavioral Impairment Checklist, that disaggregate NPS and do not frame questions in the context of obvious cognitive impairment may become commonly used in future research. However, the NPI-Q is still widely disseminated, and our aim is to improve the yield from current tools as we bridge to implementation of new tools in advancing the science of NPS. Another weakness lies in the sampling, such that most individuals attending ADRCs are nonminoritized and with higher education, which constrains generalizability to minoritized and disadvantaged groups. The UDS also does not have quantified social determinants of health or specific criterion scales for other facets of NPS (aside from depression) to validate the NPI-Q. These gaps deserve further study. Looking ahead, these results pave the way for creating reliable change indices and cutoffs for determining clinically meaningful changes over different periods of time.

At present, we hope the current results expand our armamentarium so as to improve our understanding of NPS and quality of life of individuals living with their complications.

AUTHOR CONTRIBUTIONS

DAG—Conceptualization, Methodology, Formal analysis, Data curation, Writing—original draft, Project administration, Supervision.

JCAF—Visualization, Validation, Writing review and editing, Methodology, Writing—original draft.

SESP—Writing – review and editing, Conceptualization.

JRS—Conceptualization, Methodology, Supervision, Writing—review and editing.

DATA STATEMENT

Data are available from the National Alzheimer's Coordinating Center (https://naccdata.org/) and syntax for the current analysis can be found on an Open Science Framework repository: https://osf.io/unjwz/?view_only=f1a533670bc94a3ca152062e9aa849cc

Portions of this manuscript's analyses were performed on earlier NACC data freezes, and were presented at the following scientific meetings:

- *Gonzalez, C., Obolsky, M.A., Kowalczyk, K., Soble, J.R., and González, D. A. (2024, February). Exploring the Neuropsychiatric Inventory Questionnaire domains across diagnostic categories. [Poster presentation] International Neuropsychological Society 52nd Annual North American Meeting, New York, NY.*

- *Gonzalez, C., Obolsky, M.A., Kowalczyk, K., Soble, J.R., and González, D. A. (2024, February). Investigating a four-factor behavioral and neuropsychiatric model for assessing the severity of dementia. [Poster presentation] International Neuropsychological Society 52nd Annual North American Meeting, New York, NY.*
- *Tabet, M., Jennette, K.J., Soble, J.R., González, D.A. (2021, May)* Classification accuracy of the Neuropsychiatric Inventory -Questionnaire (NPI-Q) *[Poster presentation]. American Psychiatric Association Annual Meeting, Los Angeles, CA.*

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors did not use any AI-adjacent technology in the writing process.

DISCLSOURES

The authors do not declare any conflicts of interest directly relevant to this manuscript. Other funding is listed below:

DAG—Salary is partially funded by grants from the Alzheimer's Association (AARG-22-924771) and National Institutes of Health (1OT2HL156812-01, R01NS125294-02), and he has received consulting fees from EPI-Q, Inc.

J-CAF—No conflicts to report.

SESP—No conflicts to report.

JRS—No conflicts to report.

ACKNOWLEDGMENTS

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jagp.2024.10.014.

References

1. Alzheimer's Association: 2024 Alzheimer's disease facts and figures. Alzheimers Dement J Alzheimers Assoc 2024; 20(5):3708–3821;doi:10.1002/alz.13809
2. Steinberg M, Shao H, Zandi P, et al: Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. Int J Geriatr Psychiatry 2008; 23(2):170–177; doi:10.1002/gps.1858
3. McKeith IG, Boeve BF, Dickson DW, et al: Diagnosis and management of dementia with Lewy bodies. Neurology 2017; 89(1):88–100;doi:10.1212/WNL.0000000000004058
4. Rascovsky K, Hodges JR, Knopman D, et al: Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain J Neurol 2011; 134(Pt 9):2456–2477;doi:10.1093/brain/awr179
5. Ismail Z, Smith EE, Geda Y, et al: Neuropsychiatric symptoms as early manifestations of emergent dementia: provisional diagnostic criteria for mild behavioral impairment. Alzheimers Dement 2016; 12(2):195–202;doi:10.1016/J.JALZ.2015.05.017
6. McKeith IG, Ferman TJ, Thomas AJ, et al: Research criteria for the diagnosis of prodromal dementia with Lewy bodies. Neurology 2020; 94(17):743–755;doi:10.1212/WNL.0000000000009323

7. Zapata-Restrepo L, Rivas J, Miranda C, et al: The psychiatric misdiagnosis of behavioral variant frontotemporal dementia in a Colombian sample. front neurol 2021; 12:729381;doi:10.3389/fneur.2021.729381
8. Babulal GM, Chen L, Murphy SA, et al: Neuropsychiatric symptoms and Alzheimer disease biomarkers independently predict progression to incident cognitive impairment. Am J Geriatr Psychiatry Off J Am Assoc Geriatr Psychiatry 2023; 31(12):1190–1199;doi:10.1016/j.jagp.2023.07.012
9. Mallo SC, Patten SB, Ismail Z, et al: Does the Neuropsychiatric Inventory predict progression from mild cognitive impairment to dementia? A systematic review and meta-analysis. Ageing Res Rev 2020; 58:101004;doi:10.1016/J.ARR.2019.101004
10. Peters ME, Schwartz S, Han D, et al: Neuropsychiatric symptoms as predictors of progression to severe Alzheimer's dementia and death: the Cache County Dementia Progression Study. Am J Psychiatry 2015; 172(5):460–465;doi:10.1176/appi.ajp.2014.14040480
11. Balestreri L, Grossberg A, Grossberg GT: Behavioral and psychological symptoms of dementia as a risk factor for nursing home placement. Int Psychogeriatr 2000; 12(S1):59–62;doi:10.1017/S1041610200006773
12. Wiener PK, Kiosses DN, Klimstra S, et al: A short-term inpatient program for agitated demented nursing home residents. Int J Geriatr Psychiatry 2001; 16(9):866–872;doi:10.1002/gps.437
13. Fischer CE, Ismail Z, Schweizer TA: Delusions increase functional impairment in Alzheimer's disease. Dement Geriatr Cogn Disord 2012; 33(6):393–399;doi:10.1159/000339954
14. Fischer CE, Ismail Z, Schweizer TA: Impact of neuropsychiatric symptoms on caregiver burden in patients with Alzheimer's disease. Neurodegener Dis Manag 2012; 2(3):269–277;doi:10.2217/nmt.12.19
15. Kapustin D, Zarei S, Wang W, et al: Neuropsychiatric symptom burden across neurodegenerative disorders and its association with function. Can J Psychiatry Rev Can Psychiatr 2023; 68(5):347–358;doi:10.1177/07067437221147443
16. Aarsland D, Brønnick K, Ehrt U, et al: Neuropsychiatric symptoms in patients with Parkinson's disease and dementia: frequency, profile and associated care giver stress. J Neurol Neurosurg Psychiatry 2007; 78(1):36–42;doi:10.1136/JNNP.2005.083113
17. Karttunen K, Karppi P, Hiltunen A, et al: Neuropsychiatric symptoms and quality of life in patients with very mild and mild Alzheimer's disease. Int J Geriatr Psychiatry 2011; 26(5):473–482;doi:10.1002/gps.2550
18. Ismail Z, Agüera-Ortiz L, Brodaty H, et al: The mild behavioral impairment checklist (MBI-C): a rating scale for neuropsychiatric symptoms in pre-dementia populations. J Alzheimers Dis JAD 2017; 56(3):929–938;doi:10.3233/JAD-160979
19. Cummings J: The neuropsychiatric inventory: development and applications. J Geriatr Psychiatry Neurol 2020; 33(2):73–84;doi:10.1177/0891988719882102
20. 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;doi:10.1176/JNP.12.2.233
21. González DA, Obolsky MA, Amitrano N, et al: The neuropsychiatric inventory questionnaire. In: Martin CR, Patel VB, Preedy VR, Rajendram R, eds. Handbook of the Behavior and Psychology of Disease, Switzerland: Springer Nature, 2025:1–28;doi:10.1007/978-3-031-32046-0_107-1
22. De Vito AN, Calamia M, Weitzner DS, et al: Examining differences in neuropsychiatric symptom factor trajectories in empirically derived mild cognitive impairment subtypes. Int J Geriatr Psychiatry 2018; 33(12):1627–1634;doi:10.1002/GPS.4963
23. Feghali Y, Fares Y, Abou Abbas L: Assessment of neuropsychiatric symptoms in dementia: validity and reliability of the Lebanese version of the neuropsychiatric inventory questionnaire. Appl Neuropsychol Adult 2021; 28(5):588–595;doi:10.1080/23279095.2019.1670182
24. González DA, Resch ZJ, Obolsky MA, et al: Comparing behavioral and psychological symptom structures on the neuropsychiatric inventory questionnaire. Psychol Assess 2023; 35(6):522–532;doi:10.1037/pas0001230
25. Sayegh P, Knight BG: Functional assessment and neuropsychiatric inventory questionnaires: measurement invariance across Hispanics and Non-Hispanic Whites. Gerontologist 2014; 54(3):375–386
26. Siafarikas N, Alnæs D, Monereo-Sanchez J, et al: Neuropsychiatric symptoms and brain morphology in patients with mild cognitive impairment and Alzheimer's disease with dementia. Int Psychogeriatr 2021; 33(11):1217–1228;doi:10.1017/S1041610221000934
27. Siafarikas N, Selbaek G, Fladby T, et al: Frequency and subgroups of neuropsychiatric symptoms in mild cognitive impairment and different stages of dementia in Alzheimer's disease. Int Psychogeriatr 2018; 30(1):103–113;doi:10.1017/S1041610217001879
28. Thakur B, Alvarado L, Dodoo C, et al: Ethnic differences between Hispanics and Non-Hispanic Whites in neuropsychiatric symptoms predict conversion to mild cognitive impairment. J Geriatr Psychiatry Neurol 2021; 34(6):622–631;doi:10.1177/0891988720957087
29. Travis Seidl JN, Massman PJ: Cognitive and functional correlates of NPI-Q scores and symptom clusters in mildly demented Alzheimer patients. Alzheimer Dis Assoc Disord 2016; 30(2):145–151;doi:10.1097/WAD.0000000000000104
30. Trzepacz PT, Saykin A, Yu P, et al: Subscale validation of the Neuropsychiatric Inventory Questionnaire: comparison of Alzheimer's Disease Neuroimaging Initiative and National Alzheimer's Coordinating Center cohorts. Am J Geriatr Psychiatry 2013; 21(7):607–622;doi:10.1016/J.JAGP.2012.10.027
31. Boada M, Cejudo JC, Tàrraga L, et al: [Neuropsychiatric inventory questionnaire (NPI-Q): Spanish validation of an abridged form of the Neuropsychiatric Inventory (NPI)]. Neurol Barc Spain 2002; 17(6):317–323
32. Camozzato AL, Godinho C, Kochhann R, et al: Validity of the Brazilian version of the Neuropsychiatric Inventory Questionnaire (NPI-Q). Arq Neuropsiquiatr 2015; 73(1):41–45;doi:10.1590/0004-282X20140177
33. Kim HJ, Choi KH, Kim SH, et al: Validation study of the Korean version of the brief clinical form of the neuropsychiatric inventory. Dement Geriatr Cogn Disord Extra 2016; 6(2):214–221;doi:10.1159/000445828
34. Mao HF, Kuo CA, Huang WN, et al: Values of the minimal clinically important difference for the neuropsychiatric inventory questionnaire in individuals with dementia. J Am Geriatr Soc 2015; 63(7):1448–1452;doi:10.1111/jgs.13473
35. Matsumoto N, Ikeda M, Fukuhara R, et al: [Validity and reliability of the Japanese version of the Neuropsychiatric Inventory Caregiver Distress Scale (NPI D) and the Neuropsychiatric Inventory Brief Questionnaire Form (NPI-Q)]. No To Shinkei 2006; 58(9):785–790
36. Wong A, Cheng ST, Lo ESK, et al: Validity and reliability of the neuropsychiatric inventory questionnaire version in patients with stroke or transient ischemic attack having cognitive

impairment. J Geriatr Psychiatry Neurol 2014; 27(4):247–252; doi:10.1177/0891988714532017
37. Zepeda MUP, Guerrero JAR, Carrasco OR, et al: P3-038: Validation of the neuropsychiatric inventory questionnaire in a group of Mexican patients with dementia. Alzheimers Dement 2008; 4(4S_Part_16):T527–T528;doi:10.1016/j.jalz.2008.05.1601
38. Knopman DS, Weintraub S, Pankratz VS: Language and behavior domains enhance the value of the clinical dementia rating scale. Alzheimers Dement J Alzheimers Assoc 2011; 7(3):293–299; doi:10.1016/j.jalz.2010.12.006
39. Miyagawa T, Brushaber D, Syrjanen J, et al: Utility of the global CDR® plus NACC FTLD rating and development of scoring rules: data from the ARTFL/LEFFTDS Consortium. Alzheimers Dement 2020; 16(1):106–117;doi:10.1002/alz.12033
40. Yesavage JA, Brink TL, Rose TL, et al: Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982; 17(1):37–49;doi:10.1016/0022-3956(82)90033-4
41. Krishnamoorthy Y, Rajaa S, Rehman T: Diagnostic accuracy of various forms of geriatric depression scale for screening of depression among older adults: systematic review and meta-analysis. Arch Gerontol Geriatr 2020; 87:104002;doi:10.1016/j.archger.2019.104002
42. Morris JC: Clinical Dementia Rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr 1997; 9:173–176;doi:10.1017/S1041610297004870
43. O'Bryant SE, Lacritz LH, Hall J, et al: Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the National Alzheimer's Coordinating Center database. Arch Neurol 2010; 67(6):746–749;doi:10.1001/archneurol.2010.115
44. Weintraub S, Salmon D, Mercaldo N, et al: The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. Alzheimer Dis Assoc Disord 2009; 23(2):91–101; doi:10.1097/WAD.0b013e318191c7dd
45. Weintraub S, Besser L, Dodge HH, et al: Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer Dis Assoc Disord 2018; 32(1):10;doi:10.1097/WAD.0000000000000223
46. Pfeffer R, Kurosaki T, Harrah C, et al: Measurement of functional activities in older adults in the community. J Gerontol 1982; 37(3):323–329;doi:10.1093/GERONJ/37.3.323
47. González DA, Gonzales MM, Resch ZJ, et al: Comprehensive evaluation of the Functional Activities Questionnaire (FAQ) and its reliability and validity. Assessment 2022; 29(4):748–763; doi:10.1177/1073191121991215
48. McKhann GM, Knopman DS, Chertkow H, et al: The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7(3):263–269;doi:10.1016/J.JALZ.2011.03.005
49. McKhann G, Drachman D, Folstein M, et al: 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 1984; 34(7):939–944;doi:10.1212/wnl.34.7.939
50. McKeith IG, Dickson DW, Lowe J, et al: Diagnosis and management of dementia with Lewy bodies: third report of the DLB consortium. Neurology 2005; 65(12):1863–1872;doi:10.1212/01.wnl.0000187889.17253.b1
51. Neary D, Snowden JS, Gustafson L, et al: Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 1998; 51(6):1546–1554;doi:10.1212/WNL.51.6.1546
52. Mokkink LB, Terwee CB, Patrick DL, et al: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63(7):737–745; doi:10.1016/j.jclinepi.2010.02.006
53. Vispoel WP, Morris CA, Kilinc M: Practical applications of generalizability theory for designing, evaluating, and improving psychological assessments. J Pers Assess 2018; 100(1):53–67; doi:10.1080/00223891.2017.1296455
54. Ben-Shachar MS, Lüdecke D, Makowski D: effectsize: Estimation of effect size indices and standardized parameters. J Open Source Softw 2020; 5(56):2815;doi:10.21105/joss.02815
55. Zhu CW, Schneider LS, Elder GA, et al: Neuropsychiatric symptom profile in Alzheimer's disease and their relationship with functional decline. Am J Geriatr Psychiatry 2024; 32(12):1402–1416;doi:10.1016/j.jagp.2024.06.005
56. Nelson RS, Abner EL, Jicha GA, et al: Neurodegenerative pathologies associated with behavioral and psychological symptoms of dementia in a community-based autopsy cohort. Acta Neuropathol Commun 2023; 11(1):89;doi:10.1186/s40478-023-01576-z
57. Chandler J, Georgieva M, Desai U, et al: Impact of differential rates of disease progression in amyloid-positive early Alzheimer's disease: findings from a longitudinal cohort analysis. J Prev Alzheimers Dis 2024; 11(2):320–328; doi:10.14283/jpad.2024.28