B. del Pozo-Cruz
D. Collado-Mateo
J.C. Adsuar
M.A. Garcia-Gordillo
J.M. Cordero-Ferrera
F.I. Sánchez-Martínez
J.M. Abellán-Perpiñán

A Psychometric Comparison of Different Health Utility Measures in Patients Affected by Parkinson's Disease

Executive Summary

- Parkinson's disease is a chronic disease affecting quality of life and well-being.
- Cost utility analyses is a method of determining the cost effectiveness of nursing interventions. These analyses are performed using data from preference-based questionnaires.
- Several options are available but their psychometric properties should be compared to optimize choice.
- The main purpose of this study was to evaluate the validity, sensitivity, and relative efficiency of 15D and SF-6D questionnaires in Spanish patients with Parkinson's to be used in cost-effectiveness analyses.
- Findings indicated SF-6D and 15D are adequate instruments for monitoring of patient's health during the period of rehabilitation.

In recent years there has been an increasing interest in using economic evaluation, particularly cost-effectiveness analysis, as a relevant argument to decide which new healthcare technologies should be publicly financed. A critical element in evaluating health interventions is the measurement and valuation of health outcomes. Clinical measures were used extensively in the past, but now there is a consensus regarding the importance of taking into account more comprehensive measures. In chronic diseases, especially those affecting patients' quality of life, the subjective perception of their own disease and well-being is crucial and need to be considered both by nursing professionals and by patient associations or family, since sometimes professional opinions differ from the subjective perception of patients (Murray & Lopez, 1997; Pickard & Knight, 2005). This perception, health-related quality of life (HRQoL) (Guyatt, Feeny, & Patrick, 1993), is measured through questionnaires. HRQoL shows how a healthcare condition or intervention affects a patient's health by

B. DEL POZO-CRUZ, PhD, is Professor, Department of Public Health, University of Auckland, Auckland, New Zealand.

D. COLLADO-MATEO, MSc, is PhD Student, Department of Exercise, Quality of Life and Health Economics, University of Extremadura, Caceres, Spain.

J.C. ADSUAR, PhD, is Senior Lecturer, Department of Exercise, Quality of Life and Health Economics, University of Extremadura. Caceres, Spain.

M.A. GARCIA-GORDILLO, PhD, is Facultad de Administración y Negocias, Universidad Autonoma de Chile, sede Talca, Chile.

J.M. CORDERO-FERRERA, PhD, is Professor, Department of Economics, University of Extremadura, Badajoz, Spain.

F.I. SÁNCHEZ-MARTÍNEZ, PhD, is Professor, Department of Applied Economics, University of Murcia, Murcia, Spain.

J.M. ABELLÁN-PERPIÑÁN, PhD, is Professor, Department of Applied Economics, University of Murcia, Murcia, Spain.

assessing if the patient's condition and symptoms are affecting his or her physical or mental abilities.

To have clinical usefulness, HRQoL measures must offer validity, reliability, sensitivity to change, and provide useful clinical data among other properties. Then, by comparing different measures with disease-specific measures, the most appropriate measure can be obtained for monitoring changes in a patient's HRQoL due to an intervention (Aggarwal et al., 2009).

Parkinson's disease (PD) is a chronic neurodegenerative movement disorder. Its prevalence rises with age and varies slightly depending on geographic location and gender (Pringsheim, Jette, Frolkis, & Steeves, 2014). As a consequence of both motor and non-motor symptoms (Chaudhuri, Odin, Antonini, & Martinez-Martin, 2011; Gazewood, Richards, & Clebak, 2013), HRQoL is significantly reduced in patients affected by PD. The etiology of PD remains unknown and although there are medications for reducing the progression, in long term their effectiveness may be reduced. In fact, several non-motor symptoms like dementia or sexual dysfunction could indeed be caused by medication treatment (Chaudhuri et al., 2011). Therefore, the main treatment objective in these patients is to control the symptoms to increase HRQoL. Therefore, healthcare decisions in PD should be focused on patients' HRQoL.

Specific questionnaires are more suitable for patients affected by PD because they can detect relevant clinical changes better than other generic HRQoL instruments. Unlike these profile-based instruments, generic preference-based questionnaires (SF-6D, EQ-5D, 15-D, HUI, etc.) add a valuation or preference component to the descriptive element (Brazier, Deverill, & Green, 1999). These instruments are suitable to calculate *quality-adjusted life years* (QALYs), which have become a common measure of treatment effects on health state (health gains) (Maklin et al., 2012). A QALY is a single measure that comprises mortality and morbidity; it combines the effect on survival in years and HRQoL experienced in those years (Soares, 2012). To calculate QALYs, the number of years in a specific health state is multiplied by a health factor or quality weight, which ranges from 0 (death) to 1 (full health). Cost-effectiveness studies which use QALYs as the unit for measuring health outcomes are called *cost-utility analyses* (CUA). Using a common and comprehensive measure for health gains allows comparison of cost-effectiveness ratios between different diseases and/or health interventions (Hanmer et al., 2015; Tosh, Longworth, & George, 2011).

To our knowledge, there are no previous studies comparing 15D (Sintonen, 2001) and SF-6D (Brazier, Roberts, & Deverill, 2002) in PD population. Therefore, the main aim is to evaluate the validity, sensitivity, and relative efficiency of both instruments. The result will be useful to decide which HRQoL questionnaires are the most adequate in Spanish patients affected by PD, depending on level of progression of the disease (stages I-II or III-IV of the Hoehn & Yahn scale [H&YS]).

Methods

Study design. A cross-sectional study design was performed. Data were obtained between May 1, 2012 and August 1, 2013. This study was based on a non-representative sample of different local PD associations. The sample consisted of 231 patients affected by PD living in Spain.

Patient recruitment. Patients were recruited from different local PD associations. There are 44 PD associations in Spain. From those, 23 PD associations participated in the present study, belonging to 13 of 17 different regions of Spain. Patients were included in the study if they followed the following inclusion criteria: (a) greater than 18 years of age, (b) able to answer questions by themselves, and (c) diagnosed with PD. Patients were classified using the H&YS (Hoehn & Yahr, 1967), a simple descriptive scale with five different stages of Parkinsonian motor impairment (Goetz et al., 2004). The questionnaires included in the study were (a) two generic HRQoL questionnaires: SF-36 and 15D; (b) one specific HRQoL questionnaire for PD population: PDQ-8; and (c) the H&YS. Respondents were also asked about their medical and therapeutic conditions and sociodemographic characteristics. The initial database included many other variables, such as PDQ-39, EQ-5D, and its extended version EQ-5D-5L. However, given the difficulty of presenting a global comparison of all questionnaires, these comparisons have been presented separately, avoiding the potential readers could have difficulty understanding clearly the results in one article (Garcia-Gordillo, del Pozo-Cruz, Adsuar, Sanchez-Martinez, & Abellan-Perpinan, 2014; Garcia-Gordillo et al., 2015).

Utility Measures

Short form health survey questionnaire 6 dimensions (SF-6D). The SF-6D is a preference-based single index measure of HRQoL. This tool allows researchers to calculate preference-based utilities using data from SF-36 or SF-12 questionnaires (Brazier et al., 2002; Brazier & Roberts, 2004). For this study, SF-6D utilities were derived from the SF-36 questionnaire (Brazier et al., 2002). The descriptive part of the instrument consists of six dimensions: physical functioning, role limitations, social functioning, pain, mental health, and vitality. Each dimension has four, five, or six levels. These levels indicate the degree of impairment on each dimension. The combination of levels and dimensions defines 18,000 different health states. The utility scoring algorithm for the Spanish population was estimated by Abellán-Perpiñán, Sanchez Martinez, Martinez Perez, and Mendez (2012) using a lottery-

equivalent method. This method is different from the one used in other studies, the standard gamble, which has typically been regarded as the "gold standard" in measuring health state utilities (Pickard, Wang, Walton, & Lee, 2005). Lottery-equivalent procedures reduce the bias known as "certainty effect," which means respondents tend to overvalue the riskless option against the gamble (Cohen & Jaffray, 1988). This effect is reduced because lottery-equivalent method proposes two risky alternatives (Abellan Perpinan et al., 2012).

The 15 dimensions questionnaire (15D). The 15D (Sintonen, 2001) is a generic and standardized HRQoL questionnaire that can be used both as a profile-based measure and preference-based instrument. It consists of 15 dimensions that describe health status: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity. Answers from 1-5 are available for each dimension, greater scores representing worse conditions. Each health state is defined by the combination of levels 1-5 and dimensions, and can be represented by a single utility score, which ranges from 0 (death) to 1 (full health). To our knowledge, there is only one set of utilities for 15D, which is based on the preferences of a Finnish population sample (Sintonen, 2001). It has been validated with patients affected by PD (Haapaniemi, Sotaniemi, Sintonen, & Taimela, 2004).

Parkinson's disease questionnaire (PDQ-8). The short version of PDQ-39 is the PDQ-8 (Jenkinson, Fitzpatrick, Peto, Greenhall, & Hyman, 1997). It is a profile-based questionnaire that includes eight dimensions, each of them represented by one single item. The dimensions are mobility, activities of daily living, emotional well-being, social support, cognition, communications, bodily discomfort, and stigma (Jenkinson & Fitzpatrick, 2007). Each dimension score goes from 0 (never) to 4 (always or cannot do at all). A single index score can be calculated and it ranges from 0-100, where the higher score represents worse HRQoL (Jenkinson et al., 1997). No evidence of floor or ceiling effect has been found for this questionnaire (Brazier et al., 2002; Martinez-Martin et al., 2011). The Spanish adaptation of PDQ-8 has been used in different studies (Jenkinson & Fitzpatrick, 2007; Martinez-Martin et al., 2005).

Hoehn & Yahr scale (H&YS). The H&YS was designed to be a simple scale that provides an overall measure of Parkinsonian motor impairment. The effects of PD on motor capability can be unilateral or bilateral and may affect body balance. Based on these effects, patients affected by PD can be classified from unilateral (stage 1) to bilateral symptoms (stage 2). Bilateral involvement can be without impairment of balance (stage 2) or with some postural instability (stage 3). On stage 4, patients are still able to walk or stand unassisted but the disability is severe. Patients need a wheelchair or must be bed-bound on stage 5. Different studies have found a correlation between progression in H&YS and dopaminergic loss, other standardized scales of motor disorders, disability, and quality of life (Goetz et al., 2004).

Data collection process. All participants received detailed information about the aims and were included in the study once they gave their written informed consent. Interviews were performed at the local PD association by a trained interviewer. Thirty minutes was the duration on average of the interviews. The study was approved by the ethics committee of the University of Extremadura and was developed following the ethical guidelines of the Declaration of Helsinki as revised in 2000.

Statistical analyses. Statistical analyses were performed using Statistical Package for Social Sciences, version 18.0 (SPSS Inc., Chicago, IL). The significance level was set at $p<0.05$ in all cases. Since data did not follow a normal distribution, nonparametric statistics were used for all analyses.

Descriptive statistics were computed to characterize the sample and distribution of SF-6D, 15D, and PDQ-8. Different descriptive statistics were obtained depending on the type of variable: mean, standard deviation, median, inter-quartile range, and range were calculated for continuous variables, while the number and proportion in the sample were computed for categorical variables. Patients were classified in stages from 1-4 of the H&YS. Patients in stages 1 or 2 were clustered in the "mild to moderate" severity group, while patients in stages 3 or 4 were grouped in the "severe" group. Mann-Whitney U or chi-square tests were used to compare subjects with different levels of PD severity.

Construct validation. The convergent validity of the HRQoL generic questionnaires SF-6D and 15D was evaluated by examining their association with the specific questionnaire, PDQ-8. Spearman's rank correlation coefficient (r) was calculated as the validity coefficient to assess these associations. Correlation strength was established as figures in absolute values (Cohen, 1988): strong correlation ($r\geq0.5$), moderate correlation (0.3-0.5), and weak correlation (0.2-0.3).

Another analysis was computed to further extend testing validity. A "known-groups" scheme was used to assess the discriminative validity of the two generic HRQoL questionnaires (Hattie & Cooksey, 1984). This analysis evaluated the ability of SF-6D and 15D to discriminate patients with different levels of PD severity and self-reported health status groups, in conjunction with other sociodemographic and medical variables, namely social economics status, duration of PD, ongoing therapies, and presence/absence of other

medical conditions different than PD. The effects of the dichotomous variables (including levels of PD severity) on utility scores of both questionnaires were assessed by computing Mann-Whitney U tests.

Level of agreement between SF-6D and 15D. The intra-class correlation coefficient (ICC) and Bland-Altman plot were computed to test the level of agreement between SF-6D and 15D utility score. An ICC value equal or greater than 0.7 was considered acceptable (Fayers & Machin, 2016). The x-axis of the Bland-Altman plot represents the mean of the utility score, while the y-axis represents the difference between both questionnaires, where the SF-6D was the subtrahend. A difference value close to zero suggests a strong agreement, while deviation from this value represents degree of disagreement between both questionnaires for each subject (Bland & Altman, 1986). Paired comparisons using Wilcoxon's signed rank test were performed to compare SF-6D and 15D utility scores across the sample as well as for subgroups based on sociodemographic and clinical variables. Association between both questionnaires in those conditions was also calculated with Spearman's rank correlation.

Efficiency and sensitivity of SF-6D and 15D. The relative efficiency (RE) coefficient was computed to evaluate and compare the efficiency of the SF-6D and 15D, and to identify clinically relevant differences between patients affected by PD. RE can be defined as the ratio of the square of the *t*-statistic of the comparator instrument over the square of the *t*-statistic of the reference instrument (Fayers & Machin, 2016; Gaujoux-Viala et al., 2011; Petrou & Hockley, 2005).

$$RE = \frac{(T - \text{statistic of the comparator instrument})^2}{(T - \text{statistic of the reference instrument})^2}$$

In the present study, the comparator instrument was SF-6D utility score and the reference was 15D utility score. A coefficient greater than 1 suggests SF-6D is more efficient than 15D at detecting clinically relevant differences in the conditions under analysis, while a coefficient lower than 1 means the reference instrument (15D in this study) is more efficient than the comparator instrument (SF-6D).

Additionally, the receiver operating characteristics (ROC) curves (Stucki, Liang, Fossel, & Katz, 1995) were computed to compare the sensitivity of both 15D and SF-6D questionnaires. ROC method requires an external criterion or indicator. In this study, HRQoL assessed by PDQ-8 and severity of symptoms using the H&YS were these external indicators. The area under the ROC curve is the overall probability the instrument describes patients accurately. This area under the curve (AUC) is calculated in logistic regression models using a dichotomous variable as dependent variable and instrument score as independent variable. The AUC score ranges from 0-1, with 1 the perfect discrimination, while 0.5 or less indicates the measure has no discriminatory power. Therefore, the most sensitive utility measure at detecting changes in the external indicator would obtain an AUC score closer to 1 than the less sensitive instrument. Since ROC method needs dichotomized variables, "cut-off" points for PDQ-8 and H&YS had to be established. Based on the literature (Luo et al., 2009), 5.8 and 7.4 were selected as cut-off points for PDQ-8; however, the cut-off selected is better used as a cut-off value for changed scores derived in longitudinal studies. Another cut-off point of 25.0 was selected based on the PDQ-8 median. Severity of symptoms variable was dichotomized differentiating between patients in stages I-II and those in stages III-IV.

Results

Participant characteristics. The mean age of the sample was 66.00 (±9.32) years old (range 34-86 years). A total 148 patients were classified within stages I-II (H&YS), while the rest (81 patients) were classified within stages III-IV.

Descriptive statistics of SF-6D and 15D. Main descriptive characteristics of the sample are shown in Table 1, which also contains information about the distribution of scores in the different questionnaires. Two groups were created based on H&YS: the first group including those patients in stages I-II and the second group clustering those in stages III-IV. Both groups were compared using Mann-Whitney U or chi-square tests. Statistically significant differences ($p<0.05$) between both groups were found for three sociodemographic variables: (a) mean age of patients was higher in the second group (stages III-IV); (b) patients with a household income greater than €1,800 belong more frequently to the first group (stages I-II); (c) mean years since clinical diagnosis were greater in patients of stages III-IV. The mean PDQ-8 total score was statistically higher ($p<0.001$) in patients of stages III-IV in comparison with those in stages I-II. Statistically significant differences ($p<0.001$) were found in the utility scores of 15D and SF-6D, where patients in stages III-IV obtained a score lower than those in stages I-II.

Distribution of patients in the different levels of 15D and SF-6D can be seen in Table 2. Only one subject obtained a score in the 15D and SF-6D utility indexes equal to 1 (best possible health state), while no respondent obtained the lowest score in any of the instruments. Therefore, neither 15D utility score nor SF-6D utility score showed apparent ceiling or floor effects. As can be seen in the 15D dimension distribution of Table 2, the highest (partial) ceiling effects were found in *hearing* (71.9%), *eating* (69.7%), *vision* (53.2), and *breathing* (53.2); while no floor effect was found in any dimen-

Table 1.
Characteristics of Participants and Distribution of PDQ-8, 15-D, and SF-6D Utility Scores

Variables-Based Subgroups	Total	H&YS (Stage I-II) (*n*=148)	H&YS (Stage III-IV) (*n*=81)	*p*-Value
Age (years)				0.154[a]
Median (IQR)	66 (13)	66 (12)	7 (14.50)	
Mean (SD)	66 (9.32)	65.21 (9.24)	67.53 (9.33)	
Range	34-86	34-86	43-84	
Age 34-65				0.53[a]
Median (IQR)	60 (7)	59.50 (7.50)	60 (7)	
Mean (SD)	58.10 (6.55)	57.62 (7.17)	59.06 (5.09)	
Range	34-65	34-65	43-65	
Age 66-90				0.005[a]
Median (IQR)	72 (8.5)	72 (8)	74 (9.25)	
Mean (*SD*)	72.74 (5.07)	71.75 (4.70)	74.60 (5.26)	
Range	66-86	66-86	66-84	
Gender				0.216[b]
Female (%)	77 (33.33)	46 (19.91)	31 (13.41)	
Male (%)	147 (63.63)	100 (43.29)	47 (20.34)	
Level of Studies				0.201[b]
Non-university (%)	175 (75.75)	110 (47.61)	65 (28.13)	
University (%)	51 (22.07)	37 (16.01)	14 (6.06)	
Occupational Status				0.350[b]
Employee (%)	13 (5.62)	10 (4.32)	3 (1.29)	
Retired (%)	216 (93.50)	138 (59.74)	77 (33.33)	
Household Income, € per month#				0.019[b]
From 400-1,800 (%)	53 (22.94)	29 (12.55)	24 (10.38)	
1,801 or more (%)	34 (14.71)	27 (11.68)	7 (3.03)	
Other Medical Conditions∞				0.904[b]
Yes	103 (44.58)	67 (29.00)	36 (15.58)	
No	128 (55.41)	81 (35.06)	45 (19.48)	
Ongoing Medications†				0.733[b]
Yes	214 (92.64)	133 (57.57)	68 (29.43)	
No	13 (5.62)	8 (3.46)	5 (2.16)	
Years Since Clinical Diagnosis				0.045[a]
Median (IQR)	6.00 (8)	6 (10)	7 (8.25)	
Mean (*SD*)	8.19 (7.25)	7.68 (6.57)	9.11 (8.36)	
Range	0-41	0-35	0-41	

continued on next page

Table 1. (continued)
Characteristics of Participants and Distribution of PDQ-8, 15-D, and SF-6D Utility Scores

Variables-Based Subgroups	Total	H&YS (Stage I-II) (*n*=148)	H&YS (Stage III-IV) (*n*=81)	*p*-Value
PDQ-8$_{\text{total score}}$				<0.001[a]
Median (IQR)	25.00 (28.12)	18.75 (23.43)	40.62 (28.12)	
Mean (*SD*)	28.93 (19.23)	23.21 (17.46)	39.72 (17.24)	
Range	0.00-87.50	0.00-87.50	6.25-71.88	
15D				<0.001[a]
Median (IQR)	0.77 (0.24)	0.81 (0.20)	0.65 (0.22)	
Mean (*SD*)	0.73 (0.15)	0.77 (0.14)	0.64 (0.27)	
Range	0.31-1.00	0.31-1.00	0.40-0.91	
SF-6D				<0.001[a]
Median (IQR)	0.54 (0.392)	0.66 (0.336)	0.40 (0.39)	
Mean (*SD*)	0.51 (0.28)	0.59 (0.25)	0.35 (0.27)	
Range	-0.22-1.00	-0.22-1.00	-0.22-0.985	

Values are presented as *n* (%), unless otherwise indicated; ∞ self-reported medical conditions Parkinson's disease apart; † = self-reported ongoing therapies for the control of the Parkinson's disease; 15D$_{\text{utility}}$ = utility index from the 15D; SF-6D = Short Form 6 Dimensions; PDQ-8 = Parkinson's disease questionnaire 8 items; H&YS = Hoehn & Yahr scale; IQR = inter-quartile range.
Lost values higher to 25%
Household income, € per month *n*=144 (62.33%)
[a] = Independent comparisons between groups based on the severity of the symptoms, with Mann–Whitney U test
[b] = Comparisons between groups based on the severity of the symptoms, with chi-square test

sion. The SF-6D did not show a floor effect in any of its six dimensions. Highest ceiling effect was found in the *role limitation* dimension (44.2%).

Construct validity. Correlations between summary score of the PDQ-8 and utility scores of the 15D and the SF-6D are shown in Table 3. The correlation was statistically significant (p<0.001) for both questionnaires and value of the Spearman's rank coefficient showed a strong correlation (r= -0.758 and r= -0.741 respectively). In Table 3, correlation between PDQ-8 dimensions and 15D and SF-6D dimensions can also be seen. All correlations between dimensions were weaker than correlations between utility indexes and PDQ-8 summary score, except the correlation between 15D dimension *depression* and PDQ-8 dimension *emotional well-being* (r= -0.754; p<.001). Most PDQ-8 dimensions showed a moderate correlation (r between -0.3 and -0.5) with dimensions of the two generic instruments. However, Spearman's rank correlation coefficient was always lower than 0.3 (in absolute value) for the PDQ-8 dimension *stigma*. In the same way, the stronger correlation between 15D dimension *hearing* and any PDQ-8 dimension was -0.255, which has been defined as a weak correlation.

Table 4 shows the univariate analyses for 15D and SF-6D. The variables PDQ-8 score and severity of symptoms were dichotomized to compare the utility scores of the two groups. The cut-off points for PDQ-8 score were 5.8, 7.4, and 25.0 (the median), while the severity of symptoms groups were based on stages of H&YS. As expected, both 15D and SF-6D utility scores gradually decreased with increasing PDQ-8 score and severity of symptoms.

Level of agreement between 15D and SF-6D. Paired comparisons of 15D and SF-6D utility scores using Wilcoxon's rank test are shown in Table 4. Statistically significant differences were found between both questionnaires in the whole sample (p<0.001), as well as in all the subgroups. However, an acceptable value between 15D and SF-6D were also found for all patients (ICC 0.78) and by the Spearman's rank correlation coefficient (0.765). ICC values were higher than 0.7 (which represents an acceptable value) in almost all subgroups. Only three subgroups obtained an ICC value lower than 0.7: (a) those patients whose occupational status was *employee* (*n*=13; ICC=0.63), (b) those with a PDQ-8 score lower

Table 2.
Distribution of 15-D and SF-6D Results within Each Domain (*n*=231)

15-D* (%)															
Level	**MO**	**V**	**H**	**B**	**SL**	**E**	**SP**	**EL**	**UA**	**MF**	**DI**	**DE**	**DIS**	**VI**	**SEX**
1	**42.0**	**53.2**	**71.9**	**53.2**	25.5	**69.7**	32.9	29.4	18.2	38.1	**39.4**	33.3	27.3	24.2	22.5
2	30.7	25.5	17.3	23.8	**36.4**	20.3	**44.6**	**42.9**	**46.3**	**42.0**	33.3	**35.5**	**41.1**	**40.3**	**28.6**
3	14.7	13.9	9.1	13.4	18.2	8.7	13.0	22.9	16.5	16.9	18.6	23.4	22.5	24.7	26.0
4	12.1	7.4	1.7	6.5	17.3	1.3	9.1	3.9	12.6	2.2	5.2	6.1	6.5	7.8	9.1
5	0.4	/	/	3.0	2.6	/	0.4	0.9	6.5	0.9	3.5	1.7	2.6	3.0	13.9
At ceiling (the best possible health state); *n* (%) 1 (0.4)															
At floor (the worst possible health state); *n* (%) 0 (0.0)															

SF-6D* (%)						
Level	**PF**	**RL**	**SF**	***p***	**MH**	**V**
1	2.6	**44.2**	25.5	18.2	10.0	6.1
2	15.2	11.7	22.1	13.4	22.9	21.6
3	23.4	9.1	**35.9**	**32.0**	**44.6**	**45.9**
4	4.3	35.1	10.4	16.9	19.9	17.3
5	**37.2**	/	6.1	12.6	2.6	9.1
6	17.3	/	/	6.9	/	/
At ceiling (the best possible health state); *n* (%) 1 (0.4)						
At floor (the worst possible health state); *n* (%) 0 (0.0)						

*15-D dimensions: MO = mobility; V = vision; H = hearing; B = breathing; SL = sleeping; E = eating; SP = speech; EL = elimination; UA = usual activities; MF = mental function; DI = discomfort and symptoms; DE = depression; DIS = distress; VI = vitality; SEX = sexual activities.

*SF-6D dimensions: PF = physical functioning; RL = role limitation; SF = social functioning; P = pain; MH = mental health; V = vitality. Level in mold is in bold.

15-D = 15 dimensions Health-Related Quality of Life Questionnaire; SF-6D = Short Form 6 dimensions

than 5.8 and 25.0 (*n*=14; ICC=0.47 and *n*=113; ICC=0.58), and (c) those in stages III-IV of H&YS (*n*=81; ICC=0.66). Spearman's rank correlation coefficient was statistically significant ($p<0.001$) for both across the sample and according to different sociodemographic and clinical variables, except in those patients whose occupational status was employee (*n*=13; r=0.434). In the rest of the groups, the r value ranged from 0.570-0.835 (see Table 4).

Bland-Altman analysis indicated 95% limits of agreement between 15D and SF-6D ranged from 0.160-0.605 and over 95% lies within those limits (see Figure 1). The bias (*SD*) was 0.223 (0.195). A regular difference in the comparison of 15D and SF-6D is shown in this analysis. Lower scores of SF-6D were observed across the sample. Those differences were higher at lower mean utility scores.

Efficiency and sensitivity of SF-6D and 15D. Table 5 shows the efficiency of utility indexes from 15D and SF-6D at detecting clinically relevant differences in HRQoL (assessed with PDQ-8) and in the severity of the symptoms (evaluated with the H&YS). When individuals were grouped according to PDQ-8 scores, the value of the RE coefficient was less than 1 (0.47), meaning the SF-6D (comparator) was less efficient than the 15D. The 15D was 53% more efficient than the SF-6D when the selected cut-off point of PDQ-8 was 5.8, although when the cut-off point was 7.4, the gap narrowed and the difference was only 8%. Regarding the median, 15D was 21% more efficient than SF-6D. However, when severity of symptoms was used to cluster patients, SF-6D was 2% more efficient than 15D at detecting differences between patients in stages I-II and those in stages III-IV of the H&YS. In addition, AUC scores were close to 1, which indicates the capacity of both instruments to detect clinically relevant differences in HRQoL and severity of symptoms was adequate. Comparing both questionnaires, 15D was better at detecting HRQoL differences, while SF-6D was slightly more

Table 3.
Correlation Between PDQ-8 and Two Quality of Life Questionnaires (15-D and SF 6D)

	PDQ-8								
	Mobility	**Activities of Daily Living**	**Emotional Well-Being**	**Stigma**	**Social Support**	**Cognitions**	**Communication**	**Bodily Discomfort**	**Summary Score**
15-D[a]									
Utility	-0.532**	-0.587**	-0.578**	-0.162*	-0.430**	-0.563**	-0.505**	-0.461**	-0.758**
MO	-0.572**	-0.541**	-0.340**	-0.074	-0.343**	-0.374**	-0.382**	-0.364**	-0.596**
V	-0.381**	-0.390**	-0.314**	0.024	-0.300	-0.409**	-0.322**	-0.287**	-0.463**
H	-0.231**	-0.208**	-0.203**	-0.070	-0.158*	-0.255**	-0.255**	-0.106	-0.271**
B	-0.274**	-0.330**	-0.289**	-0.055	-0.259**	-0.324**	-0.328**	-0.244**	-0.403**
SL	-0.156*	-0.249**	-0.300**	-0.080	-0.173**	-0.264**	-0.194**	-0.302**	-0.305**
E	-0.463**	-0.503**	-0.381**	-0.154*	-0.357**	-0.254**	-0.436**	-0.304**	-0.557**
SP	-0.370**	-0.341**	-0.236**	-0.127	-0.307**	-0.305**	-0.523**	-0.138*	-0.474**
EL	-0.267**	-0.326**	-0.238**	-0.069	-0.241**	-0.299**	-0.302**	-0.292**	-0.408**
UA	-0.484**	-0.600**	-0.405**	-0.115	-0.345**	-0.344**	-0.373**	-0.336**	-0.612**
MF	-0.151*	-0.191**	-0.355**	-0.027	-0.153*	-0.508**	-0.293**	-0.175**	-0.350**
DI	-0.371**	-0.377**	-0.316**	-0.120	-0.262**	-0.353**	-0.221**	-0.389**	-0.481**
DE	-0.274**	-0.278**	-0.754**	-0.244**	-0.309**	-0.438**	-0.308**	-0.325**	-0.563**
DIS	-0.275**	-0.341**	-0.635**	-0.283**	-0.273**	-0.469**	-0.281**	-0.296**	-0.560**
VI	-0.444**	-0.437**	-0.553**	-0.283**	-0.407**	-0.464**	-0.375**	-0.380**	-0.649**
SEX	0.294**	-0.293**	-0.389**	-0.045	-0.338**	-0.310**	-0.237**	-0.233**	-0.412**
SF-6D[b]									
Utility	-0.523**	-0.628**	-0.604**	-0.219**	-0.456**	-0.522**	-0.398**	-0.393**	-0.741**
PF	0.532**	0.724**	0.454**	0.198**	0.397**	0.372**	0.352**	0.301**	0.665**
RL	0.393**	0.447**	0.557**	0.237**	0.312**	0.435**	0.292**	0.288**	0.581**
SF	0.358**	0.340**	0.493**	0.271**	0.356**	0.375**	0.283**	0.223**	0.527**
P	0.372**	0.404**	0.411**	0.058	0.346**	0.449**	0.295**	0.391**	0.538**
MH	0.237**	0.300**	0.588**	0.262**	0.257**	0.401**	0.254**	0.219**	0.476**
V	0.401**	0.391**	0.417**	0.181**	0.283**	0.404**	0.220**	0.237**	0.507**

[a]15-D dimensions: MO = mobility; V = vision; H = hearing; B = breathing; SL = sleeping; E = eating; SP = speech; EL = elimination; UA = usual activities; MF = mental function; DI = discomfort and symptoms; DE = depression; DIS = distress; VI = vitality; SEX = sexual activities

[b]SF-6D: PF = physical functioning; RL = role limitation; SF = social functioning; P = pain; MH = mental health; V = vitality. PDQ-8: MO = mobility; AD = activities of daily living; EW = emotional well-being; ST = stigma; SS = social support; CG = cognition; C = communications; BD = bodily discomfort; SSc = summary score

15-D = 15 dimensions Health-Related Quality of Life Questionnaire; SF-6D = Short Form 6 dimensions; PDQ-8 = Parkinson's disease questionnaire 8 items

*$p<0.05$ (two-tailed); **$p<0.001$ (two-tailed).

Table 4.
Univariate Analyses and Comparison of 15-D and SF-6D Spain Utility Scores for All Patients and for Several Sociodemographic or Health Characteristics-Based Subgroups

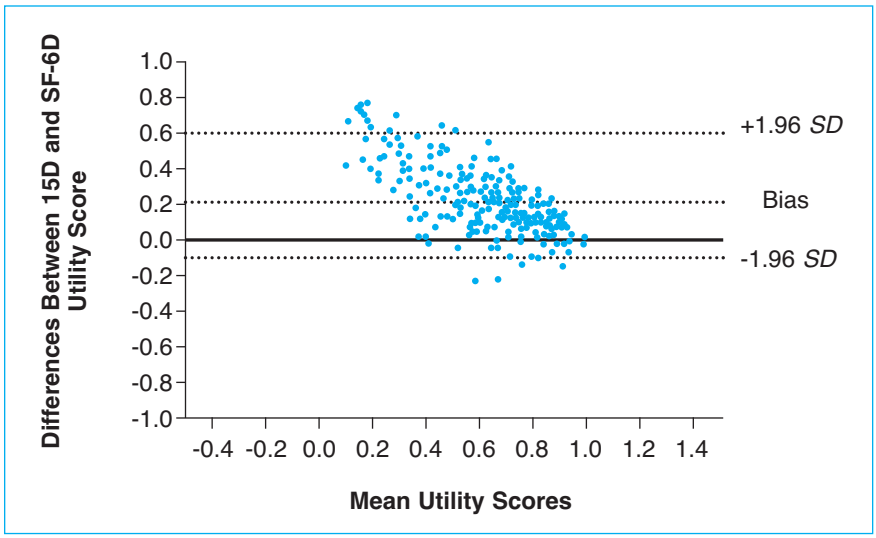
Variables-Based Subgroups		Median (SD)		Median (IQR)				Spearman's Correlation Coefficient
	n (%)	15-$D_{utility}$	SF-6$D_{utility}$	15-$D_{utility}$	SF-6$D_{utility}$	*p*-Value[a]	ICC (95% IC)	
All Patients	231	0.73 (0.15)	0.51 (0.28)	0.77 (0.24)	0.54 (0.39)	<0.001	0.78 (0.72-0.83)	0.765**
Age (years)								
34-65	103 (44.58)	0.74 (0.14)	0.50 (0.27)	0.78 (0.22)[b]	0.53 (0.36)[b]	<0.001	0.75 (0.63-0.83)	0.740**
66-90	121 (52.38)	0.72 (0.16)	0.53 (0.29)	0.75 (0.25)	0.60 (0.37)	<0.001	0.82 (0.74-0.87)	0.800**
Gender								
Female	77 (33.33)	0.68 (0.16)	0.40 (0.29)	0.70 (0.27)	0.43 (0.42)	<0.001	0.81 (0.70-0.87)	0.835**
Male	147 (63.63)	0.75 (0.15)	0.57 (0.25)	0.79 (0.23)[b*]	0.61 (0.34)[b**]	<0.001	0.77 (0.69-0.84)	0.697**
Level of Studies								
Non-university	175 (75.75)	0.71 (0.16)	0.47 (0.28)	0.73 (0.27)[b**]	0.52 (0.38)[b*]	<0.001	0.78 (0.71-0.84)	0.754**
University	51 (22.07)	0.80 (0.12)	0.62 (0.26)	0.81 (0.18)	0.68 (0.29)	<0.001	0.75 (0.57-0.86)	0.737**
Occupational Status								
Employee	13 (5.62)	0.83 (0.07)	0.70 (0.18)	0.83 (0.10)[b*]	0.69 (0.31)[b*]	0.013	0.63 (-0.20-0.88)	0.434
Retired	216 (93.50)	0.72 (0.16)	0.50 (0.28)	0.75 (0.24)	0.53 (0.40)	<0.001	0.78 (0.72-0.83)	0.764**
Other Medical Conditions∞								
Yes	103 (44.58)	0.72 (0.15)	0.49 (0.26)	0.75 (0.24)[b]	0.52 (0.37)[b]	<0.001	0.80 (0.70-0.86)	0.763**
No	128 (55.41)	0.74 (0.16)	0.52 (0.30)	0.77 (0.25)	0.59 (0.36)	<0.001	0.77 (0.68-0.84)	0.770**
Ongoing Medications†								
Yes	214 (92.64)	0.73 (0.15)	0.51 (0.28)	0.77 (0.24)[b]	0.55 (0.38)[b]	<0.001	0.78 (0.72-0.83)	0.758**
No	13 (5.62)	0.66 (0.19)	0.39 (0.33)	0.73 (0.41)	0.36 (0.50)	0.003	0.78 (0.30-0.93)	0.775**
Years Since Clinical Diagnosis								
≤10	146 (63.20)	0.76 (0.14)	0.54 (0.27)	0.79 (0.22)[b*]	0.60 (0.33)[b]	<0.001	0.75 (0.66-0.82)	0.723 **
>10	64 (27.70)	0.68 (0.17)	0.46 (0.29)	0.68 (0.32)	0.51 (0.41)	<0.001	0.80 (0.68-0.88)	0.776**
PDQ-$8_{total\ score}$								
≤5.8	14 (6.06)	0.90 (0.06)	0.79 (0.17)	0.89 (0.10)[b**]	0.82 (0.14)[b**]	0.002	0.47 (-0.64-0.83)	0.702**
>5.8	206 (89.17)	0.72 (0.15)	0.49 (0.28)	0.74 (0.24)	0.53 (0.39)	<0.001	0.78 (0.72-0.83)	0.770**
≤7.4	26 (11.25)	0.88 (0.10)	0.77 (0.19)	0.90 (0.10)[b**]	0.82 (0.14)[b**]	<0.001	0.76 (0.47-0.89)	0.570**
>7.4	194 (83.98)	0.71 (0.15)	0.74 (0.24)	0.47 (0.28)	0.52 (0.38)	<0.001	0.77 (0.70-0.83)	0.754**
≤25.0	113 (51.4)	0.83 (0.09)	0.69 (0.16)	0.85 (0.10)[b**]	0.72 (0.20)[b**]	<0.001	0.589 (0.40-0.71)	0.460**
>25.0	107 (48.6)	0.62 (0.13)	0.31 (0.26)	0.62 (0.19)	0.35 (0.39)	<0.001	0.700 (0.55-0.79)	0.687**
H&YS								
Stages I-II	148 (64.06)	0.77 (0.14)	0.59 (0.33)	0.81 (0.20)[b**]	0.66 (0.33)[b**]	<0.001	0.81 (0.74-0.86)	0.720**
Stages III-IV	81 (35.06)	0.64 (0.14)	0.35 (0.27)	0.65 (0.22)	0.40 (0.39)	<0.001	0.66 (0.48-0.78)	0.644**

Values are presented as *n* (%), unless otherwise indicated; ∞ self-reported medical conditions Parkinson's disease apart; † = self-reported ongoing therapies for the control of the Parkinson's disease; 15-$D_{utility}$ = utility index from the 15-D; SF-6D = Short Form 6 Dimensions; PDQ-8 = Parkinson's disease questionnaire 8 items; H&YS = Hoehn & Yahr scale; ICC = intra-class correlation coefficient; IQR = inter-quartile range.

[a]Paired comparisons of 15-D and SF-6D utilities scores were made with Wilcoxon's signed-rank test. *p<0.05 (two-tailed); **p<0.001 (two-tailed).

[b]Independent comparisons between dichotomized variables groups were made with Mann-Whitney U test. [b*]p<0.05; [b**]p<0.001

Figure 1.
Bland-Altman Plot for All 231 Individuals who Completed both the 15D and the SF-6D



efficient at detecting differences in severity of symptoms (see Table 5).

Discussion

The purpose of this study was to analyze and compare the validity and sensitivity of 15D and SF-6D HRQoL questionnaires in patients affected by PD, accounting for different sociodemographic and clinical variables, such as age, gender, level of studies, occupational status, household income, ongoing therapies, years since clinical diagnosis, and severity of symptoms. Previous studies have demonstrated the validity of the 15D in the PD population (Garcia-Gordillo et al., 2014; Haapaniemi et al., 2004) and have compared the 15D and SF-6D in different populations (Sorensen, Linde, Ostergaard, & Hetland, 2012). However, to our knowledge, no study has compared the psychometric properties of these two questionnaires in this specific population. Therefore, the current study is the first one that provides that relevant information.

According to the results, these values represent a strong correlation and indicate those general preference-based questionnaires are valid for the PD population. The ICC statistic showed an acceptable value across the sample (ICC=0.78), as well as for almost all subgroups. Since the utility index of SF-6D comes from SF-36, the required time to be completed by the respondents is similar for both 15D and SF-6D questionnaires.

The Bland-Altman plot suggests differences between 15D and SF-6D instruments (15D utility score exceeding SF-6D utility score) were higher when the mean utility scores were lower and tended to be lower when mean utility scores became higher. At this point, the different ranges of both utility scores should be taken into account: The 15D utility scores range from 0-1, while SF-6D utility scores range from -0.357-1. This could explain partially those differences. Wilcoxon's signed-rank test also show the difference between 15D and SF-6D. Since different valuation systems are used in each instrument, it is expected to find differences on derived utilities between each of them. These discrepancies have been observed in the rheumatoid arthritis population (Sorensen et al., 2012).

Independent comparisons between subgroups showed, as expected, the group with patients on stages I-II of the H&YS obtained statistically significant higher scores in both questionnaires than patients on stages III-IV. Patients with higher PDQ-8 scores obtained lower utility scores from both 15D and SF-6D. Those differences indicate both questionnaires are sensitive to HRQoL changes in PD.

Age is a main risk factor of PD because different processes which are essential for the function of substantia nigra neurons, including dopamine metabolism, mitochondrial DNA copy number, and protein degradation decline with aging (Brabo, Minett, & Ortiz, 2014; Reeve, Simcox, & Turnbull, 2014); however, statistically significant differences were not found between groups based on age.

Regarding floor and ceiling effects, only one respondent had a score of 1 (best possible health state) in both questionnaires, while no participants obtained the lowest utility score in any of the instruments. Therefore, neither floor effect nor ceiling effect was observed with these questionnaires in patients affected by PD. Floor and ceiling effects are extremely relevant to assess and compare adequately the improvements of any treatment (Fries, Rose, & Krishnan, 2011; Holstein, Avlund, Due, Martinussen, & Keiding, 2006; Ward, Guthrie, & Alba, 2014). It is well known that estimates of the minimal clinically important improvement are larger for subgroups of patients with more severe disease than for subgroups whose scores are closer to full health. However, recent studies suggest the baseline dependency of the minimal clinically important improvement is caused by floor and ceiling effects in the measurement process and not only by the tendency to underestimate or overestimate improvements according to their health status (van der Roer, Ostelo, Bekkering, van Tulder, & de Vet,

Table 5.
Efficiency of 15-D and SF-6D to Detect Clinically Relevant Differences in HRQoL and in the Severity of the Symptoms

				t-Test		ROC curve		
Measure	**PDQ-8**	***n***	**Mean (*SD*)**	***t*-statistic**	***p*-Value**	**RE†**	**AUC**	**95% CI**
$15\text{-}D_{utility}$	≤5.8	14	0.90 (0.06)	9.12	<0.001	1.00	0.85*	(0.78-0.93)
	>5.8	206	0.72 (0.15)					
$SF\text{-}6D_{utility}$	≤5.8	14	0.79 (0.17)	5.89	<0.001	0.47	0.83*	(0.72-0.94)
	>5.8	206	0.49 (0.28)					
$15\text{-}D_{utility}$	≤7.4	26	0.88 (0.10)	7.24	<0.001	1.00	0.85*	(0.76-0.94)
	>7.4	194	0.71 (0.15)					
$SF\text{-}6D_{utility}$	≤7.4	26	0.77 (0.19)	6.98	<0.001	0.92	0.83*	(0.74-0.92)
	>7.4	194	0.47 (0.28)					
$15\text{-}D_{utility}$	≤25	113	0.83 (0.09)	13.72	<0.001	1.00	0.90*	(0.86-0.94)
	>25	107	0.62 (0.13)					
$SF\text{-}6D_{utility}$	≤25	113	0.69 (0.16)	12.26	<0.001	0.79	0.88*	(0.84-0.92)
	>25	107	0.31 (0.26)					
Measure	**H&YS**	***n***	**Mean (*SD*)**	***t*-statistic**	***p*-Value**	**RE†**	**AUC**	**95% CI**
$15\text{-}D_{utility}$	Stages I-II	148	0.77 (0.14)	6.55	<0.001	1.00	0.74*	(0.68-0.81)
	Stages III-IV	81	0.64 (0.14)					
$SF\text{-}6D_{utility}$	Stages I-II	148	0.59 (0.25)	6.64	<0.001	1.02	0.76*	(0.69-0.82)
	Stages III-IV	81	0.35 (0.27)					

$15\text{-}D_{utility}$ = Utility index from the 15-D questionnaire

$SF\text{-}6D_{utility}$ = Spanish utility index from SF-6 dimensions

H&YS = Hoehn & Yahr scale; AUC = area under ROC curves; CI = confidence interval; RE = relative efficiency; ROC = receiver operating characteristic. HRQoL = health-related quality of life; *SD* = standard deviation

*$p<0.001$ indicates that AUC statistically significantly greater than 0.5

†Reference is $15\text{-}D_{utility}$.

2006; Ward et al., 2014). An adequate instrument must have a wide range that reduces the number of patients at the top and bottom of the scale. Only this kind of instrument is useful for calculating QALYs and CUA. Both SF-6D and 15D seem to be sensitive enough to assess HRQoL in patients affected by PD.

Since the 15D seemed to be more efficient comparing groups based on HRQoL, similar results at assessing severity of symptoms were expected. However, both RE and AUC analyses indicated SF-6D utility score was slightly more efficient and sensitive than 15D utility index. The AUC for both instruments was lower for severity of symptoms than for HRQoL (0.85 and 0.83 vs. 0.74 and 0.76). The choice of the HRQoL measurement is relevant to obtain QALYs and performing the subsequent CUA (Sach et al., 2009). Therefore, it would be interesting to compare cost-effectiveness ratios using the same HRQoL instrument and scoring algorithm in different interventions. Results presented in the current study and based on the relative efficiency of both questionnaires at assessing HRQoL changes provide helpful information for future studies on PD. The 15D seems to be more adequate for populations with better health status, while no great differences were found for the rest of the sample. Additionally, lack of floor and ceiling effects in both 15D and SF-6D questionnaires allow researchers and clinicians to choose any of them for economic evaluations and comparisons, depending on their interests and circumstances.

Limitations

This study has some limitations. The clinical measures were self-reported, thus it is possible data may result in biased estimates of prevalence of some conditions (Lubetkin, Jia, Franks, & Gold, 2005). With regard to the

size of the sample, it was collected from 23 different local PD associations from 13 of 17 regions of Spain (to a certain degree representative of the Spanish context). A population-based study with randomized and stratified sample distribution could not be performed, so voluntariness of the participation in the study could have introduced a selection bias. It would be interesting to conduct new studies with larger sample sizes which could confirm the results obtained in the current study, as well as to analyze other properties, such as longitudinal response and reliability.

Policy Implications

These results show HRQoL questionnaires may be a useful working tool for all nursing professionals. Depending on the aspects to assess (detecting clinical HRQoL changes or symptom severity), it is advantageous to use 15D or SF-6D. As a consequence, government agencies, employers, and private providers should increase the collection of quality of life data related to population health, since quality of health care can only be assessed adequately in terms of the related quality of life with the health of the nation.

Conclusions

Both 15D and SF-6D appear to be valid and sensitive preference-based generic HRQoL instruments for Spanish patients with PD. None of these questionnaires showed floor or ceiling effects. The 15D seems to be more efficient and sensitive at detecting clinical HRQoL changes, especially when patients reported good health state. On the other hand, SF-6D was slightly more efficient and sensitive to detect clinical changes in PD severity of the symptoms. **$**

REFERENCES

Abellan Perpinan, J.M., Sanchez Martinez, F.I., Martinez Perez, J.E., & Mendez, I. (2012). Lowering the 'floor' of the SF-6D scoring algorithm using a lottery equivalent method. *Health Economics, 21*(11), 1271-1285. doi:10.1002/hec.1792

Aggarwal, R., Wilke, C.T., Pickard, A.S., Vats, V., Mikolaitis, R., Fogg, L., ... Jolly, M. (2009). Psychometric properties of the EuroQol-5D and Short Form-6D in patients with systemic lupus erythematosus. *The Journal of Rheumatology, 36*(6), 1209-1216.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307-310.

Brabo, N.C., Minett, T.S., & Ortiz, K.Z. (2014). Fluency in Parkinson's disease: Disease duration, cognitive status and age. *Arquivos de Neuro-Psiquiatria, 72*(5), 349-355.

Brazier, J., Deverill, M., & Green, C. (1999). A review of the use of health status measures in economic evaluation. *Journal of Health Services Research & Policy, 4*(3), 174-184.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*(2), 271-292.

Brazier, J.E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*(9), 851-859.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, M., & Jaffray, J.Y. (1988). Certainty effect versus probability distortion: An experimental analysis of decision making under risk. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 554.

Chaudhuri, K.R., Odin, P., Antonini, A., & Martinez-Martin, P. (2011). Parkinson's disease: The non-motor issues. *Parkinsonism & Related Disorders, 17*(10), 717-723. doi: 10.1016/j.parkreldis.2011.02.018

Fayers, P.M., & Machin, D. (2016). *Quality of life: The assessment, analysis and reporting of patient-reported outcomes* (3rd ed.). Hoboken, NJ: Wiley-Blackwell.

Fries, J., Rose, M., & Krishnan, E. (2011). The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *Journal of Rheumatology, 38*(8), 1759-1764. doi:10.3899/jrheum.110402

Garcia-Gordillo, M., del Pozo-Cruz, B., Adsuar, J.C., Sanchez-Martinez, F.I., & Abellan-Perpinan, J.M. (2014). Validation and comparison of 15-D and EQ-5D-5L instruments in a Spanish Parkinson's disease population sample. *Qualily of Life Research, 23*(4), 1315-1326. doi: 10.1007/s11136-013-0569-4

Garcia-Gordillo, M.A., Del Pozo-Cruz, B., Adsuar, J.C., Cordero-Ferrera, J.M., Abellan-Perpinan, J.M., & Sanchez-Martinez, F.I. (2015). Validation and comparison of Eq-5d-3l and Sf-6d instruments in a Spanish Parkinson's disease population sample. *Nutricion Hospitalaria, 32*(6), 2808-2821. doi:10.3305/nh.2015.32.6.9765

Gaujoux-Viala, C., Rat, A.C., Guillemin, F., Flipo, R.M., Fardellone, P., Bourgeois, P., & Fautrel, B. (2011). Comparison of the EQ-5D and the SF-6D utility measures in 813 patients with early arthritis: Results from the ESPOIR cohort. *Journal of Rheumatology, 38*(8), 1576-1584. doi:10.3899/jrheum.101006

Gazewood, J.D., Richards, D.R., & Clebak, K. (2013). Parkinson disease: An update. *American Family Physician, 87*(4), 267-273.

Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., ... Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. (2004). Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations. *Movement Disorders, 19*(9), 1020-1028. doi:10.1002/mds.20213

Guyatt, G.H., Feeny, D.H., & Patrick, D.L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine, 118*(8), 622-629.

Haapaniemi, T.H., Sotaniemi, K.A., Sintonen, H., & Taimela, E. (2004). The generic 15D instrument is valid and feasible for measuring health related quality of life in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 75*(7), 976-983.

Hanmer, J., Feeny, D., Fischhoff, B., Hays, R.D., Hess, R., Pilkonis, P.A., ... Yu, L. (2015). The PROMIS of QALYs. *Health and Quality of Life Outcomes, 13*(1), 122.

Hattie, J., & Cooksey, R.W. (1984). Procedures for assessing the validities of tests using the "known-groups" method. *Applied Psychological Measurement, 8*(3), 295-305.

Hoehn, M.M., & Yahr, M.D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology, 17*(5), 427-442.

Holstein, B.E., Avlund, K., Due, P., Martinussen, T., & Keiding, N. (2006). The measurement of change in functional ability: Dealing with attrition and the floor/ceiling effect. *Archives of Gerontology and Geriatrics, 43*(3), 337-350. doi:10.1016/j.archger.2005.12.004

Jenkinson, C., & Fitzpatrick, R. (2007). Cross-cultural evaluation of the short form 8-item Parkinson's Disease Questionnaire (PDQ-8): Results from America, Canada, Japan, Italy and Spain. *Parkinsonism & Related Disorders, 13*(1), 22-28. doi:10.1016/j.parkreldis.2006.06.006

Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R., & Hyman, N. (1997). The PDQ-8: Development and validation of a short-form Parkinson's dis-

ease questionnaire. *Psychology and Health, 12*(6), 805-814.
Lubetkin, E.I., Jia, H., Franks, P., & Gold, M. R. (2005). Relationship among sociodemographic factors, clinical conditions, and health-related quality of life: Examining the EQ-5D in the U.S. general population. *Quality of Life Research, 14*(10), 2187-2196. doi:10.1007/s11136-005-8028-5
Luo, N., Tan, L.C., Zhao, Y., Lau, P.N., Au, W.L., & Li, S.C. (2009). Determination of the longitudinal validity and minimally important difference of the 8-item Parkinson's Disease Questionnaire (PDQ-8). *Movement Disorders, 24*(2), 183-187. doi:10.1002/mds.22240
Maklin, S., Rasanen, P., Laitinen, R., Kovanen, N., Autti-Ramo, I., Sintonen, H., & Roine, R.P. (2012). Quality-adjusted life-years for the estimation of effectiveness of screening programs: A systematic literature review. *International Journal of Technology Assessment in Health Care, 28*(2), 145-151. doi:10.1017/S0266462312000141
Martinez-Martin, P., Benito-Leon, J., Alonso, F., Catalan, M.J., Pondal, M., Zamarbide, I., ... de Pedro, J. (2005). Quality of life of caregivers in Parkinson's disease. *Quality of Life Research, 14*(2), 463-472.
Martinez-Martin, P., Jeukens-Visser, M., Lyons, K.E., Rodriguez-Blazquez, C., Selai, C., Siderowf, A., ... Schrag, A. (2011). Health-related quality-of-life scales in Parkinson's disease: Critique and recommendations. *Movement Disorders, 26*(13), 2371-2380. doi:10.1002/mds.23834
Murray, C.J., & Lopez, A.D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet, 349*(9063), 1436-1442. doi:10.1016/S0140-6736(96)07495-8
Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics, 14*(11), 1169-1189. doi:10.1002/hec.1006
Pickard, A.S., & Knight, S.J. (2005). Proxy evaluation of health-related quality of life: A conceptual framework for understanding multiple proxy perspectives. *Medical Care, 43*(5), 493.
Pickard, A.S., Wang, Z., Walton, S.M., & Lee, T.A. (2005). Are decisions using cost-utility analyses robust to choice of SF-36/SF-12 preference-based algorithm? *Health and Quality of Life Outcomes, 3,* 11. doi:10.1186/1477-7525-3-11
Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T.D. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders, 29*(13), 1583-1590. doi:10.1002/mds.25945
Reeve, A., Simcox, E., & Turnbull, D. (2014). Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Research Reviews, 14,* 19-30. doi:10.1016/j.arr.2014.01.004
Sach, T.H., Barton, G.R., Jenkinson, C., Doherty, M., Avery, A.J., & Muir, K.R. (2009). Comparing cost-utility estimates: Does the choice of EQ-5D or SF-6D matter? *Medical Care, 47*(8), 889-894. doi:10.1097/MLR.0b013e3181a39428
Sintonen, H. (2001). The 15D instrument of health-related quality of life: Properties and applications. *Annals of Medicine, 33*(5), 328-336.
Soares, M.O. (2012). Is the QALY blind, deaf and dumb to equity? NICE's considerations over equity. *British Medical Bulletin, 101,* 17-31. doi:10.1093/bmb/lds003
Sorensen, J., Linde, L., Ostergaard, M., & Hetland, M.L. (2012). Quality-adjusted life expectancies in patients with rheumatoid arthritis – comparison of index scores from EQ-5D, 15D, and SF-6D. *Value in Health, 15*(2), 334-339. doi:10.1016/j.jval.2011.09.010
Stucki, G., Liang, M.H., Fossel, A.H., & Katz, J.N. (1995). Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology, 48*(11), 1369-1378. doi:0895-4356(95)00054-2 [pii]
Tosh, J.C., Longworth, L.J., & George, E. (2011). Utility values in National Institute for Health and Clinical Excellence (NICE) technology appraisals. *Value in Health, 14*(1), 102-109.
van der Roer, N., Ostelo, R.W., Bekkering, G.E., van Tulder, M.W., & de Vet, H.C. (2006). Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine, 31*(5), 578-582. doi:10.1097/01.brs.0000201293.57439.47
Ward, M.M., Guthrie, L.C., & Alba, M. (2014). Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *Journal of Clinical Epidemiology, 67*(6), 689-696. doi: 10.1016/j.jclinepi.2013.10.025