

FACTOR STRUCTURE OF AN INTERNET-BASED SYMPTOM CHECKLIST AND CORRELATION WITH CONVENTIONAL RATING SCALES

C. Thomas Gualtieri

Dr Gualtieri
NC Neuropsychiatry
400 Franklin Square
1829 East Franklin Street
Chapel Hill NC 27514
919 933 2000 x 106
919 933 2830 fax
tg@ncneuropsych.com

ACKNOWLEDGEMENT, DISCLOSURES

The Neuropsych Questionnaire and related instruments were developed at The North Carolina Neuropsychiatry Clinics in Chapel Hill, Charlotte and Raleigh by Dr Gualtieri. The NPQ is not a commercial instrument and is freely available on the internet at www.ncneuropsych.com. The research was supported by North Carolina Neuropsychiatry, PA, in Chapel Hill and Charlotte. No external support was sought or received on behalf of this research.

ABSTRACT

The Neuropsych Questionnaire (NPQ) addresses two important clinical issues: how to screen patients for a wide range of neuropsychiatric disorders, quickly and efficiently; and how to acquire independent verification of a patient's complaints. The NPQ is available over the Internet in both adult and pediatric versions. The adult version of the NPQ consists of 207 simple questions about common symptoms of neuropsychiatric disorders. The NPQ reports scores patient and/or observer responses in terms of 20 symptom clusters: inattention, hyperactivity-impulsivity, learning problems, memory, anxiety, panic, agoraphobia, obsessions and compulsions, social anxiety, depression, mood instability, mania, aggression, psychosis, somatization, fatigue, sleep, suicide, pain and substance abuse.

In this report, the factor analysis with a larger set of patients to investigate the factor structure of the NPQ and report the correspondence between the NPQ and other clinical RSs in common use. This analysis generated three distinct factors: cognitive, somatic and manic – and anxiety-depression. The symptoms scales that comprise the anxiety-depression factor load equally with the mania and somatic factors but not with the cognitive factor.

In summary, the NPQ is a useful tool in neuropsychiatric practice. It also generates interesting data about the nature of symptom self-report and its relationship to specific psychiatric diagnoses.

INTRODUCTION

Capturing systematic and quantitative data about patients' symptoms can have positive impact on treatment decisions, patient outcome and quality assurance review.^{1,2} In clinical research, the systematic capture and quantitative analysis of psychiatric symptoms is often done by using rating scales (RS) and symptom questionnaires. However, conventional RS take time to administer and to score. If RS could be administered electronically clinicians might be inclined to use them more frequently.

Conventional rating scales are inexpensive, quick and simple to use. Ideally, they capture all of the symptoms or signs of a given condition; if so, they are systematic. The same or similar rating scales can be given to patients and to relevant others (parents, spouses, teachers, clinicians), permitting independent or concurrent assessment. Because rating scales address key problems, changes in scores are taken to reflect clinically meaningful change; improvement, for example, when the score goes down, deterioration if the score goes up, and remission if the score falls below a certain benchmark. Rating scales also allow for uniformity of assessment and comparability of results in different clinical sites.

Why then, aren't rating scales used more frequently by practitioners? Not because of their conceptual and psychometric weakness, although there is no shortage of those: the problem of rater bias, halo effects, leniency error, regression to the mean and the error of central tendency.³ In contrast to medical and psychological tests, rating scales are inherently subjective. They are essentially about giving a numerical value to a qualitative state, and they depend on the ability of the informant to understand the question and his or her inclination to honestly disclose information. Thus, many rating scales have a problem with poor correlation between rated and objectively observed behavior.⁴

Different RS have different metrics. A score of 20 indicates mild depression on the Beck depression inventory and severe depression on the Hamilton depression rating scale; 20 is a pathological score on the Conners Parent-Teacher Questionnaire, but is a normal score on the Brown ADD scales.

Of greater concern to practitioners, though, are the practical problems. For example, there a number of perfectly good rating scales to monitor change over time in a clinical trial of patients with depression. But what single rating scale should one use to measure the patient's depression in the first place? Comprehensive symptom inventories like the Personality Assessment Inventory (PAI) and the Behavioral Assessment for Children (BASC), and semi-structured interviews, like the Structured Clinical Interview for DSM Disorders (SCID) and the Schedule for Affective Disorders and Schizophrenia (SADS) are expensive, time-consuming and hard to score.

Even in areas where rating scale technology is well developed, like depression, anxiety and attention deficit hyperactivity disorder (ADHD), different rating scales emphasize different aspects of the condition. A good many of the questions on the Hamilton for depression (HAM-D) are about sleep and anxiety, for example, while the Beck depression inventory emphasizes cognitive symptoms of depression. Does that mean one should use both scales to diagnose depression? Because neuropsychiatric patients often have mixed or comorbid disorders, this rationale would lead one to use rating scales for the other condition(s) as well. But to deal with the content validity problem, one should use two or three rating scales for every comorbid disorder, which is clearly impossible.

We believe that rating scales are useful for guiding diagnosis and treatment in the clinical setting, but that the technology warranted an update. To that end, we we developed the Neuropsych Questionnaire (NPQ). The NPQ, is an internet-based symptom checklist, with 207 items to capture data that would ordinarily require the administration of several rating scales and symptom checklists. It generates quantitative data about the patient's status at a first evaluation visit, and, in a shorter form, at every follow-up visit. The NPQ captures the advantages

of rating scales but minimizes the inefficiencies in terms of patient load, scoring time and the cost of conventional instruments. In this paper, we report the correspondence between the NPQ and five widely used paper-and-pencil scales, a further analysis of the factor structure of the NPQ, and the relationship between NPQ scores in patients and spouses.

METHODS

THE NEUROPSYCH QUESTIONNAIRE (NPQ)

The adult version of the NPQ consists of 207 questions about common symptoms of neuropsychiatric disorders. Each item is scored as “not a problem” (0), a “mild problem” (1), a “moderate problem (2) or a “severe problem” (3). The default mode is a span of “the past two weeks” but the program can be set to other durations. The analysis we reported in (DATE) established that the 207 items clustered into 20 symptom scales; for example, 22 of the items addressed the problem of memory impairment. The patient’s scores on these 22 memory items are then averaged and multiplied by 100 to eliminate the decimal point. Thus, a memory score is generated; the highest score someone might achieve on a symptom scale is 300; that would indicate that he or she had marked every item in the relevant scale as “a severe problem.” The lowest score one might get is zero; every item in the scale is scored as “not a problem.” The 20 symptom scales in the NPQ were: attention (ATT), hyperactivity-impulsivity (HIP), learning problems (LPX), memory (MEM), anxiety (ANX), panic (PANIC), agoraphobia (AGORA), obsessions and compulsions (OC), social anxiety (SAD), depression (DEP), mood instability (MS), mania (MANIA), aggression (AGG), psychosis (PSYCH), somatization (SOMA), fatigue (FTG), sleep (SLEEP), suicide (SUI), pain (PAIN) and substance abuse (SA). The list of 207 items in the adult NPQ and the scales to which they contribute is given in Appendix 1.

In a previous publication, we described the construction of the NPQ, item selection and the generation of 20 symptom scales by factor analysis. We demonstrated that the NPQ is reliable (patients tested twice, patient-observer pairs, two observers), discriminates patients with different diagnoses and is sensitive to the effects of treatment.⁵

SUBJECTS

Beginning in November, 2006, patients at the Neuropsychiatry Clinics were routinely administered the NPQ as part of their initial evaluation and the data was stored on a secure database. Patients give written informed consent to the use of their de-identified clinical data, including the NPQ, for the purposes of research and program evaluation.

From November, 2006 through June 2009, 2883 patients were administered the NPQ. The records of patients were selected for this study if they were between 18 and 65 years of age. The patients had a wide range of neuropsychiatric diagnoses (Table 1). From this number of patients, 1353 records were selected for factor analysis. Records were selected only if they contained all of the relevant demographic data, if there were no indications of invalid or exaggerated responding⁶, and if the diagnostic category contained at least 10 patients.

In the records of these 1353 patients, 314 were identified who had also had a conventional RS administered during their initial evaluation. The RS were: the Hamilton RSs of Anxiety and Depression (clinician administered), The Beck Inventories of Anxiety And Depression, and the Brown ADHD RS. These data were applied to the concurrent validity analysis.

Among the 1353 records, there were 201 patients whose spouse had accompanied them to the initial evaluation and who had also rated the patient's symptoms on the NPQ. These data were applied to the "inter-rater" reliability analysis.

NPQ data was also available for 2764 patients, age 18-65, who had been evaluated at the clinics between July, 2009 and November, 2011. The range of neuropsychiatric diagnoses was much wider in this second sample, and all records were used, including normal subjects and patients with all diagnoses. As it happened, the demographic characteristics of the 2764 group was virtually identical to the 1399 group. (Table 1)

RESULTS

TABLE 1. DEMOGRAPHIC CHARACTERISTICS OF THE TWO SAMPLES

DXCODE	N		AGE	EDUC	%WHITE	%FEMALE	N	AGE	EDUC	%WHITE	%FEMALE
							18	43.4	15.8	0.93	0.67
								13.3	2.6		
GAD	135	mean	36.3	14.8	0.87	0.64	62	37.3	14.5	0.85	0.40
		sd	11.8	2.6				13.8	2.7		
MDD	325	mean	38.9	14.9	0.77	0.73	255	39.0	14.6	0.84	0.63
		sd	12.1	2.6				12.3	2.4		
ADHD	388	mean	30.4	15.0	0.79	0.52	256	34.8	14.7	0.90	0.58
		sd	11.4	2.5				12.0	2.6		
BPAD	103	mean	38.5	14.5	0.74	0.73	59	35.8	14.5	0.71	0.68
		sd	12.3	2.6				12.4	3.4		
SCHIZ	16	mean	37.1	14.2	0.44	0.63	10	45.1	12.9	0.50	0.75
		sd	12.4	2.2				13.3	2.5		
TBI	126	mean	43.1	13.8	0.66	0.32	77	34.9	13.6	0.82	0.53
		sd	11.5	2.9				12.5	2.6		
PAIN	58	mean	44.9	13.3	0.76	0.45	29	36.1	13.7	0.89	0.50
		sd	9.8	2.2				13.9	2.4		
PCS	33	mean	34.2	14.4	0.48	0.15					
		sd	12.1	1.7							
STROKE	10	mean	50.2	15.6	0.70	0.70					
		sd	8.2	2.9							
MOOD	59	mean	35.7	13.9	0.80	0.59	53	35.9	14.3	0.81	0.69
		sd	11.0	2.4				14.2	2.5		
COG DIS	22	mean	43.8	15.6	0.91	0.59	23	46.5	14.5	0.83	0.61
		sd	8.8	2.5				11.9	3.1		
OTHER PSYCH	54	mean	39.9	13.6	0.48	0.61	202	36.0	14.3	0.79	0.65
		sd	10.6	2.1				12.4	2.8		
OTHER NEURO	12	mean	33.5	13.9	0.25	0.75	84	41.6	13.8	0.78	0.48
		sd	12.3	1.9				13.0	2.8		
OTHER MEDICAL							9	36.9	15.3	0.67	0.57
								11.2	2.2		
OTHER COG	12	mean	59.1	15.7	1.00	0.67	16	38.4	13.2	0.77	0.57
		sd	5.9	2.9				15.6	2.8		
Total	1353	mean	36.9	14.6	0.75	0.58	2764	37.2	14.5	0.84	0.59
		sd	12.6	2.6				12.7	2.7		

FACTOR ANALYSIS

In this study, a factor analysis using extraction via Varimax rotation of the 20 scales was conducted, with the resulting factors and the original 20 symptom scales compared to appropriate ratings scales. When the 20 scales are factored, two scales (suicide and substance abuse) have extraction factors less than 0.5, so they are not included in subsequent factor analysis. The remaining scales present a clear solution for three factors (Varimax rotation):

Mania factor (MF) = mood stability, mania, aggression, psychosis

Somatic factor (SF) = somatic, fatigue, pain, sleep

Cognitive factor (CF) = attention, learning problems, memory

This left six scales – anxiety, panic, agoraphobia, obsessive-compulsive, social anxiety, depression – which loaded more or less equally with factors 1 and 2. When we factored those six scales by themselves, the result was a single-factor solution. These scales comprise a fourth factor, the Anxiety-depression factor (ADF), which is strongly related to the MF and the SF but not to the CF. Each of these four factors generate a factor score, by averaging the relevant symptom scale scores. The second patient group, sample 2764, generated the same factor structure (Table 2a).

TABLE 2A: NPQ FACTOR ANALYSIS (SAMPLE 1353)

	MF	SF	CF
ATT	.229	.186	.900
LPX	.211	.224	.854
MEM	.163	.312	.851
ANX	.622	.462	.350
PANIC	.515	.629	.125
AGORA	.448	.654	.214
OC	.597	.424	.426
SAD	.469	.457	.306
DEP	.544	.652	.230
HIP	.682	.117	.501
MS	.779	.316	.258
MANIA	.598	.279	.329
AGG	.807	.091	.125
PSYCH	.705	.438	.304
SOMA	.341	.790	.165
FTG	.138	.766	.305
SLEEP	.200	.587	.230
PAIN	.112	.827	.158

TABLE 2B. NPQ FACTOR ANALYSIS, SAMPLE 2764

	SF	MF	CF
ATT	.200	.243	.896
LPX	.200	.217	.856
MEM	.295	.157	.860
ANX	.568	.591	.279
PANIC	.681	.465	.063
AGOR	.661	.455	.195
OC	.412	.600	.367
SAD	.514	.437	.244
DEP	.652	.521	.236
HIP	.199	.701	.434
MS	.339	.751	.264
MANIA	.307	.590	.318
AGG	.140	.819	.114
PSYCH	.486	.674	.277
SOM	.779	.304	.153
FTG	.764	.094	.371
PAIN	.790	.080	.163
SLEEP	.659	.219	.209

CORRELATION OF NPQ WITH RATING SCALES

The correlation among the four factors, the symptom scales, and five commonly used rating scales (HAM-A, HAM-D, BAI, BDI, and BROWN ADHD) in the 314 patients from the 1353 sample are shown in Table 2. The first 5 rows and 5 columns report the correlations among the traditional rating scales themselves, with the rows and columns following showing the correlations between the NPQ factors/scales compared to the traditional rating scales. The highlighted cells are where one might expect to see strong correlations; e.g., between the DEP scale on the NPQ and the Hamilton Depression Rating Scale or the Beck Depression Inventory. All of the correlations in the table are significant at the level of $P < 0.01$

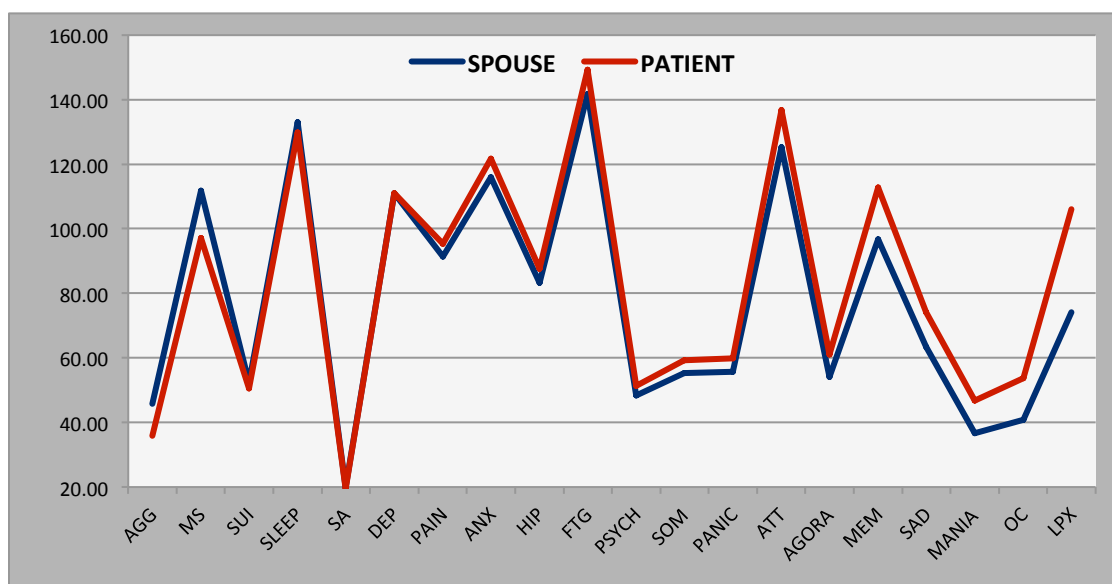
TABLE 4. CORRELATIONS BETWEEN TRADITIONAL RATING SCALES AND NPQ SCORES

	HAM-A	HAM-D	BAI	BDI	BROWN ADHD
HAM-A		.797	.597	.517	.374
HAM-D	.797		.544	.682	.391
BAI	.597	.544		.655	.458
BDI	.517	.682	.655		.530
BROWN ADHD	.374	.391	.458	.530	
SLS	.439	.477	.571	.612	.550
COGFAC		.179	.264	.320	.522
MANIAFAC	.281	.363	.382	.466	.462
SOMAFAC	.482	.469	.615	.573	.340
ANXDEPFAC	.496	.553	.602	.649	.443
ATT			.220	.311	.568
LP			.263	.286	.493
MEM	.171	.192	.256	.291	.372
ANX	.428	.461	.520	.494	.385
PANIC	.414	.396	.565	.456	.307
AGORA	.465	.444	.525	.509	.327
OC	.387	.424	.497	.553	.470
SAD	.305	.351	.381	.446	.338
DEP	.440	.617	.509	.753	.412
HIP	.204	.216	.332	.323	.433
MS	.299	.442	.360	.498	.398
MANIA	.199	.215	.283	.302	.345
AGG		.175	.172	.236	.259
PSYCH	.314	.386	.386	.508	.425
SOM	.457	.370	.610	.456	.351
FTG	.418	.475	.512	.589	.311
SLEEP	.403	.476	.392	.383	.196
SUICIDE	.324	.489	.338	.612	.297
PAIN	.395	.356	.529	.437	.246
SA					

PATIENT-SPOUSE CORRELATION

Two hundred and one of the patients in the 1353 sample had also been rated by their spouses on the same day. The intraclass correlation is 0.618 (Cronbach's alpha) and the Spearman-Brown coefficient is 0.764. The correlations between patient and spouse reports were all significant at the $P < 0.001$ level. The symptom scale scores on the NPQ by patient-spouse reports is shown in Figure 1 below.

FIGURE 1. NEUROPSYCHIATRIC QUESTIONNAIRE SYMPTOMS SCALE SCORES: PATIENT-SPOUSE REPORTERS (N=201)



While the figure appears to show good correlation between the patient and spouse, a closer look at the actual data shows that the patients and spouses perception of symptoms, is more variable however, with correlation coefficients ranging from 0.802 (PAIN) to 0.293 (MANIA) (Table 3). The patient and spouse report statistically significantly different total scores on aggression, mood stability, SAD, mania, obsessive-compulsive, and LPX.

TABLE 3. PATIENT AND SPOUSE SCORES ON THE NPQ

	SPOUSE	PATIENT	Correlation	DIFF	%	t	Sig. (2-tailed)
AGG	45.7	35.9	.526	-9.8	-0.27	2.510	.013
MS	111.9	97.1	.535	-14.7	-0.15	2.813	.005
SUI	52.4	50.5	.628	-1.8	-0.04	-.219	.827
SLEEP	132.9	129.8	.678	-3.1	-0.02	.543	.588
SA	19.8	19.4	.317	-0.4	-0.02	.107	.915
DEP	111.0	111.0	.680	0.0	0.00	.053	.958
PAIN	91.4	95.3	.802	3.9	0.04	-.191	.849
ANX	115.9	121.6	.590	5.7	0.05	-.968	.334
HIP	83.2	87.5	.442	4.3	0.05	-.810	.419
FTG	141.8	149.3	.706	7.5	0.05	-1.027	.306
PSYCH	48.4	51.4	.532	3.0	0.06	-.851	.396
SOM	55.2	59.2	.707	4.0	0.07	-.521	.603
PANIC	55.7	59.8	.612	4.1	0.07	-.773	.440
ATT	125.2	136.8	.529	11.6	0.08	-1.765	.079
AGORA	54.2	61.1	.599	6.9	0.11	-1.559	.121

MEM	96.8	112.7	.443	15.9	0.14	-1.178	.240
SAD	63.4	74.0	.560	10.7	0.14	-2.167	.032
MANIA	36.6	46.7	.293	10.2	0.22	-2.290	.023
OC	40.9	53.6	.497	12.7	0.24	-3.370	.001
LPX	74.2	106.1	.536	31.9	0.30	-6.191	.000
AVERAGE	77.8	82.9	0.561	5.1	0.056		

There is something systematic in the way patients and spouses differ in their perceptions of the patient's symptoms. As a rule, patients are more aware, or at least more expressive of learning problems, obsessive-compulsive symptoms and symptoms of mania, social anxiety, poor memory and agoraphobia while spouses are more cognizant of, or more likely to complain of symptoms of aggression and mood instability than are patients

DISCUSSION

THE DEVELOPMENT OF THE NPQ

The purpose of the NPQ is to inquire after the presence and intensity of symptoms of a wide range of neuropsychiatric disorders, to score patient responses and assemble them in the appropriate symptom scales, and to generate a report on the spot. The NPQ is not a diagnostic test, any more than any symptom checklist is a diagnostic test. It simply presents the opinions of a patient (and others) about the patient's clinical state at a point in time. Therefore, the validity of the questionnaire scores is a matter for the clinician to evaluate. If the NPQ report indicates that the patient scores in the severe range on all of the depression items, for example, that clearly merits the physician's attention. If the scores are low for depression, but the patient is clearly depressed during the examination, that also merits attention. It calls attention to the patient's appreciation of his or her problem, or at least his willingness to acknowledge it.

The NPQ is suitable for routine clinical use; the database from which the data in this report were selected comprised more than 5,000 test administrations during initial patient evaluations. It aims to complement the clinical assessment of the neuropsychiatric patient. PC and internet technology generate quantitative data that enriches the psychiatric interview. It alleviates the problem of having to inquire, arduously, after numerous potential problems, and allows the clinician to spend more time exploring the meaning of the patient's complaints.

The psychometric properties of the instrument are encouraging in some areas but troubling in others. The correlations with conventional rating scales described in table 4 indicate a degree of non-specificity. The factor structure of the NPQ indicates three discrete factors and one factor that overlaps with two of the others. In an earlier publication, we described the factor structure of the NPQ, which indicated the same pattern. In that paper we also reported its test-retest reliability and sensitivity to treatment.⁵ In this report, we describe levels of agreement between patients and their spouses when they both participate in the test during an initial evaluation. Rating scales used in clinical research usually report inter-rater reliability, involving clinicians who are trained to assess symptoms in a uniform way. Patient-spouse correlations are not comparable, although they are a form of "inter-rater" reliability. The intraclass correlation (Spearman-Brown coefficient) is 0.74, which is respectable but not compelling. The Pearson product-moment correlations range from 0.293 (mania) to 0.802 (pain) are not comparable to the IRR reported for research scales, because we are not dealing with trained raters. However the data presented in Table 4 are interesting in their own right. As a rule, spouses rate patients a bit lower than the patients rate themselves, by about 6%. Patients are more aware, or at least more expressive of learning problems, obsessive-compulsive symptoms and symptoms of mania, social anxiety, poor memory and agoraphobia. Spouses are more cognizant of, or more likely to complain of symptoms of aggression and mood instability than patients

are. When spouses and patients agree in their assessment on the NPQ, that is re-assuring; when they differ, that is clinically useful information as well, perhaps even more useful.

The NPQ generates results similar to that achieved with conventional, paper-and-pencil RS and symptom checklists. In an earlier publication, we reported correlations between NPQ scales and the relevant scales on the Personality Assessment Inventory in 144 adult patients who had taken both tests, with significant correlations for the depression symptom scale ($r = 0.67$), anxiety ($r = 0.55$); mania ($r = 0.43$), psychotic ($r = 0.62$) and aggression ($r = 0.65$).⁵ In this paper, we compared NPQ scores with five commonly used clinical rating scales. The NPQ is more similar to other patient self-rating scales like the Becks and the Brown than to clinician-rated scales like the Hamilton's. The correlation between the DEP scale on the NPQ, for example, with the Beck Depression Inventory (patient report) is 0.75 but only 0.62 with Hamilton Depression Rating which is completed by clinicians. The ANX scale correlates with the Beck Anxiety Inventory at $r = 0.52$ but with the Hamilton Anxiety Rating Scale at 0.43. The correlations are only moderate, but more respectable when one allows for the reliability of the respective instruments. So, if the reliability of the NPQ is 0.74 and the reliability of the Hamilton Depression Rating Scale is 0.85, the maximum correlation between the two instruments is the square root of the product of their reliabilities, or $(0.74 * 0.85)^{1/2}$, or 0.793. By this formula, the adjusted correlation between the NPQ and the HamD is 0.77. In any event, it appears that moderate-to-poor correlations are the rule when clinician rating scales are compared to patient self-report rating scales. For example, in a review of seven studies comparing the HamD to the BDI, the correlations ranged from 0.21 to 0.82⁷ and in nine similar studies, correlations between the HamD and the Zung Self-rating scale Depression Scale ranged from 0.22 to 0.95.^{8,9} The reliability of the NPQ is more-or-less equivalent to published data for other scales.

The structure of the NPQ is comprised of twenty symptom scales, determined by factor analysis of the 207 items. Eighteen of the twenty symptom scales load with three principal components, a cognitive factor (attention, memory and learning problems), a somatic factor (somatic complaints, pain, fatigue, sleep) and a mania factor (hyperactive-impulsive, mood stability, mania, aggression, psychosis). A fourth factor, anxiety-depression, is comprised of six symptom scales (anxiety, panic, agoraphobia, obsessions & compulsions, social anxiety and depression). These symptom scales load more or less equally with the MF and the SF, but not with the CF. The factor structure of the NPQ was found to be consistent in two large databases of patients with diverse neuropsychiatric conditions. The factor structure suggests that the occurrence of anxiety/depression symptoms occur in two contexts: one in which internalizing elements are manifest as somatic complaints, pain, fatigue and poor sleep, and another in which externalizing elements like hyperactivity, aggression, mania, etc are prominent.

One can draw only so many conclusions from a report of this nature. Although the sample sizes were large, the diversity of clinical diagnoses is both an advantage – it supports the relevance of a broad-spectrum symptom checklist – but also a disadvantage. Patients with anxiety, mood disorders and ADHD were an inordinately high proportion of patients, patients with other psychiatric conditions less well represented. By the same token, traumatic brain injury is over-represented among the neurological disorders. This is what comes of “data-mining” a convenience sample of clinic patients.

The factor structure of the NPQ is a statistical reality but is it meaningful? The four factors are all highly correlated with one another, especially the MF, SF and ADF. The high correlation among factors is probably an example of the halo effect the tendency of a patient who feels bad to describe difficulties in a wide range of different areas.¹⁰ The specificity and sensitivity of the factors and the symptom scales relative to specific diagnoses and to differential diagnosis is a topic that will be addressed in future publications.

The NPQ itself is open to criticism. RS technology generally recommends using “anchor points” to describe to the rater where a response should lie, in the most specific terms possible. Anchor points are believed to improve the reliability of a RS.¹¹ Anchor points would be unwieldy in this format; they would increase reading time. Reliability of the NPQ is achieved through item redundancy.

The optimal number of points on a RS has been the subject of earnest debate; the idea that seven-point scales are optimal dates from 1924¹² but others have suggested better results with finer-grained scales (10-20 points).¹³ Preston et al demonstrated that scales with 7, 8 or 9 points were more reliable than scales with 2, 3 or 4 points.¹⁴ But it has also been represented that “the number of response categories does not materially affect the cognitive structure derived from the results.”¹⁵ Simplicity and ease of administration, led us to select a four-point scale. The idea was that reliability would be preserved by reducing administration time for each item, and thus permitting the inclusion of multiple and even redundant items.

The items in the NPQ are not weighted. That is, every item in a scale is considered as important as every other. This created some anomalies. For example, in the beta version of the NPQ, the item “mind going blank” was deemed as important to the measurement of anxiety as “feeling nervous” or “feeling anxious.” The problem was addressed in the item analysis; items that were infrequently cited by patients, or that failed to generate sufficient weight in the factor analysis, or that correlated poorly with the scale score, were dropped from that particular scale. In the case of anxiety, “mind going blank” was eliminated but “feeling nervous” and “feeling anxious” were retained.

The scales refer to symptom clusters and not to DSM or ICDM diagnoses. That was a considered decision. The NPQ is a measurement instrument, not a diagnostic instrument. Patients’ complaints and the opinions of observers are valuable data, but diagnosis is a medical exercise. By the same token, the symptom clusters are more or less “pure.” For example, the depression scale asks only after questions about depression itself, not anxiety, fatigue or sleep difficulties, which are addressed in the relevant scales. In this wise, the NPQ is different from conventional RS, like the Hamilton Depression RS, which contains a number of anxiety-related items, or the Conners Parent-Teacher Questionnaire, which includes items related to inattention as well as hyperactivity-impulsivity.

Computerized rating instruments were originally developed for use in clinical trials.¹⁶ Computer-administered versions of clinician-administered scales are now available for the assessment of depression, anxiety, obsessive-compulsive disorder, social phobia and Alzheimer’s disease.^{17 18 19 20} They have been used in patients with severe mental illness²¹, as a screening tool for mental disorders in adolescents²² and to measure consumer satisfaction with treatment in a community mental health setting.²³ Other studies have shown “good overall interrater reliability scores”²⁴ and agreement with conventional RS.²⁵ A computerized adaptive test¹⁸ for depression showed high sensitivity and specificity when depressed patients were compared to normal controls.¹⁸ Patient reaction to computerized instruments is said to be positive, “with patients generally more honest with and often preferring the computer for assessing sensitive areas such as suicide, alcohol or drug abuse, sexual behavior, or HIV related symptoms.”¹⁶

The NPQ has been developed largely on the basis of existing RS, symptom questionnaires, and diagnostic criteria. Nevertheless, it a new test, and as such, it has to pass the essential tests of construct and content validity, test-retest and inter-rater reliability. Because it is a broad spectrum test, relevant to the problems of patients with a wide range of disparate disorders, it is necessary to evaluate the structure of the test in empirical terms. To this end, the factor analysis described here has demonstrated a structure that is statistically sound and clinically intuitive.

REFERENCES

1. Slade M, McCrone P, Kuipers E, et al. Use of standardised outcome measures in adult mental health services: randomised controlled trial. *Br J Psychiatry*. 2006;189:330–336.
2. Kesselheim AS, Ferris TG, Studdert DM. Will physician-level measures of clinical performance be used in medical malpractice litigation? *JAMA*. 2006;295(15):1831–1834.
3. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry*. 2004;161(12):2163–2177.
4. Streiner D, Norman. *Health Measurement Scales A Practical Guide to Their Development and Use*. 2nd ed. New York: Oxford University Press; 1995.
5. Gualtieri C. An Internet-based symptom questionnaire that is reliable, valid, and available to psychiatrists, neurologists, and psychologists. *MedGenMed*. 2007;9(4):3.
6. Slick DJ, Sherman EM, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol*. 1999;13(4):545–561.
7. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Qual Life Res*. 1993;2(6):367–368.
8. Hedlund J, Vieweg B. The Hamilton Rating Scale for Depression: a comprehensive review. *J Operational Psychiatry*. 1979;10:149–165.
9. McDowell I, Newell C. *Measuring Health. A Guide to Rating Scales and Questionnaires*. 2nd ed. New York: Oxford Univ Press; 1996.
10. Nisbett, R.E., Wilson, T.D. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*. 1977;35(4):250–256.
11. Dowd JB, Todd M. Does self-reported health bias the measurement of health inequalities in U.S. adults? Evidence using anchoring vignettes from the Health and Retirement Study. *J Gerontol B Psychol Sci Soc Sci*. 2011;66(4):478–489.
12. Symonds P. On the Loss of Reliability in Ratings due to the Coarseness of the Scale. *Journal of Experimental Psychology*. 1924;17:456–461.
13. Garner WR. Rating scales, discriminability, and information transmission. *Psychol Rev*. 1960;67:343–352.
14. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst)*. 2000;104(1):1–15.
15. Schutz H, Rucker M. Variable Configurations across Scale Lengths: An Empirical Study. *Educ Psychol Measurement*. 1975;35:319–24.
16. Kobak KA, Greist JH, Jefferson JW, Katzelnick DJ. Computer-administered clinical rating scales. A review. *Psychopharmacology (Berl)*. 1996;127(4):291–301.
17. Smits N, Zitman FG, Cuijpers P, Den Hollander-Gijsman ME, Carlier IVE. A proof of principle for using adaptive testing in Routine Outcome Monitoring: the efficiency of the Mood and Anxiety Symptoms Questionnaire -Anhedonic Depression CAT. *BMC Med Res Methodol*. 2012;12:4.

18. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch. Gen. Psychiatry*. 2012;69(11):1104–1112.
19. Cummings J, Gould H, Zhong K. Advances in designs for Alzheimer's disease clinical trials. *Am J Neurodegener Dis*. 2012;1(3):205–216.
20. Inoue M, Jimbo D, Taniguchi M, Urakami K. Touch Panel-type Dementia Assessment Scale: a new computer-based rating scale for Alzheimer's disease. *Psychogeriatrics*. 2011;11(1):28–33.
21. Wolford G, Rosenberg SD, Rosenberg HJ, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. *Psychiatr Serv*. 2008;59(7):769–775.
22. Fein JA, Pailler ME, Barg FK, et al. Feasibility and effects of a Web-based adolescent psychiatric assessment administered by clinical staff in the pediatric emergency department. *Arch Pediatr Adolesc Med*. 2010;164(12):1112–1117.
23. Goldstein LA, Connolly Gibbons MB, Thompson SM, et al. Outcome assessment via handheld computer in community mental health: consumer satisfaction and reliability. *J Behav Health Serv Res*. 2011;38(3):414–423.
24. Rucker J, Newman S, Gray J, et al. OPCRIT+: an electronic system for psychiatric diagnosis and data collection in clinical and research settings. *Br J Psychiatry*. 2011;199(2):151–155.
25. Chinman M, Young AS, Schell T, Hassell J, Mintz J. Computer-assisted self-assessment in persons with severe mental illness. *J Clin Psychiatry*. 2004;65(10):1343–1351.