

## PAPER

# Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments

J Marinus, C Ramaker, J J van Hilten, A M Stiggelbout

*J Neurol Neurosurg Psychiatry* 2002;**72**:241–248

See end of article for authors' affiliations

Correspondence to:  
J Marinus, Leiden University Medical Center, Department of Neurology (K5 Q), PO Box 9600, NL-2300, RC Leiden, The Netherlands;  
J.Marinus@lumc.nl

Received 27 April 2001  
In revised form 27 July 2001  
Accepted 27 September 2001

**Objective:** To compare and contrast disease-specific quality of life instruments in Parkinson's disease and assess their clinimetric properties.

**Methods:** Two reviewers independently evaluated both thoroughness and results of studies regarding clinimetric characteristics of identified scales.

**Results:** Twenty studies were found reporting on the clinimetric properties of four scales. The content validity of the Parkinson's disease questionnaire-39 item version (PDQ-39), the Parkinson's disease quality of life questionnaire (PDQL), and the "Fragebogen Parkinson LebensQualität" (Parkinson Quality of Life questionnaire; PLQ) was adequate to good, but for the Parkinson's impact scale (PIMS) it was insufficient. Construct validity of both the PDQ-39 and the PDQL was good, but for the PLQ and the PIMS this was insufficiently evaluated. Internal consistency of all scale totals and of subscale totals of the PDQL were good, whereas for the social support subscale of the PDQ-39 and four subscales of the PLQ this was inadequate. Test-retest reliability was not evaluated for the PDQL and was adequate in the other scales. Responsiveness was partially established for the PDQ-39, and not assessed for the other scales. The number of available translations, as well as the number of studies in which these instruments were used, differed considerably.

**Conclusions:** The selection of an instrument partially depends on the goal of the study. In many situations however, the PDQ-39 will probably be the most appropriate HRQoL instrument. The PDQL may be considered as an alternative, whereas the PLQ may be considered in studies involving German speaking patients with Parkinson's disease. Use of the PIMS should be considered only as a means of identifying areas of potential problems.

Quality of life (QoL) is a multidimensional concept that reflects a subjective evaluation of a person's satisfaction with life and concerns, among others, the relationships with family or relatives, a person's own health, the health of another close person, finances, housing, independence, religion, social life, and leisure activities.<sup>1</sup> Health contributes to QoL, and this domain is often referred to as "health related quality of life" (HRQoL). The World Health Organisation (WHO) describes health as a state of complete physical, mental, social, and spiritual wellbeing, and not merely the absence of disease or infirmity.<sup>2</sup> This indicates that psychological and social factors are an integral part of health. Sometimes "role functioning" is added as a separate entity to the concept of HRQoL. Bowling<sup>3</sup> takes several definitions of HRQoL into account and defines the concept as optimum levels of mental, physical, role (work, parent, carer, etc), and social functioning, including relationships, and perceptions of health, fitness, life satisfaction, and wellbeing. The HRQoL gained importance in the past 3 decades and is considered to be an important outcome measure in studies involving patients with chronic diseases. Although initially physician based evaluations were chosen as primary end points in clinical research, more recent studies often consider HRQoL as their main outcome measure.

The HRQoL can be assessed both with generic and disease specific instruments. The generic instruments (for example, medical outcomes study—short form 36, sickness impact profile) offer the possibility of comparing HRQoL across different diseases. These instruments contain items of a more general nature, and therefore lack specificity. Disease specific instruments generally tap the same domains, but the items are tailored to particular disease characteristics and may also include items dealing with side effects of therapy. Consequently, disease specific instruments better reflect the consequences of

that disease to a particular person and generally are more sensitive to change in perceived HRQoL.<sup>4</sup>

In Parkinson's disease several disease specific HRQoL instruments have become available in the past few years. Investigators who want to use such an instrument are faced with the choice between several scales, which differ in many respects. In the process of selecting the appropriate instrument, a comparison of the quality of these scales can be helpful. We therefore compared and contrasted HRQoL instruments in Parkinson's disease and evaluated their clinimetric properties.

## METHODS

### Search strategy

We reviewed the literature from 1965 to 2000 and used the following sources to identify studies of interest: Medline, Embase, SCISEARCH, the Cochrane Library, symposia reports, Parkinson's disease handbooks, and reference lists of included publications. We used the following search terms: Parkinson disease, quality of life, health status, PDQ39, PDQL, Parkinson's disease questionnaire, PIMS, Parkinson's impact scale, PLQ, Fragebogen Parkinson Lebensqualität, PDQUALIF. These terms were combined with the following terms: clinimetric, psychometric, reliability, validity, internal consistency, factor analysis, factor structure, responsiveness, and sensitivity to

**Abbreviations:** PDQ-39, Parkinson's disease questionnaire-39 item version; PDQL, Parkinson's disease quality of life questionnaire; PLQ, Fragebogen Parkinson Lebensqualität (Parkinson quality of life questionnaire); PIMS, Parkinson's impact scale; QoL, quality of life; HRQoL, health related quality of life; ICC, intraclass correlation coefficients;  $\kappa$ , Kappa

change. The list of publications regarding each scale was sent to the developer, with the request to add references in the case of incompleteness.

### Methods of the review

Two reviewers independently reviewed the identified publications according to a two step review process. Firstly, abstracts were reviewed for eligibility. Thereafter, eligible reports were judged against a set of methodological criteria, in which both *thoroughness* (methodological and statistical) and *results* of studies testing validity, reliability, and responsiveness were assessed. To this extent we used a checklist, evaluating sample characteristics, outcome measures, appropriateness of statistical analysis, and methodological quality. The method of presenting the quality of scales was adopted from McDowell and Newell.<sup>5</sup> For reliability, Cronbach's  $\alpha$  greater than 0.7 and intraclass correlation coefficients (ICCs) or  $\kappa$  greater than 0.7 were considered a good result, and studies were judged "thorough" when the appropriate statistical procedures were used and the sample size was considered to be large enough. With respect to validity, the result of content validity was considered "good" if all relevant domains were covered and "thorough" if unselected (community based) patients with Parkinson's disease were closely involved in both the generation and evaluation of items. When only outpatient samples or samples from a Parkinson's disease society were used in this phase, we considered thoroughness to be moderate and when the patients were not involved at all, we considered the thoroughness to be poor.

Discrepancies were registered and resolved by consensus with a third and fourth reviewer.

Studies were eligible when they evaluated the following clinimetric characteristics of disease specific HRQoL instruments in Parkinson's disease: validity (content validity and construct validity, including factor structure), reliability (internal consistency, test-retest reliability), and responsiveness. Content validity reflects the extent to which a scale covers all important topics or domains.<sup>6</sup> Construct validity is assessed by measuring the extent to which a scale correlates positively with other measures that address the same construct (convergent validity), or negatively with measures that address opposite constructs (divergent validity), in situations where a gold standard is not available.

Another method of construct validation is the analysis of "known groups" differences. In this method patients are grouped on the basis of some characteristic—for example, disease severity or difficulties in performing activities. Patients with higher disease severity or patients experiencing greater difficulty, are expected to have lower HRQoL.

In a factor analysis, items that correlate highly with each other group together in clusters (factors), that are considered to reflect underlying common themes. Factor analysis may be used to construct subscales, or to analyze the construct of an instrument.

Adequate internal consistency is a prerequisite for scales developed to measure one particular construct. When all the items within a scale correlate highly with each other, the scale demonstrates good internal consistency, and thus measures one underlying construct. Internal consistency is calculated using Cronbach's  $\alpha$ . Values range from 0–1, with higher scores reflecting higher internal consistency. For group comparisons in research situations the internal consistency is considered to be adequate when  $\alpha$  exceeds 0.7.<sup>7</sup>

Test-retest reliability is assessed by calculating the reproducibility of an instrument in stable patients over a relatively short time period and is best calculated by means of the  $\kappa$  coefficient, or the ICC.

Responsiveness (or sensitivity to change) is the ability of an instrument to accurately detect change when it has occurred. Responsiveness in HRQoL instruments is preferably demonstrated with both internal indicators of change (correlation

with the patient's own evaluation of change) and external indicators of change (correlation with external measures).

Other information that was gathered included the procedure of item generation, type of scale, number of items, response options, scoring method, available translations, availability of instructions, conditions for use, administration time, and frequency of missing items. Whenever information on studies or scales was unclear or incomplete, we contacted the authors with the request to provide additional information.

### RESULTS

We found 21 studies addressing five scales. Five of these studies concerned translated versions. One study, and consequently one scale (PDQUALIF),<sup>8</sup> was excluded because information on the format of this scale, as well as on the included items, was unavailable at the time of our review. Therefore, 20 studies reporting on the clinimetric properties of four scales were included in this review.

These scales were the Parkinson's disease questionnaire-39 item version (PDQ-39),<sup>9</sup> the Parkinson's disease quality of life questionnaire (PDQL),<sup>10</sup> the Parkinson's impact scale (PIMS),<sup>11</sup> and the Parkinson LebensQualität (PLQ) (Parkinson QoL) questionnaire.<sup>12</sup> Some common characteristics of the scales are considered first. Details on individual scales are discussed later, followed by a comparison of the clinimetric characteristics.

#### Disease specific HRQoL scales

The four questionnaires were developed between 1995 and 1998. The scales can be self completed by the patient, but can also easily be administered by an interviewer.

All scales can be used freely for scientific purposes, but in the case of the PDQL permission for use must be granted from the developers. The PDQ-39 and the PDQL have a licence fee for commercial use. The administration time of these scales was never formally assessed, but is expected to vary from 10 minutes (PIMS) to 15 or 20 minutes (PDQ-39, PDQL, PLQ).

All scales use a five point ordinal scoring system. The number of available translations differs considerably between scales, ranging from one (PLQ) or two (PIMS), to 10 (PDQL) or 21 (PDQ-39). The number of studies in which these instruments have been used range from one (PLQ and PIMS) to at least five (PDQL), or 18 (PDQ-39). An instruction manual for scientific users is only available for the PDQ-39 and the PLQ.

#### PDQ-39

The PDQ-39 was designed by Peto *et al.*<sup>9</sup> The scale has 39 items. Higher scores reflect lower HRQoL. The PDQ-39 has eight subscales: mobility (10 items), activities of daily living (six items), emotional wellbeing (six items), stigma (four items), social support (three items), cognitions (four items), communication (three items), and bodily discomfort (three items). Items in each subscale,<sup>13</sup> as well as in the total scale,<sup>14</sup> can be summarised into an index and transformed linearly to a 0–100 scale. A shorter summary index (PDQ-8 SI) can also be calculated.<sup>15</sup>

The scale has been formally validated in United States English,<sup>16</sup> United Kingdom English,<sup>9</sup> German,<sup>17</sup> and Spanish.<sup>18,19</sup> A French version is currently being validated.<sup>20</sup> Translations are available in Australian English, Canadian English, Canadian French, Czech, Danish, Dutch, Finnish, Hebrew, Italian, Polish, Portuguese, Russian, Swedish, Greek, Japanese, and Serbian.

#### PDQL

The PDQL was developed by de Boer *et al.*<sup>10</sup> This scale has 37 items. Higher scores reflect better HRQoL. Four subscales are discerned: parkinsonian symptoms (14 items), systemic

**Table 1** Content of HqoL scales

Items	PDQ39	PDQL	PIMS	PLQ
<b>Physical:</b>				
Walking				
Problems getting around house	6			
Walking	4, 5	24		25
Shuffling; initiation/shuffling		11		7
Transfer:				
Turning in bed		35		
Getting up		32		
Turning around while walking		14		
Self care:				
Self care				23
Washing	11			
Dressing	12			
Cutting food	15			
Holding drinks without spilling	16			
Doing up buttons/shoe laces	13			
Daily activities:				
Looking after home*	2		(5)	40
Carrying bags/shopping	3			
Doing leisure activities*	1	(8)	7	(38)
Getting around in public	7			
Transportation:				
Take part in traffic		29	6	21
Needed company when going out	8			
Motor feature:				
Mobility				27
Slowness				1
Rigidity		1		3
Dexterity		9		26
Shaking hands; tremor		6		4
Sudden uncontrolled movements		30		2
On off periods		20		14
Speech; talking	34	22		
Writing	14	16		
Other disease features:				
Generally unwell		2		
Extreme exhaustion		13		
Worn out		7		20
Painful cramps or spasms	37			
Pain in joints or body	38			6
Unexpected falling asleep	30			
Problems sleeping (at night)		19		5
Difficulty sitting still		27		
Drizzling		25		
Incontinence/frequent urinating		28		
Constipation		33		
Feeling unpleasantly hot or cold	39			
Feeling that body parts don't belong to oneself				44
<b>Mental:</b>				
Cognition				
Concentration	31	31		8
Adjusting to circumstances				32
Memory	32	34		28
Mood				
Depression	17	26	2†	11, 30
Weepy or tearful	19			
Feelings (positive)				
Feelings; confidence			1	41
Self worth			1	29
Feelings (negative)				
Being tense; stress		4	2	10
Angry or bitter	20			
Unsure due physical limitations; trust body functions		5		34
Safety; doing what you want without harming yourself			8	
Unsure around others		18		
Feeling isolated/lonely	18			
Feeling ignored	36			
Anxious/worry				
Anxious	21		2	
Worry loss cognitive capacity				18
Worry physical loss				19
Worry illness				42
Worry future; afraid progression	22	15		13, 16, 43
Worry operation		37		
Fear side effects				17
Worry falling	9			

**Table 1** continued

Items	PDQ39	PDQL	PIMS	PLQ
Other features				
Dependency on medication				15
Difficulty accepting illness		21		
Confined to house more than liked	10			
Distressing dreams/hallucinations	33			9
Disinterest/listlessness				12
Social and role functioning:				
Relationships				
Problems with close relationships	27		3	
Lacked support spouse/partner	28		3	
Lacked support family/friends	29			
Impact community relationships			4	
Sexual relationship		36	10	
Wanted to isolate oneself				35
Keeping up relationships				24
Social stigma (fear of)				
Embarrassed (in public)	25	10		
Worried about other's reaction	26			
Felt had to conceal PD	23			
Feeling that illness is noticed				36
Avoid eating/drinking in public	24			
Signing name in public		23		
Role functioning				
Work			5	
Looking after home*	(2)		5	
Financial security/support			9	
Feeling dependent of others				31
Social activities:				
Doing hobbies		3		39
Doing leisure activities*	(1)	8	(7)	38
Cancel social activities		12		
Cancel important activities				33
Less able to go on holiday		17		37
Visit exhibitions				22
Problems with communication	35			

Numbers in columns correspond with the number of the item in the scale; numbers in parentheses indicate that the item is assessed, but is primarily allocated to another section.

\*"leisure activities" and "looking after home" are present in both physical and social domain because of differences in subscale allocation among scales; †item numbers of the PIMS are sometimes presented more than once because of broad formulation.

symptoms (seven items), social function (seven items), and emotional function (nine items).

The PDQL has been formally validated in Dutch,<sup>10</sup> United Kingdom English,<sup>21</sup> German,<sup>22</sup> and French.<sup>22</sup> Translations are available in Argentinian Spanish, Belgian Dutch, Italian, Portuguese, and Spanish.

### PIMS

The PIMS was developed by Calne *et al.*<sup>11</sup> The scale has 10 items and is completed three times, 1 month apart. The items in the PIMS are broadly formulated and concern domains rather than specific situations. Higher scores reflect lower HRQoL. Stable patients only score each item once, whereas patients with fluctuations judge the negative impact for both on and off periods. The scale can be self completed, but the developers recommend that patients be advised with respect to their disease state (stable or fluctuating). Guidelines for use by physicians are available.

The scale is only available in Canadian English and Canadian French. The scale contains two optional items—sexuality and financial security—which were left unanswered in 32% and 13% of the questionnaires, respectively.

### PLQ

The PLQ was designed by van den Berg.<sup>12</sup> The scale has 44 items. Items in the scale are grouped in nine domains: depression (five items), physical achievement (five items), concentration (four items), leisure (five items), restlessness (four items), activity limitation (six items), insecurity (five items), social integration (five items), and anxiety (five items). There

are five types of standard questions and four categories of responses, worded in two directions. Responses can be recoded with a spreadsheet programme that is available from the author.

The scale has been validated only in German.

### Scale development, scoring, and time frame

The method of item generation differed between scales. In the PIMS, items were decided upon by consensus between 10 specialised nurses and tested in 167 patients. In the other three scales, patients were directly involved in the generation and evaluation of items. Although the PDQ-39 and the PLQ solely relied on patient information for item generation, items in the PDQL were also obtained from interviews with neurologists, relatives of patients, and studying the literature.

Items in the PDQ-39 were initially generated through interviews with 20 patients visiting an outpatient neurology clinic.

**Table 2** Number of items/domain

	PDQ-39	PDQL	PIMS	PLQ
Physical	19	21	2	16
Mental	12	9	3	19
Social	8	7	5	9
Total number of items	39	37	10	44

Items are allocated to a particular domain on the basis of face value, or, in the case of ambiguity, on the basis of subscale allocation or factor structure, as presented by the scale developers in original study.

**Table 3** Clinimetric characteristics of HQoL scales

Scale	Reliability		Validity			
	Intern consist	Test-retest	Content	Construct	Factorial	Responsiveness
PDQ-39 $\alpha$	Total scale: 0.84–0.94 <sup>14 23 27</sup>	$r=0.68-0.94^9$ (n=167; 3–6 days)	+++ <sup>9</sup>	<p><i>Generic HRQoL scales:</i> Range subscales PDQ-39 and SF-36: <math>r=-0.34-0.80^9</math> Summary indices of PDQ-39 and EQ-5D: <math>r_s=-0.75^{28}</math></p> <p><i>Disease specific scales:</i> H and Y: <math>r_s=0.60^{28 29}</math> H and Y v PDQ subscales, range: <math>r_s=0.16-0.72^{24 27}</math> SES: <math>r_s=-0.66^{28 29}</math> UPDRS-ME: <math>r_s=0.41^{29}</math> Columbia, range subscales: <math>r_s=0.08-0.58^{24 27}</math></p> <p><i>Other measures:</i> Beck's DI with PDQ 39 SI: <math>r_s=0.68^{28}</math> and with PDQ 39 emotional: <math>r=0.73^{25}</math> Barthel index with PDQ 39 ADL: <math>r_s=0.3^{25}</math> MMSE: <math>r_s=-0.32^{28 29}</math> <i>Analysis of group differences:</i> • by self reported severity: S<sup>9</sup> By H and Y (clinic sample) : S (except social)<sup>27</sup> or S (except emotional, stigma, social, cognitions, bodily discomfort)<sup>23</sup> By H and Y (population sample): s<sup>30</sup> (except social and stigma)</p>	<p>Eight factors<sup>9</sup> Mobility ADL Emotional Stigma Social support Cognition Communication Bodily discomfort</p>	<p>SRM for mobility and ADL: 0.55 and 0.43, respectively (n=51, 4 month interval)<sup>24</sup> Change PDQ-39 compared with change in<sup>24</sup>: Columbia: NS H and Y: NS SF-36: S Self report change: S  Four subscales (mobility, ADL, stigma, social support) show significant deterioration PDQ-39 more responsive than GHQ-28 and OPCS<sup>25</sup></p>
	$\alpha$ Subscales: 0.69–0.94 <sup>9</sup> or 0.66–0.95 <sup>27</sup> or 0.57–0.94 <sup>23</sup>					
	Item-total correlation: 0.67–0.91 <sup>9</sup>					
PDQL	$\alpha$ Total scale: 0.94 <sup>10</sup> –0.95 <sup>21</sup>	0	++ <sup>10</sup>	<p><i>Generic HRQoL scales:</i> SF24 Mental health v PDQL emotional: <math>r=0.66^{10}</math> SF24, related domains: range <math>r=0.46-0.64^{10}</math></p> <p><i>Disease specific measures:</i> "Webster contributes significant to QoL"<sup>21</sup></p> <p><i>Other measures:</i> CES D v PDQL emotional: <math>r=-0.79^{10}</math> MOS soc support survey v PDQL-social: <math>r=0.13^{10}</math> "CAMCOG/GDS 15 contribute significant to QoL"<sup>21</sup> <i>Analysis of group differences:</i> By SES, 3 levels: S<sup>10</sup> By Webster, 3 levels: S (all, except emotional)<sup>21</sup></p>	<p>Four factors<sup>10</sup>: Parkinsonian Systemic Social Emotional</p>	0
	$\alpha$ Subscales: 0.80–0.87 <sup>10</sup> or 0.77–0.87 <sup>21</sup>					
PLQ	$\alpha$ Total scale: 0.95 <sup>12</sup>	Total scale: $r=0.87^{12}$	++ <sup>12</sup>	<p><i>Generic HRQoL scales:</i> EORTC QLQ30: <math>r=0.67</math> (n=111)<sup>12</sup></p> <p><i>Disease specific measures:</i> H and Y: <math>r=0.27</math>, NS (n=21–29)<sup>12</sup> SES: <math>r=-0.27</math>, NS (n=21–29)<sup>12</sup></p> <p><i>Other measures:</i> QoL VAS: <math>r=-0.28</math>, NS (n=21–29)<sup>12</sup> ADL scale: <math>r=0.73</math> (n=111)<sup>12</sup></p>	<p>9 Subscales, 1 or 2 factors/subscale, &gt;50% variance<sup>12</sup> Depression Physical achievement Leisure Concentration Social integration Insecurity Restlessness Activity limitation Anxiety</p>	<p>2 Week interval No external criterion  (n=16)<sup>12</sup></p>
	$\alpha$ Subscales: 0.62–0.87 <sup>12</sup>	Subscales: $r=0.69-0.86^{12}$ (n=65; 14 days)				
	Correlation subscale–total scale: $r=0.73-0.86^{12}$ (n=405)					
PIMS	$\alpha$ Total scale: 0.90 <sup>11</sup>	ICC=0.72 <sup>11</sup> (n=149; 1 month)	– <sup>11</sup> (Consensus)	<p><i>Analysis of group differences:</i> By self assessed fluctuations: S (only between stable patients and fluctuating patients at their worst)<sup>11</sup></p>	<p>4 Factors, 72% variance<sup>11</sup>: Psychological Social Physical Financial</p>	0
	Correlation among factors: $r=0.10-0.46^{11}$					

n, Number of patients;  $\alpha$ , Cronbach's  $\alpha$ ; r, Pearson;  $r_s$ , Spearman; ICC, intraclass correlation coefficient; +++, good; ++, adequate; fair; –, poor; 0, no numerical results reported;  $\emptyset$ , unclear; SES, Schwab and England scale; Beck's DI, Beck's depression inventory; S, significant; NS, not significant; SRM, standardised response mean.

This resulted in a 65 item list that was reduced to 39 items on the basis of a survey in 359 patients. Items in the PDQL were generated by means of interviews with five patients and a relative, consulting neurologists, and reviewing the literature. Seventy three items were found and piloted in 13 inpatients and outpatients. Items endorsed most often or rated as most important were selected for the final 37 item version, which was tested in 384 patients. Items in the PLQ were generated by interviews with groups of patients. This resulted in 113 items that subsequently were piloted in 61 inpatients and outpatients. The questionnaire was then reduced to 44 items and tested in 405 patients (constituting a response rate of only 38%).

The PDQ-39 and the PDQL both assess the frequency with which patients experience difficulties. The PIMS assesses the impact of the disease on patient's lives, whereas the PLQ, depending on the item of interest, assesses intensity, applicability, or quality.

Scales differ considerably in the period they refer to. The PIMS does not specify a time frame, whereas the PDQL assesses the past 3 months, the PDQ-39 the past month, and the PLQ the past week. The PDQL and PLQ assess the items "as is", without asking the patient to indicate whether this was due to Parkinson's disease, whereas the PDQ-39 and the PIMS relate the items to having Parkinson's disease. In the PDQ-39 all items begin with: "Due to having Parkinson's disease, how much of the time did you have trouble with . . .". In the PIMS patients are asked to rate the negative impact of Parkinson's disease in a particular domain.

#### Content validity

The content of the scales differs considerably. We grouped items thematically on the basis of face value in domains reflecting physical, mental, and social or role functioning (table 1). Whenever there was doubt regarding the correct allocation, items were assigned to domains according to subscale allocation or factor structure as reported in the original studies. Table 2 shows that about half of the items in the PDQ-39 and the PDQL concern physical features, whereas the PIMS has only two items in this domain. In the PIMS, on the contrary, half of the items deal with the social domain. In the PLQ almost half of the items represent mental features.

In the physical domain only transportation is addressed by all scales. In the PIMS, the only other theme addressed in this domain concerns taking part in traffic. The PDQ-39, PDQL, and PLQ share items on walking, motor features, and other disease features. Transfers are addressed in detail in the PDQL, but are lacking in the PDQ-39 and PLQ. Items on self care are assessed in detail in the PDQ-39, and as an overall item in the PLQ, but are lacking in the PDQL. Many physical items in the PDQ-39 concern activities ("disabilities"), whereas in the PDQL and PLQ most items reflect impairments.

In the mental domain all scales include items on mood, feelings, and anxiousness. The PIMS does not incorporate items on cognition, whereas the other scales address both concentration and memory. The PLQ contains seven items addressing anxiousness.

In the social domain all scales address some aspect of relationships. Relationships with partner, family, or friends are only addressed in the PDQ-39 and the PIMS. Sexuality is only addressed in the PDQL and the PIMS. Social stigma is not assessed in the PIMS. Role functioning is adequately assessed in the PIMS, but only marginally in the PDQ-39 and the PLQ, whereas the PDQL does not address this theme at all.

#### Construct validity

Construct validity of both the PDQ-39 and the PDQL was thoroughly established using generic HRQoL scales, disease specific instruments, "known groups" comparisons, and other health measures. The PLQ was less thoroughly assessed. Correlations with a generic HRQoL scale and an ADL scale were

adequate, but correlations with disease specific instruments were poor, and known groups differences were not assessed. For the PIMS only known groups comparisons were performed, demonstrating significant differences between stable and fluctuating patients in their off situations (table 3).

In the PDQ-39 and the PDQL, subscales were constructed on the basis of an exploratory factor analysis. In the PLQ, the subscales were decided on before hand, and a confirmatory factor analysis was performed afterwards. The PIMS does not distinguish subscales.

The scales share factors on the physical, social, and psychological-emotional level, but are heterogeneous in other respects. The PDQ-39 has three factors that do not emerge as separate factors in other scales—that is, cognitions, communications, and stigma. The PDQL has a distinct factor addressing systemic symptoms. The PIMS has a separate "financial" factor, and the PLQ has unique factors on leisure, insecurity, restlessness, concentration, and anxiety.

#### Internal consistency

Cronbach's  $\alpha$ s for scale totals are all well over 0.8 (table 3). For subscales the  $\alpha$ s are higher than 0.7, except for social support in the United Kingdom version of the PDQ-39,<sup>9</sup> social support<sup>16,23</sup> and cognitions<sup>23</sup> in the United States version of the PDQ-39, for cognitions and bodily discomfort in the Spanish version of the PDQ-39,<sup>19</sup> and for mood, concentration, restlessness, and social integration in the PLQ.<sup>12</sup>

#### Test-retest reliability

Test-retest reliability was not assessed for the PDQL. In the PIMS an ICC of 0.72 was reported for the total score. Reproducibility of subscales was assessed for the PDQ-39 and the PLQ. Subscales with correlations lower than 0.7 concerned the social support subscale in the PDQ-39, and the anxiety subscale in the PLQ.

#### Responsiveness

Responsiveness was not established for either the PDQL or the PIMS. In the PLQ it was assessed only in a small subset of 16 patients during a period in hospital. Paired *t* tests were only significant for activity limitation and insecurity. When the tests were corrected for multiple comparisons, all nine scale changes were non-significant. Two studies reported on the responsiveness of the PDQ-39. Fitzpatrick *et al*<sup>24</sup> found moderate standardised response means for the mobility and ADL subscales in 51 patients who indicated that their situation had worsened over a period of 4 months. Change in the PDQ-39 score was significantly correlated with self reported change and change in the SF-36. In the other study, Harrison *et al*<sup>25</sup> found that four subscales of the PDQ-39 (mobility, ADL, stigma, social support) were responsive to deterioration in health state.

## DISCUSSION

Scales differed considerably in content. Probably this is largely the result of differences in the ways the items were generated and reduced, and differences among the samples involved in generating and evaluating the items may have added to the diversity. In the PDQ-39 and the PLQ items were only derived from interviews with patients, whereas in the PDQL information also from neurologists, relatives, and the literature was used. Items in the PIMS were obtained through consensus between specialised nurses.

To guarantee good content validity patients should be closely involved in both item generation and evaluation. For item generation other sources may be used as well. For the evaluation of items, however, a large sample of patients should be involved. This sample should ideally consist of patients attending a neurology clinic, patients living in nursing homes, and of unselected patients living in the community. None of

**Table 4** Quality assessment table

Scale	Reliability		Validity		
	Internal consistency	Test-retest	Content	Construct	Responsiveness
PDQ-39	+++ /+++	+++ /+++	++ /+++	+++ /+++	++ /+
PDQL	+++ /+++	0	++ /++	+++ /+++	0
PIMS	+++ /+++	+++ /+++	? /-	? /-	0
PLQ	+++ /+++	+++ /+++	++ /+++	+ /++	- /-

+++ /+++ : signs before the slash refer to *results* of validity, reliability, and responsiveness testing and signs behind the slash refer to *thoroughness* (strength of evidence) of validity, reliability, and responsiveness testing.

Results of validity, reliability, and responsiveness testing: 0 no numerical results reported, ? results not interpretable, - poor results, + fair results, ++ moderate results, +++ good results.

Thoroughness of validity, reliability, and responsiveness testing: 0 no reported evidence, ? results not interpretable, - poor evidence, + fair evidence, ++ moderate evidence, +++ good evidence.

the scales applied this method. The information on relevance of items in the item reduction phase was obtained from patients who were members of a Parkinson's disease society (PDQ-39), or from both inpatients and outpatients of a neurology clinic (PDQL, PLQ). In the PDQL only four outpatients and one patient and a relative from the Parkinson's disease society were involved in the item generation process, whereas only 13 inpatients and outpatients were involved in the evaluation process. Both the small sample sizes as well as the fact that only a clinic based sample was involved, may have affected the final make-up considerably. For instance, the item "feeling worried about a possible operation" in the PDQL was not found in other scales.

Different strategies with respect to the item reduction process—that is, psychometric or clinimetric—affected the final content as well. In the first strategy, considerations of the measurement properties of scales prevail, whereas in the second the completeness of the assessment is considered more important. In the PDQ-39 and the PLQ, the developers used a predominantly psychometric strategy. In the PDQ-39, items were omitted when they were considered redundant, had low item scale correlations, or clustered in subscales that could not be meaningfully interpreted. In the PLQ, items with low item scale correlations, non-normal frequency distributions, often missing values, floor or ceiling effects, or items that could not clearly be assigned to subscales, were removed. The developers of the PDQL however, followed a more clinimetric strategy and included all items patients considered important in the final scale. Items that loaded on more than one subscale were assigned to subscales on the basis of face validity.

When the scales are compared in more detail, the differences in content become apparent. The PDQ-39 lacks items addressing transfers and night time sleep problems in the physical domain, but covers all relevant themes in the mental domain. Role functioning is insufficiently covered. Sexuality is not addressed in this scale.

The PDQL misses items on self care in the physical section, taps all items in the mental domain, and lacks items on close relationships and role functioning in the social domain.

Our findings on the content validity of the PDQ-39 and the PDQL largely agree with Damiano *et al.*<sup>26</sup> However, their criteria list did not contain items explicitly addressing transfers and hobbies.

The PIMS lacks items on walking, transfers, self care, motor features, and other disease features in the physical domain, on cognition and other "features" in the mental domain, and on social stigma in the social domain.

The PLQ lacks items on transfers and communication in the physical domain, but in the mental domain all relevant themes are addressed. Role functioning and relationships are insufficiently covered in the social domain. The PLQ is the only scale that explicitly asks for the consequences of being dependent of others.

The construct validity of both the PDQ-39 and the PDQL are well established. For the PLQ this was less thoroughly demonstrated, whereas for the PIMS construct validation with other measures was not performed.

All scales share factors on physical, mental, and social functions. The other factors that emerged in the scales were very different.

The internal consistency for scale totals is adequate for all scales. All subscales of the PDQL show good internal consistency, whereas the social support subscale in the PDQ-39 and four subscales in the PLQ showed insufficient internal consistency. Test-retest reliability was not assessed for the PDQL and was found to be adequate for the other scales, except for the anxiety subscale in the PLQ and, again, the social support subscale in the PDQ-39.

Responsiveness was not assessed at all for either the PDQL or the PIMS. For the PLQ responsiveness was inadequately evaluated. There are indications that the PDQ-39 is capable of detecting deterioration, but for improvement this still needs to be established.

A comparison of the clinimetric qualities of the scales is presented in table 4.

Apart from methodological considerations, other issues may influence the selection of an HRQoL instrument. For instance, the time frame is of importance. When short periods are assessed (for example, 1 week in the PLQ), day to day differences may affect the total score considerably, resulting in lower comparability over time. Assessing longer periods may therefore be preferred, as is done in both the PDQ-39 (1 month) and the PDQL (3 months).

Another factor that may affect the selection of a scale is the framing of questions. The PDQL and the PLQ evaluate health "as is", regardless whether complaints were caused by Parkinson's disease or not. Both other scales relate the health state to having PD. However, it may be difficult or even impossible for patients to judge whether a particular situation (for example, sleep problems, fatigue) is caused by PD, or is the result of aging or some comorbid condition.

Other considerations that may guide the selection of an appropriate HRQoL instrument for a particular study may concern the language and the number of studies in which the instrument has been used. In this respect the PDQL, and especially the PDQ-39, are attractive candidates.

The intended sample may also influence the selection. The PLQ was tested only in a sample of patients that were members of a Parkinson's disease society, whereas the PIMS was used only in an outpatient clinic sample. The PDQL was evaluated both in a Parkinson's disease society sample and a community based sample, whereas the PDQ-39 was evaluated in all the aforementioned populations.

The number of items is not a useful criterion for selection, because the numbers hardly differ between the PDQ-39, the PDQL, and the PLQ, whereas insufficient clinimetric support exists for the only short scale, the PIMS.

In most other respects the scales differed only marginally, and therefore these factors are not expected to play a part in selecting a scale.

The selection of an instrument will partly be based on the goal of the study. For certain interventions, some domains of HRQoL are more important to assess than others, and may thus influence the selection of the instrument. In many situations, however, the PDQ-39 will probably be the most appropriate HRQoL instrument, because this scale has been tested most thoroughly, has adequate clinimetric characteristics, has been used in the largest number of studies, and is available in many languages. However, responsiveness of this scale still needs to be assessed more thoroughly, especially with respect to situations in which patients are expected to improve (for example, intervention studies). The PDQ-39 lacks items on self image, night time sleep problems, sexual activity, and transfers. Reliability of the social support subscale (test-retest and internal consistency) is inadequate. The PDQL may be considered as an alternative. Information on test-retest reliability and responsiveness however, is still lacking and the scale does not include items on self care, role functions, and close relationships. The PLQ may be considered in studies involving German speaking patients with Parkinson's disease. However, construct validity and responsiveness are insufficiently assessed, and items concerning transfers and speech are missing, whereas relationships and role functions are only scarcely addressed. Use of the PIMS should be considered only as a means of identifying areas of potential problems. Items in this scale lack specificity, whereas the content validity is insufficiently founded and construct validity and responsiveness are not assessed at all.

## ACKNOWLEDGEMENTS

JM is supported by the Netherlands' National Research Council (project No 0940-33-021) and CR is supported by the Prinses Beatrix Fund (project No 97-0205). We thank C Berne, S Calne, AGEM de Boer, JE Harrison, V Peto, M van den Berg, and MD Welsh for providing additional information.

.....

## Authors' affiliations

**J Marinus, C Ramaker, J J van Hilten**, Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands  
**A M Stiggelbout**, Department of Medical Decision Making, Leiden University Medical Center, Leiden, The Netherlands

## REFERENCES

- Bowling A**. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Soc Sci Med* 1995;41:1447-62.
- World Health Organization**. *WHO Constitution: the first two chapters*. Electronic citation. World Wide Web 2001: <http://www.ldb.vl/top/whoconst.htm>.
- Bowling A**. Health-related quality of life: a discussion of the concept, its use and measurement. In: *Measuring disease*. Buckingham: Open University Press 1995:1-19.
- Guyatt GH**, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *Canadian Medical Association Journal* 1986;134:889-95.
- McDowell I**, Newell C. *Measuring health: a guide to rating scales and questionnaires*. New York: Oxford University Press, 1996.
- Kirshner B**, Guyatt G. A methodological framework for assessing health indices. *Journal of Chronic Diseases* 1985;38:27-36.
- Bland JM**, Altman DG. Cronbach's  $\alpha$ . *BMJ* 1997;314:572.
- Welsh M**, McDermott M, Holloway R, et al. Development and testing of the Parkinson's disease quality of life scale: the PDQUALIF. *Mov Disord* 1997;12:836.
- Peto V**, Jenkinson C, Fitzpatrick R, et al. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Qual Life Res* 1995;4:241-8.
- de Boer AG**, Wijker W, Speelman JD, et al. Quality of life in patients with Parkinson's disease: development of a questionnaire. *J Neurol Neurosurg Psychiatry* 1996;61:70-4.
- Calne S**, Schulzer M, Mak E, et al. Validating a quality of life rating scale for idiopathic parkinsonism: Parkinson's impact scale. *Parkinsonism and Related Disorders* 1996;2:55-61.
- Van den Berg M**. Leben mit Parkinson: Entwicklung und psychometrische Testung des Fragebogens PLQ. *Neurol Rehabil* 1998;4:221-6.
- Peto V**, Jenkinson C, Fitzpatrick R. PDQ-39: a review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. *J Neurol* 1998;245 Suppl 1:S10-14.
- Jenkinson C**, Fitzpatrick R, Peto V, et al. The Parkinson's disease questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. *Age Ageing* 1997;26:353-7.
- Jenkinson C**, Fitzpatrick R, Peto V, et al. The PDQ-8: development and validation of a short-form Parkinson's disease questionnaire. *Psychology and Health* 1997;12:805-14.
- Bushnell DM**, Martin ML. Quality of life and parkinson's disease: translation and validation of the US Parkinson's disease questionnaire (PDQ-39). *Qual Life Res* 1999;8:345-50.
- Berger K**, Broll S, Winkelmann J, et al. Untersuchung zur Reliabilität der deutschen Version des PDQ-39: Ein krankheitsspezifischer Fragebogen zur Erfassung der Lebensqualität von Parkinson-Patienten. *Aktuel Neurol* 1999;26:180-4.
- Martinez-Martin P**, Frades PB. Quality of life in Parkinson's disease: validation study of the PDQ-39 Spanish version. The Grupo Centro for Study of Movement Disorders. *J Neurol* 1998;245 Suppl 1:S34-38.
- Martinez MP**, Frades B, Jimenez Jimenez FJ, et al. The PDQ-39 Spanish version: reliability and correlation with the short-form health survey (SF-36). *Neurologia* 1999;14:159-63.
- Mauduit N**, Schuck S, Allain H, et al. Échelles et questionnaires dans la maladie de Parkinson. *Rev Neurol (Paris)* 2000;156(suppl 2):63-9.
- Hobson P**, Holden A, Meara J. Measuring the impact of Parkinson's disease with the Parkinson's disease quality of life questionnaire. *Age Ageing* 1999;28:341-6.
- Marquis P**, Girod I, Berdeaux G, et al. Psychometric analysis of French and German versions of the Parkinson's disease quality of life questionnaire (PDQL). *Qual Life Res* 1998;7:632.
- Damiano AM**, McGrath MM, Willian MK, et al. Evaluation of a measurement strategy for Parkinson's disease: assessing patient health-related quality of life. *Qual Life Res* 2000;9:87-100.
- Fitzpatrick R**, Peto V, Jenkinson C, et al. Health-related quality of life in Parkinson's disease: a study of outpatient clinic attenders. *Mov Disord* 1997;12:916-22.
- Harrison JE**, Preston S, Blunt SB. Measuring symptom change in patients with Parkinson's disease. *Age Ageing* 2000;29:41-5.
- Damiano AM**, Snyder C, Strausser B, et al. A review of health-related quality-of-life concepts and measures for Parkinson's disease. *Qual Life Res* 1999;8:235-43.
- Jenkinson C**, Peto V, Fitzpatrick R, et al. Self-reported functioning and well-being in patients with Parkinson's disease: comparison of the short-form health survey (SF-36) and the Parkinson's disease questionnaire (PDQ-39). *Age Ageing* 1995;24:505-9.
- Schrag A**, Selai C, Jahanshahi M, et al. The EQ-5D—a generic quality of life measure—is a useful instrument to measure quality of life in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2000;69:67-73.
- Schrag A**, Jahanshahi M, Quinn N. What contributes to quality of life in patients with Parkinson's disease? *J Neurol Neurosurg Psychiatry* 2000;69:308-12.
- Schrag A**, Jahanshahi M, Quinn N. How does Parkinson's disease affect quality of life? A comparison with quality of life in the general population. *Mov Disord* 2000;15:1112-8.