

SF-36 includes less Parkinson Disease (PD)-targeted content but is more responsive to change than two PD-targeted health-related quality of life measures

Carlos A. Brown · Eric M. Cheng · Ron D. Hays ·
Stefanie D. Vassar · Barbara G. Vickrey

Accepted: 11 August 2009 / Published online: 28 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract

Objective To compare validity including responsiveness, and internal consistency reliability and scaling assumptions of a generic (SF-36) and Parkinson Disease (PD)-targeted (PDQ-39; PDQUALIF) health-related quality of life (HRQOL) measures.

Methods Ninety-six PD patients were administered for all HRQOL measures by telephonic interview at baseline and 18 months. Relative efficiency and responsiveness were compared relative to four external criteria (self-ratings of PD's daily effects, global Quality of Life, PD symptom severity, and a depression screener). We examined whether PD-targeted measures explained unique variance beyond the SF-36 by regressing criterion variables on HRQOL scales/items. Adequacy of PD-targeted measures' original scaling was explored by item-scale correlations.

Results Relative efficiency estimates were similar for generic and PD-targeted measures across all criteria. Responsiveness analyses showed that the SF-36 yielded large (>0.8) effect sizes (ES) for three of eight scales for each of two criterion variables, compared to only one large ES for any scale in either PD-targeted measure. Adjusted R^2 increased from 14 to 27% in regression models that included PD-targeted items compared to models with only SF-36 scales.

Item-scale correlations showed significant cross-loading of items across scales of the PD-targeted measures.

Conclusions SF-36 responsiveness was better than that of two PD-targeted measures, yet those measures had content that significantly explains PD patients' HRQOL.

Keywords Health-related quality of life · Parkinson Disease · PDQUALIF · PDQ-39 · Responsiveness · SF-36

Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disease after Alzheimer's disease. PD afflicts about one million Americans, or about 1% of the population over 60 years of age [1]. As a chronic and progressive disease, PD may impact a person's physical, mental and social health. PD patients may experience impairments in mood (especially depression and anxiety), orthostatic hypotension and other autonomic symptoms, sleep disturbances, and impulse control disorders, indicating the likelihood of a broad impact of this disease on health [2, 3].

Health-related quality of life (HRQOL) conceptualizes how well an individual functions and feels about his/her life. It encompasses physical, mental, and social dimensions of health [4]. There are two main types of HRQOL measures: generic and disease-targeted instruments, which differ in their form, content, and intended purpose. Generic HRQOL measures enable comparisons across populations, regardless of whether they have a particular condition [5]. The 36-Item Short-Form Health Survey (SF-36) is the most widely used HRQOL survey instrument in the United States [5, 6]. Its reliability and construct validity have been supported in studies of a number of other patient populations.

The views expressed in this article are those of the author(s) and do not necessarily represent the views of the Department of Veterans Affairs.

C. A. Brown · E. M. Cheng (✉) · S. D. Vassar · B. G. Vickrey
UCLA/VA Greater Los Angeles Healthcare System,
Los Angeles, CA, USA
e-mail: eric.cheng@va.gov

R. D. Hays
UCLA, Los Angeles, CA, USA

The SF-36 includes eight health concepts judged as the most affected by disease and treatment, selected from 40 concepts assessed in the Medical Outcome Study [6].

Disease-targeted measures for several neurological conditions, such as multiple sclerosis and epilepsy, may provide additional key content over generic measures, tapping domains of HRQOL important to persons with these conditions [7, 8]. The most widely used PD-targeted HRQOL measure is the Parkinson's Disease Questionnaire (PDQ-39), developed first as a 65-item questionnaire piloted on 359 individuals with PD attending a neurology outpatient clinic [9]. After testing for basic acceptability and comprehension, the number of questionnaire items was reduced to 39 items by a factor analysis, distributed across eight scales. The PDQ-39 has proved to have satisfactory reliability and construct validity in relation to other measures but limited evidence of responsiveness [10].

Another PD-targeted HRQOL measure, the Parkinson's Disease Quality of Life (PDQUALIF) scale, was initially developed and evaluated in a cross-sectional study of 233 outpatient clinic attendees with physician-confirmed idiopathic PD [11]. Movement disorder specialists ranked a list of 73 indicators relevant for quality of life (QOL) in PD, and the top 32 ranked indicators were included in the measure. More than any other PD-targeted measure, the PDQUALIF taps many non-motor symptoms of PD including fatigue, sleep, autonomic dysfunction, and sexual function.

Patient-reported outcome measures are increasingly recognized as important for longitudinal studies including clinical trials of new treatments (http://www.fda.gov/fdac/features/2006/606_patients.html), and a review of a range of disease-targeted measures found overall better ability to detect change in clinically relevant domains relative to generic measures [12]. Yet, disease-targeted measures require investment of resources to develop and evaluate relative to "off-the-shelf" existing generic measures in widespread use, such as the SF-36. Thus, it is critical to compare generic and disease-targeted measures on their responsiveness to change in HRQOL over time. To date, responsiveness indices (effect sizes) have been reported only for the eight scales of the PDQ-39 scales, and in that study, only a few scales detected any effect [10].

Our goals were to compare these two PD-targeted HRQOL measures with the widely used SF-36 on responsiveness, construct validity, internal consistency reliability, and scaling assumptions. Because the PDQ-39 is the most widely used PD-targeted HRQOL measure, and because the PDQUALIF was specifically intended to tap not only motor but also non-motor aspects of QOL in PD, we selected for inclusion these two PD-targeted measures out of the small group of existing PD-targeted measures at the time the study began [13]. We hypothesized that reliability would be comparable but that the PD-targeted measures

would have better construct validity and responsiveness than the generic SF-36.

Methods

Sample

A convenience sample of patients who were 18 years old or older and English-speaking were recruited from the Greater Los Angeles VA Healthcare System Movement Disorders Clinic and from the University of California Los Angeles (UCLA) Movement Disorders Clinic. At UCLA, study flyers were handed to PD patients by their movement disorder physician at the time of the patient's visit. At the time of check-out from a regular appointment in the VA Movement Disorders Clinic, patients with PD were informed of the study and offered the flyer. Recruiting clinicians and staff were asked to provide information about the study only to patients without diagnosed dementia. In both sites, the flyer contained information on how to contact the study team through a toll-free telephone number. If the patient expressed interest during the time of check-out, the clinic clerk requested approval from the patient for the research team to initiate contact with the patient. Ninety-six patients provided verbal informed consent and were enrolled and completed the baseline telephone interview. The study was approved by the Institutional Review Boards at the VA Greater Los Angeles Healthcare System (project number PD1-01-158-1), and at UCLA (approval number G050405204).

Study design

The baseline telephone interview took place between March 2005 and February 2006, and the follow up interview between December 2006 and March 2007. The interval from baseline-to-follow up telephone interview had a mean of 17.9 months (range equals 11.1–24 months), a median of 17.9 months, and a standard deviation of 4.2 months. Measures were administered in the same way at both baseline and follow-up to avoid differential effects due to mode of administration [14].

Measures

Generic-HRQOL measure

The SF-36 (version 1.0) has 36 items covering eight scales: Physical Functioning, Role Limitations due to Physical Health, Role Limitations due to Emotional Problems, Pain, Emotional Well-Being, Energy, General Health, and Social Functioning. A Physical Health Composite score (PCS)

and a Mental Health Composite score (MCS) can be derived from the SF-36 scales. The SF-36 is most commonly self-administered by mail survey or administered by telephone interview [15, 16].

PD-targeted HRQOL measures

The PDQ-39 has 39 items covering eight scales: Mobility, Activities of Daily Life, Emotional Well-Being, Stigma, Social Support, Cognitions, Communication, and Bodily Discomfort [9]. An overall score is constructed as the average of the eight scale scores. The PDQ-39 has been administered by telephone, with comparable levels of missingness, reliability, and construct validity to self-administration [3]. The PDQUALIF has 33 items covering seven scales; Social and Role Function; Self Image and Sexuality; Sleep; Outlook; Physical Functioning, Independence; and Urinary Function [11]. An overall score is the average of the seven scale scores.

In this study, all the PD-targeted HRQOL scales were scored on a 0–100 possible range with 0 representing the worst possible score and 100 the best possible score.

Criterion variables for evaluating validity of HRQOL measures

We used four criterion variables to assess validity of the HRQOL measures.

Criterion variable #1: “How PD affects you on a day-to-day basis?” This single item global rating of difficulty with day-to-day activities was developed specifically for PD based on interviews with PD specialist clinicians, patients, caregivers, and on a literature review; it has support for construct validity in terms of anticipated associations with depression, cognition, and PD severity in a community-based PD sample [17]. Subjects are asked to indicate one choice that “best describes how your Parkinson’s disease has affected your day-to-day activities in the last month:” (1) no difficulties, (2) mild difficulties, (3) moderate difficulties, (4) high levels of difficulties, or (5) extreme difficulties. Each choice is followed by a detailed example.

Criterion variable #2: “Current rating of overall QOL on scale of 1 to 10.” Subjects chose an integer between 1 (worst possible QOL, as bad as or worse than being dead) and 10 (the best possible QOL). This variable was adapted from other measures [18].

Criterion variable #3: “Rating of PD symptoms in the past 6 months”. In order to assess symptom severity, subjects were asked to rate the severity of their symptoms as (1) no symptoms, (2) mild symptoms, (3) moderate symptoms, and (4) severe symptoms.

Criterion variable #4: “Patient Health Questionnaire (PHQ)-9 Scoring Categories”. This nine-item self-rated

depression screener is directly mapped on the Diagnostic and Statistical Manual-IV (DSM-IV) criteria for major depression [19]. It has been evaluated in large studies of primary care patients and used in a recent large trial of depression care in the elderly [20]. Three categories can be derived based on responses to the nine items: (1) depression treatment may be not needed, (2) clinical judgments about treatment on duration of symptoms and functional impairments, and (3) warrants treatment for depression.

We hypothesized that the PD-targeted HRQOL measures would be more highly associated than the SF-36 with the two criterion variables that elicited ratings of day-to-day difficulties with PD (criterion variable #1) and PD symptoms (criterion variable #3). Because the two PD-targeted HRQOL measures each had one summary score and the SF-36 had separate physical and mental health composite scores, we hypothesized that the SF-36 mental health composite score would be more highly associated with the PHQ-9 (criterion variable #4) than all the three other summary scores. We had no a priori hypotheses with respect to the global QOL rating (criterion variable #2) and summary scores of PD-targeted versus generic measures, nor did we make any formal a priori hypotheses about individual scale scores on any measure and the four criterion variables.

Socio-demographic and clinical characteristics included gender, age, race/ethnicity, marital status, education, and employment. We also collected self-reported Activities of Daily Living (ADL) via the Unified Parkinson’s Disease Rating Scale (UPDRS) [21].

Data collection

Telephone interviews were administered by trained research assistants who followed protocols for quality of data collection by interview. Participants were paid \$10 for each interview. The research assistant obtained verbal consent over the telephone. Data were directly entered into an electronic spreadsheet. Reasons for the 38 non-respondents (39.6% non-response) at the follow-up interview 1–2 years later include: unreachable despite multiple attempts by phone ($n = 19$), phone number disconnected ($n = 6$), asked to not be contacted again after the first survey ($n = 4$), declined ($n = 4$), unable to participate because of stroke/dementia ($n = 3$), deceased ($n = 1$), and other ($n = 1$).

Data analysis

Data were analyzed using SAS version 9.1 (SAS® software, Version 9.1, SAS Institute, Cary, NC).

Mean scores, standard deviations, ranges, and percentages of patients scoring the minimum = 0 (floor), and maximum = 100 (ceiling) possible scores were examined.

Internal consistency reliability of each multi-item scale was assessed using Cronbach's alpha [22]. Reliability of composite scores was estimated using Mosier's formula [23]. We categorized scales as reliable if Cronbach's alpha was greater than or equal to 0.70, a widely used threshold for adequate reliability in group comparisons [24, 25].

Relative validity is reported as the ratio of the *F*-statistic of each scale of the three HRQOL measures to the *F*-statistic of a designated reference scale, usually the smallest *F*-statistic among the scales of the three HRQOL measures [26]. For a given criterion variable, the scale with the highest *F*-ratio is thus most sensitive to differences across categories of that criterion variable; for a fixed level of power, relative validity (*F*-ratio) values "are equivalent to the ratio of sample sizes that would be required to detect the known group difference using one measure versus the other" [27]. For each of the four criterion variables (baseline distributions in Table 1 and Appendix Table 7), we used analysis of variance (ANOVA) based *F*-statistics to compare mean HRQOL scale scores across different patient groups, based on patient's categorization across different levels within that criterion variable [27].

Responsiveness was assessed using standard methods [28]. We examined the decline in responses between the baseline and the follow-up interviews for criterion variable #1 ("How PD affects you on a day-to-day basis") and for criterion variable #2 ("Current rating of overall QOL"). We excluded responses from subjects who "improved" rather than combining them with "declined" because prior studies suggest that the magnitude of responsiveness is different for these two groups and higher among "declined" [29], and PD is a disease of progressive decline [29]. We examined the distribution of the change in responses and used the existing literature and clinical judgment to set the threshold for a change in each criterion variable. For criterion variable #1 of how PD affects you daily, each of the five response choices were developed as clinically distinct and meaningful [17]; thus, we set a threshold of true change to be a change of at least one level. For criterion variable #2 of overall QOL, there are 10 response choices with anchors only at each extreme; we set a threshold of true change to be at least two levels based on our judgment. Selection of these thresholds was made a priori. Unchanged was defined as responses on the second interview that did not meet threshold for a true change from the first interview. The three most widely used responsiveness indices were calculated: effect size (ES), standardized response mean (SRM), and the Guyatt responsiveness statistic (GRS) [30]. For these indices, the numerator is the mean change in scale score for the declined group. The denominators are the standard deviation of the baseline scale score of the declined group (ES), the standard deviation of change in scale score for the

change group (SRM), and the standard deviation of change in scale score for the unchanged group (GRS) [27]. Because each of these indices look at change for the declined group, we supplemented them by computing the *F*-statistic for the difference in change scores between the declined and unchanged groups. We categorized ES as large (greater than or equal to 0.80), medium (between 0.50 and 0.79), small (between 0.20 and 0.49), and not detectable (less than 0.2) according to well-known published benchmarks [31] and focused on ES in our interpretation, because such established benchmarks exist. (There is one published report providing regression equations linking different responsiveness indices [32].)

We used multivariate models to determine whether PD-targeted measures captured important HRQOL content beyond the SF-36. Each of the four criterion variable served as the dependent variable in a multivariate model, and the eight SF-36 scales served as the independent variables (Model 1). We then forced in the SF-36 scales that were significant at $P < 0.10$ in Model 1 and allowed items from the two PD-targeted HRQOL measures to enter at $P < 0.05$ (Model 2), using stepwise regression. We compared the improvement in adjusted R^2 from Model 1 to Model 2 for each of the criterion variables.

In order to evaluate the original scoring of the two-PD targeted measures, we estimated using baseline data from all the 96 subjects, the item-scale correlations from multitrait scaling analyses [33]; computing product-moment correlations between items and scales, correcting for the overlap of the item with the scale where applicable. We inspected the correlation matrix for potential lack of item discrimination across scales by highlighting those correlations that were $\frac{1}{2}$ standard error below or any amount above the correlation of the item with the scale in which it was placed.

Results

Sample characteristics

The mean age of the 96 enrolled PD patients was 72 years, 88% were white. More than three-quarters (84%) were male (see Table 1). Sixty-five percent were currently married; 63% held a Bachelor's degree or higher; 66% were retired and not working, and 58% reported moderate or severe PD symptoms.

For criterion variable #1 (How PD affects you on a day-to-day basis), 54% reported moderate, high, or extreme difficulties. For criterion variable #2 (rating of overall QOL), 61% rated their QOL as 7 or higher on the 1–10 QOL scale.

We compared characteristics of the 58 participants who completed the follow up interview to the 38 who did not

Table 1 Sample characteristics

	Total <i>N</i> (%) Mean (SD) <i>N</i> = 96	Patients with follow-up data <i>N</i> (%) Mean (SD) <i>N</i> = 58	Patients without follow-up data ^a <i>N</i> (%) Mean (SD) <i>N</i> = 38	<i>P</i> -value
Male	81 (84.4)	46 (79.3)	35 (92.1)	0.09
Age	71.6 (10.9)	70.9 (10.2)	72.6 (11.9)	0.44
Race				
White	84 (87.5)	53 (91.4)	31 (81.6)	0.03
Black	3 (3.1)	3 (5.2)	0 (0.0)	
Hispanic	7 (7.3)	1 (1.7)	6 (15.8)	
Asian	1 (1.0)	0 (0.0)	1 (2.6)	
Other	1 (1.0)	1 (1.7)	0 (0.0)	
Marital status				
Married	62 (64.6)	35 (60.3)	27 (71.1)	0.34
Separated	2 (2.1)	1 (1.7)	1 (2.6)	
Divorced	18 (18.8)	14 (21.1)	4 (10.5)	
Widowed	9 (9.4)	4 (6.9)	5 (13.2)	
Never married	5 (5.2)	4 (6.9)	1 (2.6)	
Highest degree				
None/less than high school	3 (3.1)	2 (3.5)	1 (2.6)	0.78
High school or GED	22 (22.9)	11 (19.0)	11 (29.0)	
Associate's degree	11 (11.5)	8 (13.8)	3 (8.0)	
Bachelor's degree	38 (39.6)	23 (39.7)	15 (39.5)	
Graduate/professional degree	22 (22.9)	14 (24.1)	8 (21.1)	
Years of schooling	15.7 (2.4)	15.7 (2.5)	15.5 (2.3)	0.68
Employment				
Working full time/part time	18 (18.8)	6 (10.3)	12 (31.6)	0.06
With a job and not working for other reasons	1 (1.0)	1 (1.7)	0 (0.0)	
Unemployed and looking for work	1 (1.0)	0 (0.0)	1 (2.6)	
Disabled and not working	13 (13.5)	9 (15.5)	4 (10.5)	
Retired and not working	63 (65.6)	42 (72.4)	21 (55.3)	
UPDRS ADLs scale (Range: 0–48) 0 = best state	14.5 (7.6)	14.2 (7.8)	15.1 (7.3)	0.54
Recruited at VA	54 (56.2)	32 (55.2)	22 (57.9)	0.79
How Parkinson's disease affects you on a day-to-day basis?				
No difficulties	10 (10.4)	8 (13.8)	2 (5.3)	0.17
Mild difficulties	34 (35.4)	23 (39.7)	11 (28.9)	
Moderate difficulties	38 (39.6)	17 (29.3)	21 (55.3)	
High levels of difficulties	12 (12.5)	10 (17.2)	2 (5.3)	
Extreme difficulties	2 (2.1)	0 (0.0)	2 (5.3)	
On a scale of 1–10, where 10 is the best possible quality of life and 1 is the worst possible quality of life (as bad or worse than being dead) overall, how would you rate your quality of life?				
1	1 (1.0)	1 (1.7)	0 (0.0)	0.33
2	1 (1.0)	0 (0.0)	1 (2.6)	
3	4 (4.2)	1 (1.7)	3 (7.9)	
4	5 (5.2)	3 (5.2)	2 (5.3)	
5	15 (15.6)	11 (19.0)	4 (10.5)	
6	11 (11.5)	5 (8.6)	6 (15.8)	

Table 1 continued

	Total <i>N</i> (%) Mean (SD) <i>N</i> = 96	Patients with follow-up data <i>N</i> (%) Mean (SD) <i>N</i> = 58	Patients without follow-up data ^a <i>N</i> (%) Mean (SD) <i>N</i> = 38	<i>P</i> -value
7	31 (32.3)	18 (31.0)	13 (34.2)	
8	21 (21.9)	13 (22.4)	8 (21.1)	
9	6 (6.3)	5 (8.6)	1 (2.6)	
10	1 (1.0)	1 (1.7)	0 (0.0)	

^a Mean number of months between baseline and follow up survey is 17.9 (SD = 4.2; range = 11.1–24.0 months; IQR = 13.6–22.1 months). Two-sample *t*-test was used for age and years of schooling. Chi-square test was used for gender, race, marital status, highest degree, employment, and recruited at VA. Wilcoxon rank-sum test was used for all others

(all the 96 participants had baseline data). The only significant difference was ethnicity, with a higher proportion of white participants in the follow up.

Descriptive statistics and reliability

There were noteworthy floor effects for the SF-36 Role Limitations—Physical scale (51% of sample scored the possible minimum) and ceiling effects for the SF-36 Role Limitations—Emotional scale (75% of the sample scored the possible maximum, see Table 2). On the PDQ-39, there was a ceiling effect for the Social Support scale (54% of the sample scored the possible maximum). On the PDQUALIF, there was substantial ceiling effects for the Independence scale (60% of the sample scored the possible maximum).

Internal consistency reliability was satisfactory for all the eight SF-36 scales (Cronbach's alpha > 0.70). However, coefficient alpha for three of the seven PDQUALIF scales (Physical Function, Outlook, and Sleep scales) fell below the 0.70 threshold for adequate reliability to make group comparisons (ranging from 0.52 to 0.60). Likewise, alphas for two of eight PDQ-39 scales (Cognition and Bodily Discomfort scales) were 0.59 and 0.68.

Relative validity

Criterion variable #1: how PD affects you on a day-to-day basis

Ten patients had no difficulties, 34 reported mild difficulties, 38 patients had moderate difficulties and 14 patients had high levels or extreme difficulties (see Table 3). The PDQUALIF Social/Role Function scale and the PDQ-39 Mobility scale had the highest relative validity (13.3 and 11.7). The level of discrimination across the four categories of the criterion variable was higher for the overall score of the two PD-targeted measures (relative validity = 9.3 for

the PDQUALIF, and 10.6 for the PDQ-39) compared to either composite score of the SF-36 (relative validity = 5.5 for SF-36 PCS, 2.9 for SF-36 MCS).

Criterion variable #2: rating of overall QOL

Twenty-eight patients who rated their overall QOL as 8, 9, or 10 were combined into one group, 42 patients whose ratings were 6 or 7 were combined into another group, and 26 patients whose ratings were 1–5 were combined into a third group (see Table 4). The highest relative validity was observed for the SF-36 Emotional Well Being scale (relative validity = 15.44). The SF-36 MCS had higher relative validity than the overall scores of the two PD-targeted HRQOL measures.

Similarly, for the other two criterion variables, the overall scores of the two PD-targeted measures did not perform appreciably better than the SF-36 PCS and MCS. (See Appendix Tables 8 and 9 for details.)

Responsiveness of HRQOL measures

For criterion variable #1 (“How PD affects you on a day-to-day basis”), 20 patients reported at least one level of worsening on the second interview and were categorized as “declined” and 23 patients were categorized as “unchanged.” The highest ES for any overall or composite score was for the SF-36 PCS (ES = −0.86), corresponding to a large ES (see Table 5).

For criterion variable #2 (rating of overall QOL), 16 patients reported at least two levels of worsening on the second interview and were categorized as “declined” versus 35 patients who rated within one point of baseline and were categorized as “unchanged” (see Table 6). The SF-36 MCS (ES = −1.06) had the highest ES, again corresponding to a large ES.

We found three of the eight SF-36 scales had a large ES for each criterion variable examined. In contrast, only the

Table 2 Descriptive statistics and reliability of HRQOL scales ($N = 96$)^a

Baseline	Number of items	Mean	SD	Percent scoring minimum (= 0)/maximum (= 100)	Cronbach's alpha
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)					
Physical Function	10	58.0	30.8	4.2/3.1	0.94
Role Limitations—Physical	4	28.4	35.3	51.0/10.4	0.81
Role Limitations—Emotional	3	76.4	41.9	21.9/75.0	0.98
Pain	2	63.2	26.5	0.0/15.6	0.85
Emotional Well-Being	5	69.7	18.1	1.0/0.0	0.86
Energy	4	47.0	22.8	3.1/0.0	0.92
General Health	5	52.4	22.0	2.1/1.0	0.76
Social Function	2	61.0	33.3	6.3/29.2	0.98
SF-36 composite scores (<i>T</i> -scores)					
Physical Health (PCS)	–	36.0	10.5	n/a	0.93 ^b
Mental Health (MCS)	–	49.1	12.0	n/a	0.97 ^b
PDQUALIF (range: 0–100, where 100 is best quality of life)					
Social/Role Function	9	52.6	26.0	1.0/1.0	0.88
Self-Image and Sexuality	7	60.3	23.4	0.0/4.2	0.80
Sleep	3	64.6	28.3	2.1/21.1	0.60
Outlook	4	61.1	21.3	0.0/3.2	0.55
Physical Function	5	70.9	17.4	0.0/2.1	0.52
Independence	2	79.2	32.9	11.6/60.0	0.89
Urinary Function	2	39.3	35.4	27.4/10.5	0.85
Total score	–	61.2	18.4	–	0.93 ^b
PDQ-39 (range: 0–100, where 100 is best quality of life)					
Mobility	10	62.2	29.3	2.1/5.2	0.93
Activities of Daily Living	6	65.3	25.0	2.1/5.2	0.89
Emotional Well-Being	6	74.3	21.5	1.0/9.4	0.86
Stigma	4	74.6	28.2	1.0/34.4	0.88
Social Support	3	83.2	25.9	2.1/54.2	0.85
Cognitions	4	73.4	20.4	1.0/10.4	0.68
Communication	3	75.8	23.4	1.0/29.2	0.75
Bodily Discomfort	3	62.2	22.7	1.0/7.3	0.59
Total score	–	71.4	18.0	–	0.96 ^b

^a All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b Calculated using Mosier's formula

PDQUALIF Social/Role Function scale had a large ES for the criterion variable on PD's day-to-day effects; the ES for the other six PDQUALIF and eight PDQ-39 scales were not as large for either criterion variable.

Contributions of PD-targeted and generic HRQOL content to explaining criterion variables

Using criterion variable #1 (PD's day-to-day effects) as the dependent variable, the following three SF-36 scales entered the multivariate model at $P < 0.10$: Social

Functioning, Physical Functioning, and Role Limitations—Physical (Model 1). In Model 2, the following three PDQUALIF items from the PD-targeted measures entered the model at $P < 0.05$ after forcing in the above three SF-36 scales: Financial strain (Self-Image/Sexuality scale), Adjust to change (Social Role Function scale), Sleep with partner (Sleep scale); the following two PDQ-39 items also entered: Getting around house (Mobility scale) and Memory (Cognition scale). The adjusted R^2 improved from 0.48 in Model 1 to 0.65 in Model 2.

Table 3 Relative validity of HRQOL scales by how PD affects on day-to-day basis rating ($N = 96$)^a

Scale	How Parkinson's disease affects you on a day-to-day basis?				<i>F</i> -ratio ^b	Relative validity ^c
	No difficulties ($n = 10$)	Mild difficulties ($n = 34$)	Moderate difficulties ($n = 38$)	High levels of or extreme difficulties ($n = 14$)		
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)						
Physical Function	88.5 ^d	71.6 ^d	48.2 ^e	29.6 ^f	15.72	6.72
Role Limitations—Physical	67.5 ^d	41.9 ^e	16.5 ^f	0.0 ^f	14.64	6.26
Role Limitations—Emotional	90.0 ^d	88.2 ^d	7.6 ^d	42.9 ^e	4.79	2.05
Pain	77.5 ^d	64.9 ^d	67.5 ^d	37.1 ^e	6.97	2.98
Emotional Well-Being	76.4 ^d	76.9 ^d	67.6 ^d	53.1 ^e	7.66	3.27
Energy	69.5 ^d	56.0 ^e	39.2 ^f	30.4 ^f	12.14	5.19
General Health	64.5 ^d	57.9 ^{d,e}	47.9 ^{e,f}	42.5 ^f	3.44	1.47
Social Function	88.8 ^d	79.4 ^d	49.7 ^e	26.8 ^f	19.13	8.18
SF-36 composite scores (<i>T</i> -scores)						
Physical Health (PCS)	47.7 ^d	39.2 ^e	33.2 ^e	27.1 ^f	12.86	5.50
Mental Health (MCS)	53.4 ^d	53.4 ^d	47.3 ^d	39.2 ^e	6.68	2.85
PDQUALIF (range: 0–100, where 100 is best quality of life)						
Social/Role Function	80.8 ^d	67.5 ^e	42.5 ^f	21.4 ^g	31.07	13.28
Self-Image and Sexuality	82.1 ^d	71.1 ^d	53.3 ^e	35.9 ^f	16.71	7.14
Sleep	85.0 ^d	70.6 ^{d,e}	57.2 ^e	54.5 ^e	4.00	1.71
Outlook	75.0 ^d	71.9 ^d	53.3 ^e	44.7 ^e	11.44	4.89
Physical Function	83.5 ^d	75.3 ^{d,e}	69.7 ^e	53.5 ^f	8.49	3.63
Independence	100.0 ^d	87.7 ^d	79.9 ^d	38.5 ^e	11.84	5.06
Urinary Function	55.0 ^d	47.1 ^{d,e}	29.3 ^e	36.5 ^{d,e}	2.34 ^c	1.00
Total score	80.2 ^d	70.2 ^e	55.0 ^f	40.7 ^g	21.64	9.25
PDQ-39 (range: 0–100, where 100 is best quality of life)						
Mobility	94.0 ^d	77.8 ^e	52.4 ^f	28.0 ^g	27.4	11.71
Activities of Daily Living	88.8 ^d	75.1 ^d	60.2 ^e	38.4 ^f	15.43	6.59
Emotional Well-Being	85.4 ^d	85.2 ^d	67.3 ^e	58.6 ^e	9.63	4.12
Stigma	81.9 ^d	86.0 ^d	72.2 ^d	48.2 ^e	7.57	3.24
Social Support	91.7 ^d	93.6 ^d	79.6 ^d	61.9 ^e	6.58	2.81
Cognitions	81.3 ^d	80.3 ^d	69.6 ^{d, e}	61.2 ^e	4.34	1.85
Communication	91.2 ^d	86.6 ^d	68.4 ^e	56.6 ^e	11.95	5.11
Bodily Discomfort	74.2 ^d	69.4 ^d	59.4 ^d	43.5 ^e	6.33	2.71
Total score	86.4 ^d	81.8 ^d	66.2 ^e	49.5 ^e	24.8	10.60

^a All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b One way between group ANOVAs of HRQOL scale and day-to-day effects of PD

^c Reference scale = PDQUALIF—Urinary Function (*F*-ratio = 2.34)

^{d, e, f, g} Means within a row with different letters differ significantly ($P < 0.05$; Duncan multiple range)

Using criterion variable #2 (rating of overall QOL), the following four SF-36 scales entered the model at $P < 0.10$: Role Limitations—Physical, General Health, Role Limitations—Emotional, and Emotional Well-Being (Model 1). Then in Model 2, the following three PDQUALIF items from the PD-targeted measures entered at $P < 0.05$ after forcing in to the above four SF-36 scales: Sexual ability (Self-image/Sexuality scale), Future and Ask for help (both

from the Outlook scale), and Independent hygiene (Independence scale); the following three PDQ-39 items also entered: Confined to the house (Mobility scale), Isolated and lonely (Emotional/Well-Being scale), and Concentration (Cognition scale). The adjusted R^2 improved from 0.45 in Model 1 to 0.65 in Model 2. (See Appendix Table 10 for stepwise results for these two criterion variables and also for criterion variables # 3 and # 4.)

Table 4 Relative validity of HRQOL scales by quality of life rating ($N = 96$)^a

Scale	On a scale of 1–10, where 10 is best possible quality of life and 1 is the worst possible quality of life (as bad or worse than being dead) overall, how would you rate your quality of life?				
	Values 8–10 ($n = 28$)	Values 6–7 ($n = 42$)	Values 1–5 ($n = 26$)	F -ratio ^b	Relative validity ^c
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)					
Physical Function	71.8 ^d	58.2 ^d	42.7 ^e	6.74	3.46
Role Limitations—Physical	50.9 ^d	26.8 ^e	6.7 ^f	13.42	6.88
Role Limitations—Emotional	100.0 ^d	83.3 ^d	39.7 ^e	21.45	11.00
Pain	73.6 ^d	63.5 ^{d,e}	51.5 ^e	5.06	2.59
Emotional Well-Being	80.7 ^d	73.4 ^e	51.9 ^f	30.10	15.44
Energy	62.9 ^d	45.1 ^e	33.1 ^f	15.29	7.84
General Health	65.4 ^d	52.9 ^e	37.7 ^f	13.42	6.88
Social Function	82.1 ^d	62.5 ^e	35.6 ^f	18.00	9.23
SF-36 composite scores (T -scores)					
Physical Health (PCS)	41.3 ^d	35.0 ^e	31.8 ^e	6.47	3.32
Mental Health (MCS)	56.7 ^d	51.0 ^e	37.7 ^f	28.05	14.38
PDQUALIF (range: 0–100, where 100 is best quality of life)					
Social/Role Function	67.8 ^d	52.8 ^e	35.2 ^f	13.04	6.69
Self-Image and Sexuality	74.6 ^d	59.6 ^e	45.6 ^f	12.75	6.54
Sleep	76.2 ^d	64.9 ^e	51.0 ^e	5.77	2.96
Outlook	72.3 ^d	64.1 ^d	43.3 ^e	17.73	9.09
Physical Function	80.4 ^d	68.7 ^e	64.2 ^e	7.13	3.66
Independence	86.6 ^d	80.4 ^{d,e}	69.0 ^e	1.95 ^e	1.00
Urinary Function	53.1 ^d	32.1 ^e	36.0 ^e	3.26	1.67
Total score	73.0 ^d	60.4 ^e	49.2 ^f	14.28	7.32
PDQ-39 (range: 0–100, where 100 is best quality of life)					
Mobility	77.1 ^d	63.2 ^e	44.6 ^f	9.93	5.09
Activities of Daily Living	75.9 ^d	65.6 ^e	53.4 ^f	6.05	3.10
Emotional Well-Being	85.9 ^d	77.9 ^e	55.9 ^f	19.56	10.03
Stigma	86.4 ^d	77.1 ^e	57.9 ^f	8.25	4.23
Social Support	96.7 ^d	86.3 ^d	63.8 ^e	14.75	7.56
Cognitions	80.1 ^d	71.4 ^{d,e}	69.2 ^e	2.33	1.19
Communication	88.5 ^d	75.0 ^e	63.5 ^f	9.09	4.66
Bodily Discomfort	73.2 ^d	59.9 ^e	53.9 ^e	5.79	2.97
Total score	82.9 ^d	72.0 ^e	57.8 ^f	18.29	9.38

^a All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are T -scores (mean = 50, SD = 10) calculated against a reference population

^b One way between group ANOVAs of HRQOL scale and quality of life rating

^c Reference scale = PDQUALIF—Independence (F -ratio = 1.95)

^{d, e, f} Means within a row with different letters differ significantly ($P < 0.05$; Duncan multiple range)

Evaluation of scoring of PD-targeted measures

Item-scale correlations using baseline data from the sample of 96 enrollees revealed that 18 of 32 PDQUALIF items correlated within 0.05 (one-half the standard error in this dataset) below or correlated more highly with other scales than the scales they were supposed to represent. Particularly problematic were the Outlook and Physical Functioning scales: three out of four Outlook scale items

correlated more highly on another scale than Outlook, and all the five Physical Functioning items correlated more highly with another scale than the Physical Functioning scale (see Appendix Table 11).

Item-scale correlations for five of the eight PDQ-39 scales in general provided support for the arrangement of items by scale using our criteria for item discrimination across scales. However, all three items from the Bodily Discomfort scale, two out of the four items from the

Table 5 Responsiveness indices: declined and unchanged groups based on “How Parkinson’s disease affects you on a day-to-day basis?”^a

	Baseline mean in declined group (SD)	Average change in declined group (SD)	Average change in the unchanged group (SD)	Effect size statistic	SRM	Guyatt	<i>F</i> (<i>P</i> value)
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)							
Physical Function	70.0 (26.7)	−22.8 (22.0)	−1.6 (22.9)	−0.85 ^b	−1.04	−0.99	9.54 (0.004)
Role Limitations—Physical	47.5 (41.3)	−16.3 (43.9)	6.9 (29.4)	−0.39 ^d	−0.37	−0.55	4.22 (0.05)
Role Limitations—Emotional	90.0 (30.8)	−11.7 (49.9)	−2.9 (43.7)	−0.38 ^d	−0.23	−0.27	0.38 (0.54)
Pain	68.4 (24.5)	−1.5 (30.0)	2.2 (26.5)	−0.06	−0.05	−0.06	0.18 (0.67)
Emotional Well-Being	77.8 (13.8)	−1.2 (19.0)	−3.0 (18.9)	−0.09	−0.06	−0.06	0.09 (0.76)
Energy	66.3 (17.2)	−19.5 (22.3)	−3.0 (23.7)	−1.13 ^b	−0.87	−0.82	5.44 (0.03)
General Health	69.8 (10.4)	−18.0 (17.7)	−4.1 (13.8)	−1.72 ^b	−1.02	−1.31	8.31 (0.006)
Social Function	81.9 (24.5)	−12.5 (39.5)	1.1 (28.2)	−0.51 ^c	−0.32	−0.44	1.72 (0.20)
SF-36 composite scores (<i>T</i> -scores)							
Physical Health (PCS)	41.3 (8.9)	−7.7 (10.1)	0.8 (6.9)	−0.86 ^b	−0.76	−1.11	10.51 (0.002)
Mental Health (MCS)	55.5 (8.5)	−2.5 (10.7)	−1.6 (10.8)	−0.29 ^d	−0.23	−0.22	0.08 (0.78)
PDQUALIF (range: 0–100, where 100 is best quality of life)							
Social/Role Function	69.4 (23.9)	−19.1 (21.7)	−1.7 (21.1)	−0.80 ^b	−0.88	−0.91	6.98 (0.01)
Self-Image and Sexuality	69.5 (22.9)	−5.9 (19.6)	7.8 (19.3)	−0.26 ^d	−0.30	−0.31	5.18 (0.03)
Sleep	72.5 (29.0)	−10.0 (23.8)	−12.8 (28.9)	−0.35 ^d	−0.42	−0.35	0.12 (0.73)
Outlook	74.7 (18.3)	−3.8 (18.5)	1.4 (19.1)	−0.21 ^d	−0.21	−0.20	0.79 (0.38)
Physical Function	73.0 (19.9)	−3.3 (19.7)	−11.1 (13.5)	−0.16	−0.17	−0.24	2.32 (0.14)
Independence	84.4 (31.9)	−20.6 (34.5)	−1.14 (32.0)	−0.65 ^c	−0.60	−0.64	3.61 (0.07)
Urinary Function	43.1 (31.0)	2.5 (36.9)	0.6 (44.4)	−0.08	0.07	0.06	0.02 (0.88)
Total score	69.5 (17.2)	−8.6 (15.2)	−2.4 (17.7)	−0.50 ^c	−0.57	−0.49	1.44 (0.24)
PDQ-39 (range: 0–100, where 100 is best quality of life)							
Mobility	75.4 (26.4)	−16.9 (26.1)	5.4 (16.2)	−0.64 ^c	−0.65	−1.04	11.62 (0.002)
Activities of Daily Living	68.3 (27.5)	−13.5 (27.1)	2.0 (15.1)	−0.49 ^d	−0.50	−0.89	5.59 (0.02)
Emotional Well-Being	85.0 (20.9)	−7.1 (14.3)	−1.5 (15.2)	−0.34 ^d	−0.49	−0.47	1.51 (0.23)
Stigma	80.9 (27.8)	4.7 (22.2)	10.9 (21.1)	0.17	0.21	0.22	0.88 (0.35)
Social Support	88.3 (27.1)	0.4 (34.0)	2.7 (18.2)	0.02	0.01	0.02	0.08 (0.78)
Cognitions	74.1 (17.4)	−7.2 (21.8)	−1.4 (12.9)	−0.41 ^d	−0.33	−0.56	1.18 (0.29)
Communication	83.1 (23.5)	−14.2 (16.2)	2.2 (21.4)	−0.60 ^c	−0.88	−0.66	7.8 (0.008)
Bodily Discomfort	63.8 (25.6)	−7.5 (23.6)	2.5 (20.6)	−0.29 ^d	−0.32	−0.36	2.22 (0.14)
Total score	77.4 (19.7)	−7.7 (15.9)	2.9 (11.1)	−0.39 ^d	−0.48	−0.69	6.44 (0.02)

^a Declined group declined at least one category from baseline to follow-up ($n = 20$); Unchanged group remained in the same category at follow-up ($n = 23$). All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b Large effect size (effect size greater than or equal to 0.80)

^c Medium effect size (effect size between 0.20 and 0.49)

^d Small effect size (effect size between 0.50 and 0.79)

Cognition scale, and one out of the three items from the Communication scale loaded similarly or more highly on other scales than on the scale in which they are placed (see Appendix Table 12).

Discussion

We analyzed and compared the psychometric properties of a widely used generic measure of HRQOL, the SF-36, and

two PD-targeted measures, the PDQ-39 and the PDQUALIF measures. While relative validity was somewhat better for the PD-targeted measures than the SF-36 on criterion variables that asked specifically about activities limited by PD or about PD symptoms, we found greater support for the responsiveness of the SF-36 than for the PD-targeted measures on both external criterion variables, including the variable on difficulties in day-to-day activities due to PD. Despite better responsiveness of the generic measure, however, multivariate regression models showed that items

from the PD-targeted measures tap into additional HRQOL content not covered by the SF-36 scales. An analysis of the PD-targeted measures revealed multiple problems with items correlating as highly or more highly with other scales than with the scale they were intended to represent, potentially accounting for the unanticipated finding of superior responsiveness of the SF-36 compared to the PD-targeted measures.

Few studies have compared the psychometric properties of the SF-36 with a PD-targeted measure. The responsiveness of the SF-36 and PDQ-39 was tested among 132 PD patients by administering it at baseline and at 4 months and asking a criterion question of whether there was change in the effect of PD on everyday life [10]. In that study, none of the PDQ-39 had a large ES, the PDQ-39 mobility scale showed a

Table 6 Responsiveness indices: declined and unchanged groups based on “On a scale of 1 to 10, where 10 is best possible quality of life and 1 is the worst possible quality of life (as bad or worse than being dead) overall, how would you rate your quality of life?”^a

	Baseline mean in declined group (SD)	Average change in declined group (SD)	Average change in the unchanged group (SD)	Effect size statistic	SRM	Guyatt	<i>F</i> (<i>P</i> value)
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)							
Physical Function	59.1 (29.2)	−12.6 (23.9)	−7.0 (24.3)	−0.43 ^d	−0.53	−0.52	0.58 (0.45)
Role Limitations—Physical	50.0 (44.7)	−15.6 (49.1)	4.5 (33.5)	−0.35 ^d	−0.31	−0.47	2.94 (0.09)
Role Limitations—Emotional	93.8 (25.0)	−12.5 (29.5)	−7.6 (51.2)	−0.50 ^c	−0.42	−0.24	0.13 (0.72)
Pain	69.8 (23.3)	−9.5 (32.8)	8.9 (29.7)	−0.40 ^d	−0.29	−0.32	3.95 (0.05)
Emotional Well-Being	80.5 (7.4)	−6.0 (19.3)	0.9 (16.4)	−0.81 ^b	−0.31	−0.37	1.75 (0.19)
Energy	60.6 (20.4)	−17.2 (24.2)	−4.7 (24.7)	−0.84 ^b	−0.71	−0.70	2.84 (0.10)
General Health	66.9 (12.8)	−15.9 (14.4)	−4.7 (18.9)	−1.24 ^b	−1.10	−0.84	4.46 (0.04)
Social Function	83.6 (26.9)	−19.5 (39.0)	6.4 (32.8)	−0.72 ^c	−0.50	−0.59	6.10 (0.02)
SF-36 composite scores (<i>T</i> -scores)							
Physical Health (PCS)	38.7 (9.8)	−5.7 (10.8)	−0.1 (9.5)	−0.58 ^c	−0.52	−0.60	3.52 (0.07)
Mental Health (MCS)	57.3 (5.0)	−5.3 (10.9)	−0.3 (10.5)	−1.06 ^b	−0.49	−0.50	2.44 (0.13)
PDQUALIF (range: 0–100, where 100 is best quality of life)							
Social/Role Function	67.2 (26.4)	−18.4 (22.7)	−1.6 (21.6)	−0.70 ^c	−0.81	−0.85	6.43 (0.02)
Self-Image and Sexuality	62.1 (25.4)	−6.9 (19.0)	4.3 (19.1)	−0.27 ^d	−0.36	−0.39	3.78 (0.06)
Sleep	71.4 (30.3)	−16.9 (25.9)	−10.4 (27.9)	−0.56 ^c	−0.65	−0.61	0.63 (0.43)
Outlook	66.8 (19.9)	0.8 (11.8)	2.5 (10.1)	0.04	0.07	0.08	0.10 (0.75)
Physical Function	71.6 (21.7)	−4.4 (17.5)	−4.9 (16.6)	−0.20 ^d	−0.25	−0.27	0.01 (0.93)
Independence	84.4 (34.3)	−18.0 (36.2)	6.4 (37.9)	−0.52 ^c	−0.50	−0.48	4.67 (0.04)
Urinary Function	43.0 (32.6)	−5.5 (35.4)	1.1 (42.0)	−0.17	−0.16	−0.13	0.29 (0.59)
Total score	66.6 (17.6)	−9.9 (13.5)	−0.4 (17.2)	−0.56 ^c	−0.73	−0.58	3.83 (0.06)
PDQ-39 (range: 0–100, where 100 is best quality of life)							
Mobility	68.9 (27.4)	−13.7 (26.1)	4.7 (22.8)	−0.50 ^c	−0.52	−0.60	6.59 (0.01)
Activities of Daily Living	64.8 (27.0)	−10.7 (20.5)	1.8 (26.4)	−0.40 ^d	−0.52	−0.41	2.80 (0.10)
Emotional Well-Being	83.9 (18.3)	−4.9 (8.8)	−3.0 (17.1)	−0.27 ^d	−0.56	−0.29	0.18 (0.67)
Stigma	77.7 (24.4)	4.3 (18.9)	13.2 (21.5)	0.18	0.23	0.20	2.03 (0.16)
Social Support	85.9 (28.5)	−2.1 (20.9)	−0.1 (23.5)	−0.07	−0.10	−0.09	0.08 (0.78)
Cognitions	75.4 (16.5)	−3.9 (22.6)	−2.1 (13.3)	−0.24 ^d	−0.17	−0.29	0.12 (0.73)
Communication	77.9 (23.2)	−4.2 (17.7)	−5.7 (20.1)	−0.18	−0.23	−0.20	0.07 (0.79)
Bodily Discomfort	64.6 (24.2)	−2.1 (18.6)	1.7 (20.9)	−0.09	−0.11	−0.10	0.38 (0.54)
Total score	74.9 (16.4)	−4.7 (11.5)	1.3 (13.3)	−0.29 ^d	−0.41	−0.35	2.38 (0.13)

^a Declined group declined two or more categories from baseline to follow-up ($n = 16$); Unchanged group had change of one point or less from baseline to follow-up ($n = 35$). All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b Large effect size (effect size between 0.20 and 0.49)

^c Medium effect size (effect size between 0.50 and 0.79)

^d Small effect size (effect size greater than or equal to 0.80)

medium ES, the ADL and Social Support scales had a small ES, and the other five PDQ-39 scales did not detect any change. The three PDQUALIF scales that had less than adequate internal consistency reliability in our study also did not perform well in the original study introducing the PDQUALIF (in the original study, a fourth scale Urinary Function also had Cronbach's alpha below 0.7) [11]. Using Hoehn and Yahr stage as a criterion variable for their relative validity analysis, they found that the *F*-statistic for the overall PDQUALIF (10.8) was only a bit higher than the SF-36 PCS (9.1) though considerably better than the SF-36 MCS (1.9). The poor performance of the SF-36 MCS is likely due to Hoehn and Yahr's emphasis on balance and mobility. When we used the PHQ-9 as a criterion variable in our study, the SF-36 MCS outperformed overall scores of the two PD-targeted HRQOL measures, as would be anticipated given that depression is a stronger component of mental health than physical health. Emphasis of some criterion variables on certain aspects of HRQOL was also observed in our study with respect to the PD day-to-day activities criterion variable, for which relative validity and responsiveness were stronger for physical and social functioning scales of all the measures than with scales tapping mental health.

The following limitations to our study should be noted. We recruited a convenience sample of 96 PD patients, and the portion of our analyses involving longitudinal data (responsiveness) was based on a subset of the 60% of the enrolled sample for whom we were able to collect follow-up data. Some of the sample sizes for subgroups in the responsiveness analyses were relatively small, and we recommend that our findings with regard to responsiveness be confirmed in samples having larger subgroups who changed. Power to detect a difference would have been increased with a larger sample size; for example, we observed almost significant *F*-statistic of 3.52 for the responsiveness of the SF-36 physical health summary score in Table 6. With a larger sample size we may have found this test statistic to be statistically significant.

There was a higher proportion of men than in the general PD population because about half of this study's sample was recruited from a VA. Another potential limitation is that criterion variables were all self-reported, and it would have been useful to also include a clinical measure such as the motor UPDRS, an examination recorded by a trained clinician, or the Hoehn and Yahr stage. While we administered all the measures using the same modality at different points in time, data

regarding the adequacy of telephone administration of the PDQUALIF is unknown.

The results of this study suggest that both generic and disease-targeted measures contribute important information about HRQOL. In the future, both generic and disease-targeted items tapping the same domain could be included together in an item bank and administered using computer adaptive testing [34].

Conclusion

A comparison of the psychometric properties between a generic and two PD-targeted HRQOL measures provides evidence for superior or equivalent responsiveness of the generic HRQOL measure over the PD-targeted HRQOL measures. However, the PD-targeted measures account for additional content beyond the generic HRQOL measure alone. The empirical findings related to lack of superior responsiveness of the PD-targeted measures relative to the SF-36 may in part be explained by inadequate scaling of the original PD-targeted measures.

The findings of this study provide support for use of a combination of generic and disease-targeted HRQOL measures in future research.

Acknowledgments We acknowledge with thanks Michele Maines, MSG, Sunberri Murphy, and Amelia Mittleman for bibliographic support, and Harvey Lopez and Jessica Herrera for their administrative assistance. Erin Jacob assisted in the data collection. We thank Drs. Jeff Bronstein, Yvette Bordelon, and Indu Subramanian for recruitment of patients. The research presented here was supported by the Department of Veterans Affairs, Veterans Health Administration, VA Health Services Research & Development Service project number PDI 01-158, and by NIH/NINDS NS038367 for the UCLA UDALL Parkinson Disease Center of Excellence. Eric Cheng was supported by an NINDS career development award (K23NS058571).

Conflict of interest None of the authors have a financial relationship with the organizations that sponsored the research. All authors have full control of all primary data and agree to allow the journal to review their data if requested.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

See Tables 7, 8, 9, 10, 11 and 12.

Table 7 Criterion variables #3 and #4

	Total <i>N</i> (%) Mean (SD) <i>N</i> = 96	Patients with follow-up data <i>N</i> (%) Mean (SD) <i>N</i> = 58	Patients without follow-up data <i>N</i> (%) Mean (SD) <i>N</i> = 38	Wilcoxon <i>P</i> -value
Overall, during the last 6 months, how would you rate the severity of your Parkinson's disease symptoms?				
No symptoms	1 (1.0)	1 (1.7)	0 (0.0)	0.13
Mild symptoms	39 (40.6)	27 (46.6)	12 (31.6)	
Moderate symptoms	45 (46.9)	24 (41.4)	21 (55.3)	
Severe symptoms	11 (11.5)	6 (10.3)	5 (13.2)	
PHQ-9 categories for planning and monitoring treatment				
May not need depression treatment (scores: 0–4)	37 (38.9)	25 (43.9)	12 (31.6)	0.11
Based on clinical judgment about treatment on duration of symptoms and functional impairment (scores: 5–14)	47 (49.5)	28 (49.1)	19 (50.0)	
Warrants treatment for depression using antidepressant, psychotherapy, and/or a combination treatment approach (scores: 15–27)	11 (11.6)	4 (7.0)	7 (18.4)	

Table 8 Relative validity of HRQOL scales by rating of Parkinson's disease symptoms (*N* = 96)^a

Scale	Rating of severity of Parkinson's disease symptoms			<i>F</i> -ratio ^b	Relative validity ^c
	No or mild symptoms (<i>n</i> = 40)	Moderate symptoms (<i>n</i> = 45)	Severe symptoms (<i>n</i> = 11)		
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)					
Physical Function	77.0 ^d	47.7 ^e	30.9 ^f	20.22	5.04
Role Limitations—Physical	51.3 ^d	12.2 ^e	11.4 ^e	20.23	5.04
Role Limitations—Emotional	90.0 ^d	68.2 ^{d,e}	60.6 ^e	4.01 ^c	1.00
Pain	72.8 ^d	60.4 ^d	39.6 ^e	8.38	2.09
Emotional Well-Being	76.5 ^d	66.8 ^{d,e}	57.1 ^e	6.81	1.70
Energy	56.9 ^d	42.6 ^e	29.6 ^f	9.17	2.29
General Health	60.6 ^d	49.6 ^d	34.1 ^e	7.98	1.99
Social Function	81.9 ^d	48.6 ^e	35.2 ^e	19.95	4.98
SF-36 composite scores (<i>T</i> -scores)					
Physical Health (PCS)	42.5 ^d	32.5 ^e	26.1 ^f	21.47	5.35
Mental Health (MCS)	53.1 ^d	46.9 ^{d,e}	42.9 ^e	4.77	1.19
PDQUALIF (range: 0–100, where 100 is best quality of life)					
Social/Role Function	69.5 ^d	42.5 ^e	30.6 ^e	22.55	5.62
Self-Image and Sexuality	75.5 ^d	52.1 ^e	36.8 ^f	24.42	6.09
Sleep	75.0 ^d	58.0 ^e	52.5 ^e	5.30	1.32
Outlook	71.6 ^d	55.8 ^d	42.5 ^e	12.44	3.10
Physical Function	79.4 ^d	65.0 ^e	64.0 ^e	9.56	2.38
Independence	89.4 ^d	76.1 ^d	52.5 ^e	6.00	1.50
Urinary Function	51.3 ^d	33.6 ^{d,e}	17.5 ^e	5.19	1.29
Total score	73.1 ^d	54.7 ^e	42.3 ^f	24.60	6.13
PDQ-39 (range: 0–100, where 100 is best quality of life)					
Mobility	82.4 ^d	51.6 ^e	31.9 ^f	29.22	7.29
Activities of Daily Living	79.3 ^d	58.7 ^e	41.3 ^f	17.27	4.31
Emotional Well-Being	84.7 ^d	69.9 ^e	54.2 ^f	13.01	3.24

Table 8 continued

Scale	Rating of severity of Parkinson's disease symptoms				<i>F</i> -ratio ^b	Relative validity ^c
	No or mild symptoms (<i>n</i> = 40)	Moderate symptoms (<i>n</i> = 45)	Severe symptoms (<i>n</i> = 11)			
Stigma	84.5 ^d	69.4 ^{d,e}	59.7 ^e		5.21	1.30
Social Support	95.4 ^d	78.2 ^e	59.9 ^f		12.06	3.01
Cognitions	83.3 ^d	66.9 ^e	63.6 ^e		9.71	2.42
Communication	90.1 ^d	66.9 ^e	59.1 ^e		19.02	4.74
Bodily Discomfort	72.3 ^d	55.6 ^e	52.3 ^e		7.93	1.98
Total score	84.0 ^d	64.7 ^e	52.7 ^f		31.29	7.80

^a All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b One way between group ANOVAs of HRQOL scale and PD symptom severity

^c Reference scale = SF-36 Role Limitations—Emotional (*F*-ratio = 4.01)

^{d, e, f} Means within a row with different letters differ significantly ($P < 0.05$; Duncan multiple range)

Table 9 Relative validity of HRQOL scales by PHQ-9 scoring categories ($N = 96$)^a

Scale	PHQ-9 scoring categories			<i>F</i> -ratio ^b	Relative validity ^c
	Depression treatment may not be needed (<i>n</i> = 37)	Clinical judgment about treatment on duration of symptoms and functional impairments (<i>n</i> = 47)	Warrants treatment for depression (<i>n</i> = 11)		
SF-36 v. 1.0 (range: 0–100, where 100 is best quality of life)					
Physical Function	73.7 ^d	49.8 ^e	44.6 ^e	8.88	1.55
Role Limitations—Physical	50.7 ^d	17.6 ^e	2.3 ^e	16.78	2.93
Role Limitations—Emotional	100.0 ^d	75.9 ^e	6.1 ^f	40.29	7.04
Pain	77.0 ^d	58.5 ^e	40.9 ^f	12.18	2.13
Emotional Well-Being	81.0 ^d	68.7 ^e	42.6 ^f	43.63	7.63
Energy	66.2 ^d	40.5 ^e	24.6 ^f	26.50	4.63
General Health	62.3 ^d	49.6 ^e	35.9 ^f	8.80	1.54
Social Function	84.1 ^d	52.4 ^e	25.0 ^f	26.88	4.70
SF-36 composite scores (<i>T</i> -scores)					
Physical Health (PCS)	41.6 ^d	33.1 ^e	32.4 ^e	10.06	1.76
Mental Health (MCS)	56.8 ^d	48.4 ^e	28.3 ^f	51.97	9.09
PDQUALIF (range: 0–100, where 100 is best quality of life)					
Social/Role Function	72.5 ^d	44.1 ^e	22.0 ^f	37.13	6.49
Self-Image and Sexuality	74.5 ^d	54.2 ^e	39.0 ^f	17.59	3.08
Sleep	82.7 ^d	57.3 ^e	34.9 ^f	21.97	3.84
Outlook	75.34 ^d	55.9 ^e	34.6 ^f	28.75	5.03
Physical Function	78.5 ^d	68.1 ^e	57.7 ^f	8.43	1.47
Independence	90.9 ^d	75.5 ^d	55.8 ^e	6.03	1.05
Urinary Function	53.4 ^d	32.5 ^e	21.6 ^e	5.72 ^c	1.00
Total score	75.4 ^d	55.4 ^e	37.9 ^f	41.04	7.18
PDQ-39 (range: 0–100, where 100 is best quality of life)					
Mobility	82.0 ^d	52.1 ^e	40.0 ^e	20.11	3.52
Activities of Daily Living	78.8 ^d	60.6 ^e	40.9 ^f	14.71	2.57
Emotional Well-Being	89.3 ^d	69.8 ^e	43.2 ^f	38.00	6.64
Stigma	86.0 ^d	70.8 ^d	53.9 ^e	7.18	1.26
Social Support	94.8 ^d	79.4 ^e	63.6 ^f	8.60	1.50
Cognitions	82.8 ^d	70.6 ^e	51.1 ^f	13.99	2.45
Communication	88.1 ^d	71.5 ^e	53.0 ^f	13.92	2.43

Table 9 continued

Scale	PHQ-9 scoring categories				
	Depression treatment may not be needed (<i>n</i> = 37)	Clinical judgment about treatment on duration of symptoms and functional impairments (<i>n</i> = 47)	Warrants treatment for depression (<i>n</i> = 11)	<i>F</i> -ratio ^b	Relative validity ^c
Bodily Discomfort	74.8 ^d	57.4 ^e	41.7 ^f	14.34	2.51
Total score	84.6 ^d	66.5 ^e	48.4 ^f	35.89	6.27

^a All the scales scored on a 0 (worst) to 100 (best) possible range, except for the SF-36 PCS and MCS, which are *T*-scores (mean = 50, SD = 10) calculated against a reference population

^b One way between group ANOVAs of HRQOL scale and PHQ-9 scoring categories

^c Reference scale = PDQUALIF—Urinary Function (*F*-ratio = 5.72)

^{d, e, f} Means within a row with different letters differ significantly (*P* < 0.05; Duncan multiple range)

Table 10 Stepwise regression (*N* = 96)

Model 1: Independent variable = SF-36 scales			Model 2: Independent variable = significant SF-36 scales from Model 1 (<i>P</i> < 0.10) + PDQUALIF items + PDQ-39 items		
<i>R</i> ² ; Adjusted <i>R</i> ²	Scales	<i>P</i> value	<i>R</i> ² ; Adjusted <i>R</i> ²	Scales	<i>P</i> value
Criterion variable #1: How PD affects on day-to-day basis			Criterion variable #1: How PD affects on day-to-day basis		
0.50; 0.48	Social Function	0.001	0.67; 0.65	Social Function ^a	0.09
	Physical Function	0.003		Physical Function ^a	0.005
	Role Limitations—Physical	0.01		Role Limitations—Physical ^a	0.004
				PDQUALIF items	
				Financial strain	Self-image/sexuality 0.04
				Adjust to change	Social Role Function 0.003
				Sleep with partner	Sleep 0.03
				PDQ-39 items	
				Getting around house	Mobility 0.006
				Memory	Cognitions 0.002
Criterion variable #2: Quality of life rating			Criterion variable #2: Quality of life rating		
0.47; 0.45	Role Limitations—Physical	0.02	0.69; 0.65	Role Limitations—Physical ^a	0.19
	General Health	0.01		General Health ^a	0.002
	Role Limitations—Emotional	0.04		Role Limitations—Emotional ^a	0.59
	Emotional Well-Being	0.06		Emotional Well-Being ^a	0.004
				PDQUALIF items	
				Sexual ability	Self-image/sexuality 0.005
				Future	Outlook <0.001
				Ask for help	0.02
				Independent hygiene	Independence 0.02
				PDQ-39 items	
				Confined to the house	Mobility 0.001
				Isolated and lonely	Emotional Well-Being 0.006
				Concentration	Cognitions 0.006
Criterion variable #3: Rating of severity of PD symptoms			Criterion variable #3: Rating of severity of PD symptoms		
0.42; 0.40	Physical Function	0.006	0.71; 0.67	Physical Function ^a	0.10
	Social Function	0.03		Social Function ^a	0.93
	General Health	0.07		General Health ^a	0.24

Table 10 continued

Model 1: Independent variable = SF-36 scales			Model 2: Independent variable = significant SF-36 scales from Model 1 ($P < 0.10$) + PDQUALIF items + PDQ-39 items		
R^2 ; Adjusted R^2	Scales	P value	R^2 ; Adjusted R^2	Scales	P value
	Role Limitations—Physical	0.07		Role Limitations—Physical ^a	0.02
			PDQUALIF items		
			Imbalance	Social/Role Function	0.001
			Usual share of work in home		0.02
			Self-concept/image	Self-Image/Sexuality	0.009
			Sexual ability		0.005
			PDQ-39 items		
			Memory	Cognitions	<0.001
			Hallucinations		0.008
Criterion variable #4: PHQ-9 scoring categories			Criterion variable #4: PHQ-9 scoring categories		
0.65; 0.64	Emotional well being	0.008	0.81; 0.78	Emotional well being ^a	0.51
	Pain	0.009		Pain ^a	0.02
	Energy	0.001		Energy ^a	0.03
	Role Limitations—Emotional	0.007		Role Limitations—Emotional ^a	0.003
			PDQUALIF items		
			Financial strain	Self-Image/Sexuality	0.002
			Sleep maintenance	Sleep	0.001
			Outlook	Outlook	<0.001
			PDQ-39 items		
			Close relationship	Social Support	0.001
			Concentration	Cognitions	0.008

^a Forced into Model 2 based on significance in Model 1

Table 11 Multitrait scaling of PDQUALIF items ($N = 96$)

	Social/Role Function	Self-Image and Sexuality	Sleep	Outlook	Physical Function	Independence	Urinary Function
Social/Role Function							
Work in home	0.72 ^a	0.53	0.46	0.48	0.54	0.66	0.26
Social life I	0.56 ^a	0.49	0.19	0.42	0.26	0.14	0.03
Imbalance	0.60 ^a	0.51	0.30	0.42	0.49	0.57	0.34
Social life II	0.79 ^a	0.69	0.39	0.60	0.45	0.31	0.21
Burden	0.60 ^a	0.58	0.38	0.58	0.42	0.33	0.27
Fatigue	0.61 ^a	0.52	0.52	0.52	0.43	0.35	0.28
Social isolation	0.80 ^a	0.66	0.39	0.57	0.47	0.34	0.24
Travel	0.32 ^a	0.17	0.25	0.20	0.44	0.28	0.07
Adjust to change	0.55 ^a	0.45	0.37	0.61	0.34	0.28	0.12
Self Image and Sexuality							
Self concept	0.38	0.53 ^a	−0.04	0.38	0.35	0.14	0.05
Sexual ability	0.35	0.33 ^a	0.23	0.18	0.25	0.33	0.44
Family relationship	0.63	0.62 ^a	0.33	0.61	0.48	0.18	0.23
Sexual desirability	0.65	0.67 ^a	0.28	0.67	0.44	0.32	0.21
Communication I	0.51	0.56 ^a	0.37	0.42	0.35	0.25	0.17
Communication II	0.50	0.56 ^a	0.33	0.34	0.30	0.29	0.24
Financial strain	0.37	0.48 ^a	0.19	0.26	0.36	0.27	0.15

Table 11 continued

	Social/Role Function	Self-Image and Sexuality	Sleep	Outlook	Physical Function	Independence	Urinary Function
Sleep							
Sleep initiation	0.53	0.38	0.63 ^a	0.46	0.38	0.33	0.41
Sleep maintenance	0.43	0.26	0.56 ^a	0.37	0.27	0.17	0.48
Sleep with partner	<i>0.16</i>	<i>0.17</i>	0.14 ^a	<i>0.12</i>	<i>0.21</i>	<i>0.20</i>	0.22
Outlook							
Future	<i>0.51</i>	0.44	0.30	0.52 ^a	0.16	0.09	0.18
Maintain independence	<i>0.13</i>	0.03	<i>0.15</i>	−0.11 ^a	<i>0.17</i>	0.29	0.05
Ask for help	<i>0.41</i>	0.49	0.23	0.37 ^a	0.29	0.05	0.11
Outlook	0.68	0.53	0.42	0.66 ^a	0.37	0.28	0.21
Physical Function							
Neuropathy	0.37	0.29	0.20	0.13	0.35 ^a	0.27	0.11
Dizziness	<i>0.31</i>	0.32	0.05	<i>0.31</i>	0.23 ^a	0.07	−0.06
Swallow	<i>0.41</i>	0.49	0.31	<i>0.32</i>	0.37 ^a	<i>0.43</i>	0.15
Driving	0.33	<i>0.14</i>	<i>0.14</i>	<i>0.13</i>	0.18 ^a	<i>0.23</i>	0.04
Constipation	<i>0.31</i>	<i>0.34</i>	<i>0.41</i>	<i>0.25</i>	0.22 ^a	<i>0.39</i>	0.50
Independence							
Independent hygiene	0.51	0.39	0.28	0.23	0.49	0.73 ^a	0.29
Independent food prep	0.43	0.32	0.29	0.24	0.40	0.73 ^a	0.22
Urinary Function							
Nocturia	0.20	0.29	0.47	0.18	0.18	0.20	0.80 ^a
Uro frequency	0.33	0.32	0.49	0.24	0.32	0.32	0.80 ^a

Bold values = highest item-scale correlation for a given item, unless the highest item-scale correlation is with the scale in which that item is placed. Italics values = item-scale correlation is less than a half of standard error (0.05) below or is higher than the correlation of that item with the scale in which it is placed

^a Item's original scale

Table 12 Multitrait scaling of PDQ-39 items ($N = 96$)

	Mobility	Activities of Daily Living	Emotion Well-Being	Stigma	Social Support	Cognitions	Communication	Bodily Discomfort
Mobility								
Leisure activities	0.65 ^a	0.58	0.54	0.37	0.51	0.49	0.49	0.42
Looking after home	0.79 ^a	0.63	0.36	0.28	0.22	0.36	0.46	0.56
Carry shopping bags	0.87 ^a	0.67	0.33	0.23	0.22	0.25	0.42	0.46
Walk half mile	0.83 ^a	0.64	0.30	0.20	0.33	0.28	0.37	0.46
Walk 100 yards	0.79 ^a	0.59	0.25	0.23	0.24	0.22	0.32	0.41
Getting around house	0.66 ^a	<i>0.62</i>	0.24	0.21	0.27	0.26	0.42	0.38
Getting around public	0.73 ^a	0.62	0.37	0.31	0.36	0.37	0.53	0.48
Accompanied when out	0.62 ^a	0.52	0.25	0.16	0.15	0.18	0.25	0.42
Worried about falling	0.70 ^a	0.56	0.36	0.30	0.28	0.39	0.42	0.46
Confined to house	0.64 ^a	0.51	0.50	0.28	0.44	0.41	0.43	0.44
Activities of Daily Living								
Washing	0.76	0.76 ^a	0.24	0.13	0.24	0.38	0.43	0.43
Dressing	0.75	0.84 ^a	0.25	0.19	0.25	0.41	0.39	0.48
Buttons and laces	0.64	0.74 ^a	0.22	0.16	0.24	0.44	0.32	0.42
Writing	0.46	0.62 ^a	0.26	0.17	0.22	0.39	0.50	0.33
Cutting food	0.63	0.85 ^a	0.31	0.26	0.27	0.48	0.53	0.49

Table 12 continued

	Mobility	Activities of Daily Living	Emotion Well-Being	Stigma	Social Support	Cognitions	Communication	Bodily Discomfort
Holding drink	0.50	0.57 ^a	0.45	0.40	0.27	0.43	0.45	0.51
Emotional Well-Being								
Depressed	0.33	0.29	0.76 ^a	0.44	0.54	0.45	0.49	0.43
Isolated and lonely	0.51	0.41	0.71 ^a	0.55	0.61	0.46	0.52	0.50
Weepy and tearful	0.39	0.32	0.78 ^a	0.50	0.54	0.46	0.52	0.50
Angry and bitter	0.23	0.24	0.58 ^a	0.46	0.46	0.37	0.38	0.41
Anxious	0.35	0.21	0.59 ^a	0.50	0.48	0.29	0.34	0.27
Worried about future	0.27	0.19	0.63 ^a	0.42	0.40	0.30	0.27	0.30
Stigma								
Conceal PD	0.14	0.07	0.42	0.58 ^a	0.36	0.18	0.13	0.31
Eating/drinking public	0.45	0.44	0.60	0.70 ^a	0.48	0.39	0.43	0.50
Embarrassed	0.25	0.21	0.48	0.80 ^a	0.41	0.23	0.35	0.35
Others reactions	0.28	0.19	0.59	0.89 ^a	0.46	0.26	0.39	0.41
Social Support								
Close relationships	0.40	0.30	0.64	0.47	0.73 ^a	0.42	0.47	0.55
Support spouse	0.29	0.23	0.50	0.42	0.69 ^a	0.32	0.37	0.37
Support friends/family	0.33	0.28	0.59	0.43	0.76 ^a	0.28	0.43	0.34
Cognitions								
Fall asleep day	0.25	0.28	0.17	0.09	0.15	0.45 ^a	0.17	0.28
Concentration	0.35	0.40	0.47	0.37	0.33	0.46 ^a	0.40	0.43
Memory	0.38	0.47	0.46	0.26	0.37	0.59 ^a	0.50	0.46
Hallucinations	0.19	0.33	0.32	0.17	0.25	0.36 ^a	0.23	0.41
Communication								
Speech	0.45	0.50	0.35	0.24	0.33	0.32	0.69 ^a	0.36
Communicate	0.43	0.43	0.39	0.28	0.31	0.33	0.73 ^a	0.29
Ignored	0.41	0.41	0.67	0.48	0.63	0.53	0.42 ^a	0.53
Bodily Discomfort								
Muscle cramps/spasms	0.32	0.40	0.29	0.43	0.37	0.43	0.30	0.42 ^a
Aches and pains	0.45	0.34	0.29	0.26	0.38	0.31	0.20	0.40 ^a
Hot or cold	0.50	0.46	0.55	0.33	0.30	0.49	0.50	0.37 ^a

Bold values = highest item-scale correlation for a given item, unless the highest item-scale correlation is with the scale in which that item is placed. Italics values = item-scale correlation is less than a half of standard error (0.05) below or is higher than the correlation of that item with the scale in which it is placed

^a Item's original scale

References

- Nutt, J. G., & Wooten, G. F. (2005). Diagnosis and initial management of Parkinson's disease. *New England Journal of Medicine*, 353(10), 1021.
- Lang, A. E., & Lozano, A. M. (1998). Parkinson's disease—second of two parts. *New England Journal of Medicine*, 339(16), 1130.
- Damiano, A. M., Snyder, C., Strausser, B., & Willian, M. K. (1999). A review of health-related quality-of-life concepts and measures for Parkinson's disease. *Quality of Life Research*, 8(3), 235–243.
- Ustun, T. B., Chatterji, S., Bickenbach, J., Kostanjsek, N., & Schneider, M. (2003). The international classification of functioning, disability and health: A new tool for understanding disability and health. *Disability and Rehabilitation*, 25(11–12), 565–571.
- Hays, R. D., Hahn, H., & Marshall, G. (2002). Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Archives of Physical Medicine and Rehabilitation*, 83(12 Suppl 2), S4–S9.
- Ware, J. E. (1993). *SF-36 health survey: Manual and interpretation guide*. Boston: The Health Institute, New England Medical Center.
- Vickrey, B. G., Hays, R. D., Genovese, B. J., Myers, L. W., & Ellison, G. W. (1997). Comparison of a generic to disease-targeted health-related quality-of-life measures for multiple sclerosis. *Journal of Clinical Epidemiology*, 50(5), 557–569.
- Vickrey, B. G., Hays, R. D., Graber, J., Rausch, R., Engel, J., Jr., & Brook, R. H. (1992). A health-related quality of life instrument

- for patients evaluated for epilepsy surgery. *Medical Care*, 30(4), 299–319.
9. Peto, V., Jenkinson, C., & Fitzpatrick, R. (1998). PDQ-39: A review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. *Journal of Neurology*, 245(Suppl 1), S10–S14.
 10. Fitzpatrick, R., Peto, V., Jenkinson, C., Greenhall, R., & Hyman, N. (1997). Health-related quality of life in Parkinson's disease: A study of outpatient clinic attenders. *Movement Disorders*, 12(6), 916–922.
 11. Welsh, M., McDermott, M. P., Holloway, R. G., Plumb, S., Pfeiffer, R., & Hubble, J. (2003). Development and testing of the Parkinson's disease quality of life scale. *Movement Disorders*, 18(6), 637–645.
 12. Wiebe, S., Guyatt, G., Weaver, B., Matijevic, S., & Sidwell, C. (2003). Comparative responsiveness of generic and specific quality-of-life instruments. *Journal of Clinical Epidemiology*, 56(1), 52–60.
 13. Den Oudsten, B. L., Van Heck, G. L., & De Vries, J. (2007). The suitability of patient-based measures in the field of Parkinson's disease: A systematic review. *Movement Disorders*, 22(10), 1390–1401.
 14. Hays, R. D., Kim, S., Spritzer, K. L., Kaplan, R. M., Tally, S., Feeny, D., et al. (2009). Effects of mode and order of administration on generic health-related quality of life scores. *Value Health*. doi:10.1111/j.1524-4733.2009.00566.x.
 15. Jenkinson, C., Fitzpatrick, R., & Peto, V. (1999). Health-related quality-of-life measurement in patients with Parkinson's disease. *Pharmacoeconomics*, 15(2), 157–165.
 16. Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics*, 17(1), 13–35.
 17. Hobson, J. P., Edwards, N. I., & Meara, R. J. (2001). The Parkinson's disease activities of daily living scale: A new simple and brief subjective measure of disability in Parkinson's disease. *Clinical Rehabilitation*, 15(3), 241–246.
 18. Hadorn, D. C., & Hays, R. D. (1991). Multitrait-multimethod analysis of health-related quality-of-life measures. *Medical Care*, 29(9), 829–840.
 19. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
 20. Lowe, B., Unutzer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. *Medical Care*, 42(12), 1194–1201.
 21. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. (2003). The unified Parkinson's disease rating scale (UPDRS): Status and recommendations. *Movement Disorders*, 18(7), 738–750.
 22. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
 23. Mosier, C. (1943). On the reliability of a weighted composite. *Psychometrika*, 8(3), 161–168.
 24. Riazi, A., Hobart, J. C., Lamping, D. L., Fitzpatrick, R., Freeman, J. A., Jenkinson, C., et al. (2003). Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. *Journal of Neurology, Neurosurgery and Psychiatry*, 74(6), 710–714.
 25. Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
 26. Liang, M. H., Larson, M. G., Cullen, K. E., & Schwartz, J. A. (1985). Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis and Rheumatism*, 28(5), 542–547.
 27. Fayers, P. M., & Hays, R. D. (2005). *Assessing quality of life in clinical trials*. New York: Oxford University Press.
 28. Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11(3), 193–205.
 29. Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*, 4(1), 70.
 30. Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, 40(2), 171–178.
 31. Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. London: Academic Press.
 32. Kim, S., Hays, R. D., Birbeck, G. L., & Vickrey, B. G. (2003). Responsiveness of the quality of life in epilepsy inventory (QOLIE-89) in an antiepileptic drug trial. *Quality of Life Research*, 12(2), 147–155.
 33. Hays, R. D., & Wang, E. (1992). Multitrait scaling program: MULTI. In *Proceedings of the seventeenth annual SAS users group international conference* (pp. 1151–1156).
 34. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care*, 45(5 Suppl 1), S22–S31.