
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/130561/

Deposited on: 15 November 2016
Title:
Assessment of the construct validity and responsiveness of preference-based quality of life measures in people with Parkinson’s: a systematic review

Authors:
Yiqiao Xin, Emma McIntosh

Concise and informative title:
Measuring preference-based quality of life in Parkinson’s

Affiliation and address:
Health Economics and Health Technology Assessment (HEHTA), Institute of Health and Wellbeing (IHW), University of Glasgow. 1 Lilybank Gardens, Glasgow G12 8RZ.

Corresponding author:
Yiqiao Xin
Yiqiao.xin@glasgow.ac.uk
Tel: +44 (0) 141 330 6245
Fax: +44 (0) 141 330 5018
Abstract:

Purpose

Generic preference-based quality of life (PbQoL) measures are sometimes criticised for being insensitive or failing to capture important aspects of quality of life (QoL) in specific populations. The objective of this study was to systematically review and assess the construct validity and responsiveness of PbQoL measures in Parkinson’s.

Methods

Ten databases were systematically searched up to July 2015. Studies were included if a PbQoL instrument along with a common Parkinson’s clinical or QoL measure was used, and the utility values were reported. The PbQoL instruments were assessed for construct validity (discriminant and convergent validity) and responsiveness.

Results

Twenty-three of 2,758 studies were included, of which the majority evidence was for EQ-5D. Overall, good evidence of discriminant validity was demonstrated in the Health Utility Index (HUI)-3, EQ-5D-5L, EQ-5D-3L, 15D, HUI-2 and Disability and Distress Index (DDI). Nevertheless, HUI-2 and EQ-5D-3L were shown to be less sensitive among patients with mild Parkinson’s. Moderate to strong correlations were shown between the PbQoL measures (EQ-5D-3L, EQ-5D-5L, 15D, DDI, and HUI-II) and Parkinson’s-specific measures. Twelve studies provided evidence for the assessment of responsiveness of EQ-5D-3L and one study for 15D, among which six studies reached inconsistent results between EQ-5D-3L and the Parkinson’s-specific measures in measuring the change overtime.

Conclusions

The construct validity of the PbQoL measures was generally good, but there are concerns regarding their responsiveness to change. In Parkinson’s, the inclusion of a Parkinson’s-specific QoL measure or a generic but broader scoped mental and well-being focused measure to incorporate aspects not included in the common PbQoL measures is recommended.

Keywords
Quality of life

Utility

Parkinson's

Systematic review

PDQ-39

EQ-5D

Construct validity

Responsiveness
Abbreviations

AQoL  Assessment of Quality of Life
CBA   Cost-benefit analysis
CS-PBM Condition-specific preference-based measure
CUA   Cost-utility analysis
DDI   Disability and Distress Index
EQ-5D  EuroQoL EQ-5D
HAD   Hospital Anxiety and Depression scale
HUI   Health Utilities Index
H&Y   Hoehn and Yahr scale
ICER  Incremental cost-effectiveness ratio
MCID  Minimal clinically important difference
NICE  National Institute for Health and Care Excellence
PbQoL Preference-based quality of life
PDQ-39 Parkinson’s Disease Questionnaire-39-item
PDQ-39-SI Parkinson’s disease questionnaire-39-item-Summary Index
PDQL  Parkinson’s Disease Quality of Life questionnaire
PDQUALIF Parkinson’s disease QUALity LIFe scale
PwP   People with Parkinson’s
QALY  Quality-adjusted life-years
QoL   Quality of life
RCT   Randomised controlled trials
SF-6D  Short-Form 6-Dimension
SF-36  Short-Form 36-item
SG    Standard gamble
TTO   Time trade-off
UPDRS Unified Parkinson’s Disease Rating Scale
VAS   Visual analogue scale
**Introduction**

Health state utilities or preference-based quality of life (PbQoL) values are an important parameter in economic evaluations due to their role in the calculation of quality-adjusted life-years (QALYs) for economic evaluations. Typically, incremental QALYs are combined with incremental costs to calculate the incremental cost-effectiveness ratio (ICER) in cost-utility analysis (CUA) [1]. CUA is the preferred form of economic evaluation of government advisory bodies such as the UK’s National Institute for Health and Care Excellence (NICE) for priority setting across disease areas [2].

To generate QALY’s, PbQoL measures are needed. PbQoL measures often comprise a descriptive system (i.e., attributes (or dimensions) and levels) and a value set. The value set typically reflects the preferences of a representative population sample for each of the health states defined by the profile of attributes and levels. These values are commonly elicited using methods such as the standard gamble (SG) [3,4] and time trade-off (TTO) [5]. PbQoL measures typically contain generic attributes, thus facilitating comparative analysis across health areas to assist priority setting. Widely used examples of generic measures include the EuroQol EQ-5D (EQ-5D 3L and 5L versions) [6,7], Short-Form 6-Dimension (SF-6D) [8], and Health Utilities Index (HUI) [9,10]. The EQ-5D is recommended by UK’s NICE to be used in the reference case of economic evaluations [11].

However, the validity of applying such generic measures in some specific populations is the subject of some debate. Generic measures have sometimes been found to be less sensitive to detect changes in quality of life (QoL) in specific populations, for example mental health [12], schizophrenia [13], cancer [14], Alzheimer’s disease [15] and dementia [16]. One suggestion is that the generic attributes making up these measures may not be sufficiently relevant to the specific populations [17]. Longworth et al. [18] valued three condition-specific ‘bolt-on’ attributes as extensions to the EQ-5D related to hearing, tiredness and vision, and found that the ‘bolt-on’ attributes had a significant impact on the values of the health states. Another reason posited for the limitation of the generic measures is that the values attached to the health states are generated from the general public (as recommended by NICE) rather than the specific population in the health states. It is argued that the general public does not have the same experience of the disease as patients and thus cannot reveal the true preference of the specific population being evaluated [19]. A further cited limitation is the discrepancies in utility values when measured with different preference-based instruments [20-24]. Richardson et al. [25] compared the utilities in patients from seven disease areas and compared them with values from healthy members.
from the public using six instruments, including the EQ-5D, SF-6D, HUI3, 15D, Quality of Well-Being, and Assessment of Quality of Life (AQoL). The results revealed that the magnitude of utility difference varied with the choice of instrument by more than 50% for every disease group. Such evidence raises concerns about the external comparability of the values generated by different measures and their ability to reflect true QoL in patients affected by certain conditions.

In comparison with generic QoL measures, condition-specific QoL measures are designed to be more sensitive in their ability to capture the impact of specific diseases or conditions on QoL of the population being affected. However, the QoL scores generated from such condition-specific measures are, by definition, restricted to the specific condition-specific profile of attributes and levels and as such cannot be compared meaningfully with scores obtained from other condition-specific QoL measures. Furthermore, those condition-specific QoL measures are typically not valued, i.e., not preference-based, and hence their use is restricted to ‘within-disease’ priority setting, i.e., cost-effectiveness analysis rather than broader priority setting frameworks such as CUA and cost-benefit analysis (CBA). The summary scores from condition-specific measures are typically unweighted aggregates (additive summation of scores to responses) rather than incorporating preference weights to responses. For example, in Parkinson’s, the Parkinson’s Disease Questionnaire-39-item (PDQ-39) is a common condition-specific non-preference-based QoL questionnaire for use in people with Parkinson’s (PwP). Its summary index (PDQ-39-SI) is calculated by averaging the eight attribute scores [26,27]. Despite accurately measuring the key condition attributes in PwP, this instrument cannot be used in CUA due to the lack of valuation of attributes. Without such ‘valuation’ or ‘inclusion of preferences’ for the health states, no information on how much society would be willing to pay for improvements in scores is obtained. In recent years research has begun to bridge the condition-specific measures/attributes with valuations, examples of which include condition-specific preference-based measures (CS-PBM) [28] and adding condition-specific ‘bolt-on’ attributes to EQ-5D [18]. Despite issues around comparability across disease areas [29], such research is an attempt to complement the limitations of current methods.

Parkinson’s is the second most common neurodegenerative disorder in elderly people, after Alzheimer’s disease [30]. QoL in PwP is affected by motor and non-motor symptoms, as well as medication side effects [31-37]. Utility values in PwP were shown to be the lowest among 29 chronic conditions being evaluated [38]. To our knowledge, there are three published reviews of QoL measures in Parkinson’s [39-41]. Martinez-Martin et al. [39] assessed and classified the generic and specific
health-related QoL scales by psychometric quality to three groups, ‘recommended,’ ‘suggested,’ or ‘listed.’ Soh et al. [40] grouped the commonly used health-related QoL measures into ‘health utility,’ ‘health status,’ and ‘well-being’ and overviewed the use of these measures. Dodel et al. [41] discussed several approaches in economic evaluations in Parkinson’s including the utility instrument. In this study, EQ-5D, SF-6D, 15D, and HUI were assessed according to six criteria of psychometric properties, based on which the authors recommended the use of EQ-5D and HUI to generate utilities along with SG and TTO. However, these studies are not scoped exceptionally for PbQoL, and details were not provided for the assessment of psychometric properties due to the limited space given to PbQoL. Providing these details will benefit the interpretation of the recommendations considering that the process for assessment of psychometric properties is context-sensitive in that the choice of external criteria may have substantial impact on the judgement of the properties.

The objective of this systematic review was to identify, summarize, and assess the psychometric properties including construct validity and responsiveness of PbQoL measures in PwP.

Methods

Search strategy

Electronic databases were searched to identify studies which measured preferences in PwP. The databases included PubMed, MEDLINE, EMBASE, CINAHL, PsycINFO, Applied social sciences Index and Abstracts (ASSIA), Social service abstracts (CSA), AgeInfo, Database of Abstracts of Reviews of Effects (DARE), and NHS EED database. The initial search was conducted in November 2013 and updated in July 2015. A search strategy was developed together with an expert information scientist to maximize the chance of retrieving potential relevant studies (Appendix in ESM).

Inclusion/exclusion criteria

Studies were included if the utility value for people with Parkinson’s (PwP) was measured using a PbQoL instrument and sufficient data were provided to allow the assessment of construct validity and responsiveness. Studies that were eligible for the assessment of convergent validity and responsiveness must also contain a reference measure. The reference measure could be another PbQoL measure, non-preference-based QoL measure, or commonly used clinical measures in Parkinson’s. There are two commonly used clinical measures of Parkinson’s, Unified Parkinson’s
Disease Rating Scale (UPDRS) and Hoehn & Yahr scale (H&Y). The UPDRS assesses clinical status of Parkinson’s in four domains including, mood and cognition, activities of daily living, motor symptoms severity, and complications of treatment [42]. The H&Y describes progression of motor function in Parkinson’s population, ranging from stage I (mildest) to stage V (most severe) [43].

For the assessment of discriminant validity, at least two groups had to be available, divided based on clinical characteristics related to Parkinson’s. PbQoL measure index scores had to be available for those groups. For the assessment of convergent validity, correlation coefficients should be reported between the PbQoL measure and the reference measure. For the assessment of responsiveness, at least two measurements or difference over a period of time (e.g., baseline and primary end point) of both PbQoL measure and the reference measure should be reported. Given this, studies were therefore excluded if the population being measured were patients without a confirmed diagnosis of Parkinson’s; the utilities of PwP were not measured, measured but not reported, not appropriately presented (e.g., EQ-5D index value not on a -0.59-1 scale), or not adequately presented for the assessment purpose; and a full result published later covering the shorter term result in previous papers.

**Data extraction**

After screening (YX), included studies were reviewed and the following study characteristics were extracted (YX): first author and publication year, country, study type, number of participants, clinical characteristics, and length of follow-up (when applicable). Moreover, for the purpose of assessing psychometric properties, study objectives, methods, the measures used, and their scores were also extracted.

**Assessment of construct validity and responsiveness**

Construct validity and responsiveness of the PbQoL approaches used in the included studies were assessed (YX) with methods used in previous studies [18]. **Construct validity** represents the ability that an instrument measures the construct it is intended to measure [44,45]. Construct validity is typically assessed by examining both discriminant validity and convergent validity [18,45-49]. **Discriminant validity** is the extent to which a measure can discriminate across groups that are theoretically known to differ [45,50]. This method is also known as the ‘known group method’ [50]. In this review, we examined to what extent the utility values distinguished between patients with different clinical characteristics of Parkinson’s, with the premise that the QoL of the patients were expected to differ according to these
characteristics. Good evidence of discriminant validity deemed to be demonstrated by a statistically significant difference (e.g., \( t \) test). Given that statistical significance is dependent on sample size, appropriate differences with near significance were also considered as weaker evidence for discriminant validity. Convergent validation is another test of construct validity which is defined as the extent to which one measure correlates with another measure of the same or similar construct \([45,46,50,51]\). In this research, convergent validity is deemed to be demonstrated if the test measure is highly correlated (correlation coefficient \( r \geq 0.5 \)) with a measure of similar concept. A very high correlation \( r > 0.7 \) is not expected in this research due to the inherent difference between the different types of QoL measures. Of the studies that used two or more QoL approaches, we examined the correlation between the approaches; this included both PbQoL and non-preference-based QoL measures. In this assessment, correlations above 0.5 were considered as strong, between 0.3 and 0.5 as moderate, and below 0.3 as weak. Responsiveness is the capacity of an instrument to accurately detect a change when it has occurred over a longitudinal time period \([52,53]\). We examined the extent to which PbQoL measures were able to detect changes in health states overtime as measured by clinical measures or Parkinson’s-specific QoL measures. The change could be due to the health intervention or natural progression of Parkinson’s. As with discriminant validity, good evidence of responsiveness is demonstrated with shown or nearly shown statistically significant difference between the baseline and longest follow-up time point.

Results

A total of 2758 records were retrieved after removing duplicates. Titles and abstracts were screened to identify relevant studies, and 2536 records were excluded based on eligibility criteria. Full text of the remaining 222 studies was further screened from which 23 studies were included in this review. A flowchart of the screening process is shown in Fig. 1.

Included studies were classified into two groups based on their study type for our assessment: Group A: cross-sectional studies \([54-63] \) (including two case-control studies \([59,63] \) ) for assessing discriminant and convergent validity \( (n = 10, \) Tables 1, 2); Group B: longitudinal studies \([64-76] \) for assessing responsiveness \( (n = 13, \) Table 3).
Among the included studies, one study specifically targeted people with early Parkinson’s [69], three targeted advanced Parkinson’s [70,73,76], and the remaining studies recruited PwP with a wide range of severity levels. Five studies explored the relationship between QoL and specific symptoms of Parkinson’s, including apathy [54], depression [56,62], life stress [56], presence of dyskinesia [57], presence of ‘wearing off’ period of drugs [57], sweating dysfunction [63]. Among the longitudinal studies, there were seven RCTs [64,66,67,69,73,75,70], five prospective self-comparison study [65,68,71,74], and one cohort study [72]. Three studies conducted CUA [69,70,76], and one study conducted cost-consequence analysis [75]. Two studies measured patients’ natural progression over a period [71,68]. Eleven studies conducted various interventions, including drugs [69,65,70,73], provision of community-based nurse specialists [66], provision of instructions of clinical guidelines to neurologists [67], standardised pharmaceutical care [72], adherent therapy [64], deep brain stimulation (DBS) surgery [76], and multidisciplinary rehabilitation [74].

Among the PbQoL measures, the EQ-5D was predominantly used, reported in 20 studies [54,55,41,60-69,71-76,57], while the HUI-3 was reported in two studies [59,56], HUI-2 in one [62], 15D in two [55,70], and the Disability and distress index (DDI) (often referred to as the Rosser Index) in one [62]. The DDI, developed by Rosser and colleagues in 1970s, is comprised of eight levels of disability (loss of function and mobility) and four levels of subjective distress, describing 29 disability/distress states [77,78]. One single index score is available for each state, which is generated through valuation process using ranking and relative magnitude of severity exercise [79]. The 15D is a less commonly used instrument developed in Finland [80]. It was chosen in the Norwegian and Swedish studies due to its wider spectrum aspects of QoL, higher sensitivity with five levels on each attribute and availability of value sets in the specific country where the study was conducted [81,82]. Among the reference measures for the assessment of psychometric properties, the PDQ-39 was the most widely used Parkinson’s-specific QoL measure, reported in 9 studies [62-64,66,67,70,71,75,76], followed by the short version of the PDQ-39, the PDQ-8 in 6 studies [55,57,58,61,68,72], the Parkinson’s Disease QUAility of LiFe scale (PDQUALIF) was used in one study [69], the Parkinson’s Disease Quality of Life questionnaire (PDQL) [71] in one, and the generic QoL instrument, the SF-36 in one [75]. The measures used in each of the included studies are presented in Table 4. The characteristics of the QoL measures in the included studies are summarized in Table 5. For transparency, we presented the evidence used for the
assessment of discriminant validity in Table 1, convergent validity in Table 2 and responsiveness in Table 3, along with the study characteristics.

**Assessment of construct validity and responsiveness**

**Assessment of discriminant validity**

Four studies provided adequate evidence for the assessment of the discriminant validity of the EQ-5D-3L [54,57,62,63], two studies for the HUI-3 [56,59], one study for the EQ-5D-5L and 15D [55], and one study for the DDI and HUI-II [62]. For the EQ-5D-3L, groups were defined by the presence of apathy ('with' or 'without') [54], the presence of dyskinesia ('with' or 'without') [57], the presence of 'wearing off' period ('with' or 'without') [57] and a case-control design ('PwP with sweating disturbances', or 'healthy controls') [63]. EQ-5D-3L index scores achieved statistically significant differences between the above-defined groups. One remaining study by Siderowf et al. [62] assessed the ability of EQ-5D-3L, DDI and HUI-2 to discriminate between clinically different groups as defined by a list of criteria. It was found that all of the three measures could differentiate between groups with upper (severe) and lower (mild) halves of UPDRS score ($p < 0.001$) and between first (mildest) and fourth (most severe) quartiles ($p < 0.001$); however, no difference was found in the EQ-5D-3L and HUI-2 between groups with first and second quartiles of UPDRS scores ($p = 0.88$, $p = 0.85$ respectively) while a statistically significant difference was shown in the DDI ($p = 0.03$). All three measures were found to be sensitive to symptoms including falling, freezing, visual hallucinations and depression with a statistically significant unadjusted mean difference between groups divided based on these symptoms ($p < 0.05$). However, no difference was found between groups stratified by dyskinesia or fluctuations for all the three measures, and HUI-2 failed to show difference between groups with and without swallowing difficulty ($p = 0.20$) [62].

For the HUI-3, one case-control study identified a statistically significant difference between PwP and general population, with the HUI-3 score being 0.56 (95% CI 0.48, 0.63) and 0.87 (95% CI 0.87, 0.88) respectively [59]. Another study reported a statistically significant and clinically important difference in HUI-3 values between the groups with and without depression after adjusting for several confounders such as age, sex, duration of Parkinson’s etc. [56]. This study also evaluated the impact of life stress on HUI-3 utility values and identified statistically significant adjusted mean difference between not at all/not very stressful and quite a bit/extremely stressful (adjusted mean difference 0.19 ($p < 0.05$)), but
no difference found between a bit stressful and quite a bit/extremely stressful groups (0.14, \( p < 0.05 \)) [56].

One study reported EQ-5D-5L and 15D values for groups with varied severity of Parkinson’s stratified with H&Y stages, and both instruments showed a statistically significant difference between the groups [55].

**Assessment of convergent validity**

Six studies presented correlation coefficients between a PbQoL measure and a reference measure for the assessment of convergent validity [55, 62, 61, 60, 57, 58]. The EQ-5D-3L score showed strong correlation (\( r = -0.75 \)) with the PDQ-8 summary score [57], moderate to strong correlation with H&Y staging (\( r = -0.32 \) [57], \( r = -0.53 \) [58]), and moderate to strong correlation with the UPDRS total score (absolute \( r \) ranging from 0.39 [57] to 0.72 [58, 61]).

Two studies compared multiple PbQoL measures in terms of their correlations with Parkinson’s-specific QoL measures, and the results were mixed [55, 62]. Garcia-Gordillo et al. [55] found that the utility score from the 15D had a stronger correlation than the EQ-5D-5L with PDQ-8 summary score, with coefficients being -0.710 and -0.679, respectively. The authors explained that this could be due to the broad attributes of 15D such as leisure activities, housework, communication, worries about the future, which are likely to be substantially affected by Parkinson’s [55]. Siderowf et al. [62] compared DDI, EQ-5D-3L, and HUI-II and found that the utility score from EQ-5D-3L correlated strongest with PDQ-39 while DDI showed the weakest correlation. Specifically, they found that the EQ-5D-3L correlated strongest with ADL attribute (\( r = -0.69 \)) and weakest with social support (\( r = -0.27 \)), HUI-II correlated strongest with mobility (\( r = -0.62 \)) and weakest with stigma (\( r = -0.12 \)), and DDI correlated strongest with mobility and ADL (\( r = -0.42 \) for both) and weakest with stigma (\( r = 0.067 \)) [62].

**Assessment of responsiveness**

Thirteen studies provided adequate information to allow an assessment of responsiveness of the PbQoL measures, including 12 studies for the EQ-5D-3L [64-69, 71-76] and one study for the 15D [70]. The one 15D study, by Nyholm et al. [70] demonstrated improved QoL in the duodenal levodopa infusion arm compared to conventional oral polypharmacy arm on both PDQ-39 and 15D (both \( p < 0.01 \)). Six studies showed consistency between the EQ-5D-3L and the reference measures in terms of
the evidence for whether there was a statistically significant change overtime; the reference measures included UPDRS part II ADL [65], PDQ-39 [66,67,76], PDQ-8 and H&Y [68], and HAD depression [74].

The agreement between the EQ-5D and reference measures in the remaining six studies was concerned with various degrees [64,69,71-73,75]. Daley et al. [64] reported statistically significant better QoL as shown on PDQ-39 summary score, mobility, ADL, emotional well-being, cognition, communication and bodily discomfort after adherence therapy as compared to routine care in a RCT, but the change in EQ-5D-3L was small and not statistically significant (mean difference 0.07, 95% CI -0.1, 0.2). Similarly, Schroder et al. [72] detected an improvement ($p = 0.034$) in PDQ-8 score in the group with standardised community pharmaceutical care for eight months and deterioration ($p = 0.019$) in the group with usual care, but the statistically significant difference was not shown in EQ-5D-3L score for either groups. Stocchi et al. [73] compared adjunctive ropinirole prolonged release and immediate release in a RCT and reported an improved UPDRS total motor score ($p = 0.022$), but a non-significant improved UPDRS ADL score ($p = 0.270$) and EQ-5D-3L score ($p = 0.165$). Reuther et al. [71] evaluated the change in QoL and clinical measures over one year among 145 PwP and found that clinical scores deteriorated (H&Y, $p = 0.000$, and UPDRS, $p = 0.019$); however the scores of PDQ-39 and PDQL improved (PDQ-39, $p = 0.000$, and PDQL, $p = 0.030$), and there was no difference in the EQ-5D ($p = 0.488$). In contrast, two studies showed statistically significant change overtime in the EQ-5D but not in the reference measures [69,75]. Noyes et al. [69] compared pramipexole and levodopa in a RCT over four years and did not detect a difference in PDQUALIF, but EQ-5D showed a difference between the arms from year 2 to 3 (0.048, $p = 0.03$) and 3 to 4 (0.071, $p = 0.04$). Wade et al. [75] compared multidisciplinary rehabilitation program versus usual care, in which statistically significant difference was shown between the arms in the SF-36 physical score and EQ-5D score, while no difference found for PDQ-39 and SF-36 mental score.

**Discussion**

This study systematically reviewed and assessed the psychometric properties of PbQoL measures in PwP. The EQ-5D-3L was found to be predominantly used as the PbQoL measure in Parkinson’s while the PDQ-39 was the most widely used Parkinson’s-specific QoL measure among included studies. EQ-5D-3L has achieved statistically significant difference between the known groups divided based on
clinical characteristics in most studies, but it may have limited sensitivity to detect differences in QoL among patients with mild Parkinson’s as evidenced by the subgroup analysis in the included studies [62]. Good evidence of discriminant validity has also been demonstrated in the HUI-3, EQ-5D-5L, 15D, HUI-2, and DDI despite limited evidence being available to allow the assessment. HUI-2 may be less sensitive among patients with mild Parkinson’s as it cannot differentiate between patients with first and second quartile UPDRS scores [62]. In terms of convergent validity, overall moderate to strong correlations were shown between the PbQoL measures (EQ-5D-3L, EQ-5D-5L, 15D, DDI, and HUI-II) and Parkinson’s-specific QoL measures/clinical measures. It was found that the EQ-5D-3L, DDI, and HUI-II all correlated strongest with the physical attributes (i.e., mobility and ADL) of PDQ-39 and weakest with mental and well-being attributes (i.e., social support and stigma). For responsiveness, most evidence was found for the EQ-5D-3L. The agreement between EQ-5D-3L and the Parkinson’s-specific QoL/clinical measures varied across studies. Half of the studies showed that EQ-5D scores reflected changes in clinical status overtime as shown on the reference measures, while the other half failed to reach consistent conclusions between the measures.

There is evidence from this review that the generic PbQoL measures correlate more strongly with the physical attributes than mental/well-being attributes of PDQ-39. Parkinson’s is a chronic, progressive condition which has been shown to affect mental/well-being aspects of QoL and as such it is important to include appropriate valuations for improvements in such attributes within priority setting decisions. The importance of these mental/well-being aspects is demonstrated by consistent presence of such attributes within Parkinson’s-specific QoL measures and by previous literature examining the effect of the mental and well-being aspects on PwP’s QoL [83,33]. With approximately half of the domains in PDQ-39/PDQ-8, PDQUALIF, and PDQL relating to aspects other than physical health, such domains, e.g., social communication, stigma/self-image, emotional functioning, cognition, and outlook, are highly likely to have a substantial impact on PwP’s QoL. A recent systematic review found that depression was the most frequently identified determinant of health-related QoL in PwP among all the demographic and clinical factors [84]. Therefore, sufficient incorporation of valuations for these broader attributes is crucial when measuring PbQoL in Parkinson’s. The utilities from the PbQoL measures generally discriminated well between groups and correlated well with Parkinson’s clinical and QoL measures. However, the inconsistency in findings of responsiveness between those measures cautioned that the change shown on clinical measures may not necessarily lead to the same change in QoL scores.
Reuther et al. [71] assumed that there might be other undetected factors leading to the opposite change of QoL scores to the clinical measures. One reason might be the fact that clinical measures such as H&Y and UPDRS focus mostly on the physical symptoms of Parkinson's while QoL measures are subjective to individuals and based on overall experience of health and well-being. This may also help explain our finding that the PbQoL measures that focused on physical health should be theoretically able to discriminate between groups defined by clinical factors. Besides this, as clinical status or objective health status is usually one of the primary predictors of QoL, it is reasonable to expect that PbQoL measures would display discriminant and convergent validity.

Responsiveness of PbQoL measures is crucial to economic evaluations. In a bid to measure resource use and QALYs, economic evaluations often need to be carried out longitudinally over an appropriate and meaningful time horizon depending upon the intervention being assessed. Previous studies have suggested that the results of economic evaluations are sensitive to the change of utility values when chronic conditions or long-term sequelae are involved [85]; Parkinson’s is one of those conditions. Therefore, lack of definite evidence of responsiveness may critically undermine the results of CUA analysis in Parkinson’s and thus decision making as QALY gains may differ depending on the derivation of utility values. To overcome the limited responsiveness of generic PbQoL measures in certain populations, CS-PBM have been developed in recent years, e.g., in patients with asthma [86] and urinary incontinence [87]. Researchers were concerned that CS-PBM would lose the ability of comparability across disease areas, sometimes insensitive in measuring the side-effects which have differed symptoms from the condition, and lack of comprehensiveness in people with comorbidities due to the narrow scope [29,88]. However, the development of CS-PBM is argued to be valuable as it enriches the database of utilities measured by different approaches in a disease area where it exists limitations with current methods [29] and may provide valuable supplements to existing generic measures [88].

There are a number of limitations of this research. Previous studies have argued that given that no gold standard has been established for measuring PbQoL, the test of validity can only provide a reference of a measure’s performance rather than leading to a rigorous conclusion [89]. Our study assumed that the PDQ-39 or other Parkinson’s-specific measures was a ‘benchmark’ since those measures were designed specifically for Parkinson’s and hence they should be the most relevant measures to Parkinson’s. Another related limitation of the assessment methods relates to the test of convergent
validity. Correlating the PbQoL against another non-preference QoL measure is arguably not the best test of convergent validity since the former is a weighted/valued measure while the latter is not. Despite this, as both instruments were designed to measure QoL, the trend of the scores (i.e., higher value represents better QoL) should be similar and therefore the validity of the test should still provide useful information. The third limitation is that floor and ceiling effects were not assessed in this study. It was found that the EQ-5D and HUI-2 have limited ability to discriminate between patients with varied levels of mild Parkinson’s. This may be related to the ceiling effect of the EQ-5D and HUI-2 as found in other studies [90-92,24]. This ceiling effect, if present, will affect the discriminant validity and responsiveness of the PbQoL measure so that it cannot discriminate between people who all produce 1 (full health) but have different QoL in real life. Similarly, the indicator for convergent validity, the correlation coefficient will become lower if there are ceiling effects, because when the reference score is higher, the PbQoL would not change along since it is capped at 1. This effect however may not have large impact in a Parkinson’s population in general. This is because the QoL for this population is usually at low middle to upper middle range as shown in the included studies, and thus it is not likely to have large proportion of responses of full health. A final note is that the ‘minimal clinically important difference’ (MCID) was not specified in the criteria for responsiveness due to the lack of information regarding how much MCID could be in the Parkinson’s population for the PbQoL measures. There is one published study assessing MCID for PDQ-39 and suggested that the MCID differs across dimensions [93]. One conference abstract estimated MCID for EQ-5D based on the PDQ-39 scores and the UPDRS to be 0.11 (range: 0.08 - 0.14) and 0.10 (range: 0.04 - 0.17), respectively [94]. As no other information was found regarding the MCID for PbQoL in Parkinson’s, MCID was not used in our assessment criteria. Nevertheless, our criteria were not rigid on ‘statistically significant difference’ considering the sample size issue and thus ‘nearly significance’ was also accepted.

Conclusion

The construct validity of the PbQoL measures identified in this review was generally good, but there were concerns regarding their responsiveness to the change in QoL overtime. Given the current requirement in countries such as the UK to report QALY (typically using the EQ-5D instrument) as the
preferred outcome measure in economic evaluations, it is therefore important to ensure adequately broader estimation of PwP’s utilities for resource allocation decisions in Parkinson’s. The development of methods to incorporate broader aspects into health care decision making may represent a valuable research development in this area. In addition, incorporation of the Parkinson’s-specific QoL measures would be beneficial alongside a generic PbQoL measure in longitudinal studies as to sensitively capture the full impact of QoL by Parkinson’s.
**Conflict of Interest:** Dr Emma McIntosh is funded by a Parkinson’s UK Senior Fellowship. Yiqiao Xin declared no conflict of interest.

**Ethical approval:** This manuscript is a systematic review which only contains data from previously published studies. No clinical trials were conducted nor patient data were collected for this research.
Database search after removing duplicates
n = 2,758

Studies excluded (title and abstract screening)
n = 2,536

Full text screened
n = 222

Studies excluded (n = 199):
Reasons:
- Economic modelling with utility data from other sources (n=48)
- Reviews, methodology, protocol (n = 48)
- Cost study (n=16)
- Measured QoL but did not value (n=23)
- Updated paper existed (n=12)
- Measured utilities but did not report (n=13)
- Measured utilities of carers of PwP rather than PwP (n=5)
- Diagnosis of Parkinson’s was not confirmed in the patient group (n=3)
- Insufficient data to assess psychometric properties (n=31)

Included (n=23):
- Cross-sectional studies (n=10)
- Longitudinal studies (n = 13)

Figure 1. Flowchart of study screening
Table 1

Characteristics of included studies – assessment of discriminant validity

<table>
<thead>
<tr>
<th>Study</th>
<th>Publication year</th>
<th>Country</th>
<th>No. of participants</th>
<th>Stage of Parkinson’s (Early or Advanced)</th>
<th>Other characteristics</th>
<th>Study eligibility criteria</th>
<th>Study type</th>
<th>Group define criteria (C) and groups (G)</th>
<th>Evidence for discriminant validity: mean (standard deviation)</th>
<th>Discriminant validity assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benito-Leon et al. [54]</td>
<td>2012</td>
<td>Spain</td>
<td>557</td>
<td>Both</td>
<td>Cross-sectional</td>
<td>C: presence of apathy defined as Lille Apathy Rating Scale. G1: Noapathetic; G2: Apathetic</td>
<td>UPDRS motor: G1: 17.1 (8.5); G2: 24.8 (11.3); p &lt; 0.001; H&amp;Y: Higher proportion of early stages in G1; p &lt; 0.001</td>
<td>EQ-5D: G1: 0.83 (0.17); G2: 0.64 (0.26); p &lt; 0.001. All attributes of EQ-5D showed sig</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garcia-Gordillo et al. [55]</td>
<td>2013</td>
<td>Spain</td>
<td>133</td>
<td>Both</td>
<td>Cross-sectional</td>
<td>C: H&amp;Y. G1: H&amp;Y stages 1-2; G2: H&amp;Y stages 3-4</td>
<td>PDQ-8: G1: 18.30 (11.83); G2: 31.58 (19.56); p &lt; 0.001</td>
<td>EQ-5D-5L: G1: 0.70 (0.18); G2: 0.53 (0.28); p &lt; 0.001. 15D: G1: 0.81 (0.10); G2: 0.70 (0.17); p = 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jones et al. [56]</td>
<td>2009</td>
<td>Canada</td>
<td>259</td>
<td>Both</td>
<td>Cross-sectional</td>
<td>C: depression. G1/G2: without/with depression; C: life stress. G1/G2/G3: not at all/a bit/very stressful</td>
<td>NA</td>
<td>HUI-3: G1: 0.49 (95% CI 0.39, 0.59); G2: 0.20 (95% CI 0.03, 0.37); p (G1 vs. G2) &lt; 0.05. G1': 0.42 (95% CI 0.29, 0.55); G2': 0.38 (95% CI 0.24, 0.51); G3': 0.23 (95% CI 0.10, 0.36); p (G1' vs. G3') &lt;0.05; p (G2' vs. G3') &gt;0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo et al. [57]</td>
<td>2009</td>
<td>Singapore</td>
<td>206</td>
<td>Both</td>
<td>Cross-sectional</td>
<td>C: presence of dyskinesia. G1: no dyskinesia; G2: with dyskinesia. C: presence of ‘wearing off’ periods. G1': no ‘wearing off’; G2': with ‘wearing off’</td>
<td>NA</td>
<td>EQ-5D: p (G1 vs. G2) &lt; 0.01; p (G1' vs. G2') &lt; 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Publication year</td>
<td>Country</td>
<td>No. of participants</td>
<td>Study eligibility criteria</td>
<td>Study type</td>
<td>Group define criteria(C) and groups (G)</td>
<td>Evidence for discriminant validity: mean (standard deviation)</td>
<td>Discriminant validity assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pohar et al. [59]</td>
<td>2009</td>
<td>Canada</td>
<td>261</td>
<td>- Data from Canadian Community Health Survey</td>
<td>Cross-sectional, case-control</td>
<td>C: presence of Parkinson's. G1: With Parkinson's; G2: general population</td>
<td>Age: G1: 68.9 (95% CI 66.6, 71.2); G2: 44.8 (95% CI 44.8, 44.9); p &lt; 0.05. No. of medical conditions: G1: 3.0 (95% CI 2.5, 3.4); G2: 1.5 (95% CI 1.5, 1.5); p &lt; 0.05</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siderowf et al. [62]</td>
<td>2002</td>
<td>US</td>
<td>97</td>
<td>Without cognitive impairment</td>
<td>Cross-sectional</td>
<td>C: total UPDRS score. G1 and G1*: upper and lower halves; G2 and G2*: 1st and 2nd quartiles; G3 and G3*: 1st and 4th quartiles. C: depression. G4 and G4*: with and without depression; and a various motor &amp; non motor symptoms</td>
<td>NA</td>
<td>EQ-SD: Diff (G1vs.G1*:0.24; p &lt; 0.001; Diff (G2vs.G2*):0.09;p = 0.88; Diff (G3vs.G3*):0.40;p &lt; 0.001; Diff (G4vs.G4*):0.26;p &lt; 0.001. DDI: Diff (G1vs.G1*):0.09;p = 0.007; Diff (G2vs.G2*):0.01;p = 0.03; Diff (G3vs.G3*):0.17;p = 0.02; Diff (G4vs.G4*):0.17;p &lt; 0.001. HUI-3: Diff (G1vs.G1*):0.15;p = 0.001; Diff (G2vs.G2*):0.008;p = 0.85; Diff (G3vs.G3*):0.25;p = 0.001; Diff (G4vs.G4*):0.17;p &lt; 0.001.</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Swinn et al. [63]</td>
<td>2003</td>
<td>UK</td>
<td>77</td>
<td>Patients with sweating disturbances, without marked cognitive impairment or confusion</td>
<td>Cross-sectional, case-control</td>
<td>Case-control. G1: PwP with sweating disturbances; G2: healthy controls</td>
<td>PDQ-39: G1: 41.7 (19.5); G2: NA</td>
<td>EQ-SD: G1: 0.47; G2: 0.85; p &lt; 0.005</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Assessment result for discriminant validity: ✓ evidence available to demonstrate that the PbQoL measure was able to show statistically significant difference between the known groups that were expected to differ as shown by the reference measure; o some evidence available but still uncertain whether PbQoL measure can show statistically significant difference between the known groups that were expected to differ; * evidence showing the PbQoL measure failed to differentiate between the known groups.

MMSE - Mini-Mental State Examination, H & Y Hoehn & Yahr scale, HAD Hospital Anxiety and Depression Scale, SCOPA-Motor Scales for Outcomes in Parkinson's disease – Motor examination, UPDRS Unified Parkinson's Disease Rating Scale, Diff mean difference between groups. Sig statistically significance. C criteria, G group, NA not available.

*Reference measure could be either another PbQoL measure, Parkinson's-specific QoL measure, or (if the former two not available) clinical measures.
### Table 2

**Characteristics of included studies – assessment of convergent validity**

<table>
<thead>
<tr>
<th>Study</th>
<th>Publ ication year</th>
<th>Country</th>
<th>No. of participants</th>
<th>Study eligibility criteria</th>
<th>PbQoL measure(s)</th>
<th>Evidence for convergent validity</th>
<th>Convergent validity assessment¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garci-Gordillo et al.</td>
<td>2013</td>
<td>Spain</td>
<td>133</td>
<td>Both</td>
<td>EQ-5D-5L, 15D</td>
<td>PDQ-8</td>
<td>15D/PDQ-8: -0.710. EQ-5D-5L/PDQ-8: -0.679</td>
</tr>
<tr>
<td>Luo et al. [57]</td>
<td>2009</td>
<td>Singapore</td>
<td>31</td>
<td>Both</td>
<td>EQ-5D</td>
<td>PDQ-8 SI, H&amp;Y, UPDRS motor</td>
<td>EQ-5D/PDQ-8: -0.75. EQ-5D/H&amp;Y: -0.32. EQ-5D/UPDRS motor: -0.39</td>
</tr>
<tr>
<td>Rodrigue-Blazquez et al.</td>
<td>2010</td>
<td>Spain</td>
<td>387</td>
<td>Both</td>
<td>EQ-5D</td>
<td>SCOPA-AUT</td>
<td>EQ-5D/SCOPA-AUT: -0.49</td>
</tr>
<tr>
<td>Rodrigue-Blazquez et al.</td>
<td>2013</td>
<td>Argentina, Cuba, Mexico, US, and Spain</td>
<td>435</td>
<td>Both</td>
<td>EQ-5D</td>
<td>MDS UPDRS M-EDL</td>
<td>EQ-5D/UPDRS M-EDL: -0.72</td>
</tr>
<tr>
<td>Siderowf et al. [62]</td>
<td>2002</td>
<td>US</td>
<td>97</td>
<td>Both</td>
<td>EQ-5D, DDI, H&amp;Y</td>
<td>PDQ-39 all sub-attributes, UPDRS</td>
<td>EQ-5D/PDQ-39 all attributes; from -0.27 (social support) to -0.69 (ADL). EQ-5D/UPDRS total: -0.61. HU1 PDD-39 all attributes; from -0.12 (sligma) to -0.82 (mobility). HU1UPDRS total: -0.59. DDI/PDQ-39 all attributes; from 0.067 (stigma) to -0.42 (mobility/ADL). DDIUPDRS total: -0.40</td>
</tr>
</tbody>
</table>

Assessment result for convergent validity: ‘✓’ evidence available to demonstrate that PbQoL measure and the reference measure were highly related (r ≥ 0.5); ‘ ’ the PbQoL measure and the reference measure were moderately correlated (0.3 ≤ r < 0.5); ‘ ’ the PbQoL measure and the reference measure were weakly correlated (r < 0.3). NMSS Non-Motor Symptoms Scale, SCOPA-AUT SCales for Outcomes in Parkinson’s disease – AUTonomnic, ADL Activities of Daily Living, H&Y Hoehn & Yahr stage, MDS UPDRS M-EDL Movement Disorders Society Unified Parkinson’s Disease Rating Scale – Motor Experiences of Daily Living section, r correlation coefficient

* Reference measure could be either another PbQoL measure, Parkinson’s-specific QoL measure, or (if the former two not available) clinical measures.
### Characteristics of included studies – assessment of responsiveness

<table>
<thead>
<tr>
<th>Study</th>
<th>Publication year</th>
<th>Country</th>
<th>No. of participants</th>
<th>Study eligibility criteria</th>
<th>Study type and time horizon</th>
<th>Intervention (I) and comparator (C) or: before (B) and after (A)</th>
<th>Evidence for responsiveness – change from baseline to primary endpoint: Mean change (standard deviation)</th>
<th>Responsiveness assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daley et al. [64]</td>
<td>2014</td>
<td>UK</td>
<td>76</td>
<td>Both On anti-parkinsonian drug(s), no dementia</td>
<td>RCT, 12 wks</td>
<td>I: adherence therapy; C: routine care</td>
<td>PDQ-39: l: -6.8 (6.4); C: 2.3 (7.4); Diff.: 9.0 (95% CI -12.2, -5.8); ( p &lt; 0.001 )</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D: l: 0.04 (0.3); C: -0.03 (0.3); Diff.: 0.07 (95% CI -0.1, 0.2); ( p = 0.055 )</td>
<td>×</td>
</tr>
<tr>
<td>Ebersbach et al. [65]</td>
<td>2010</td>
<td>Germany</td>
<td>61</td>
<td>Both Responsive to levodopa, had not responded to or did not tolerate entacapone, age 30-80, H&amp;Y 2-4, on stable medication for ≥ 4 wks</td>
<td>Before and after self-comparison, 4 wks</td>
<td>B and A: tolcapone targeting sleep quality</td>
<td>UPDRS part II (ADL): B*: 15.1 (7.1); A*: 10.8 (7.0); ( p &lt; 0.0001 )</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D: B*: 0.562 (0.234); A*: 0.678 (0.206); ( p = 0.0001 )</td>
<td>✓</td>
</tr>
<tr>
<td>Jarmann et al. [66]</td>
<td>2002</td>
<td>UK</td>
<td>1859</td>
<td>Both On anti-parkinsonian drug(s)</td>
<td>RCT, 2 yrs</td>
<td>I: provision of community based nurses specialists; C: no provision. B and A: Also analysed deterioration over 2 yrs of all participants</td>
<td>PDQ-39: B and A: all sub-attributes: ( p &lt; 0.05 ); Diff.: 0.47 (95% CI -2.72, 3.66); ( p = 0.77 )</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D: B and A: -0.10 (-0.12, -0.08); ( p &lt; 0.001 ); Diff.: -0.02 (95% CI -0.06, 0.02); ( p = 0.30 )</td>
<td>✓</td>
</tr>
<tr>
<td>Larisch et al. [67]</td>
<td>2011</td>
<td>Germany</td>
<td>386</td>
<td>Both Not reported</td>
<td>Cluster RCT, 9 mths</td>
<td>I: providing instructions of clinical practice guidelines to neurologists; C: without instructions</td>
<td>PDQ-39: l: 1.8 (11.2); C: 1.1 (11.5); ( p^2=0.7591 )</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D: l: -0.001 (0.195); C: 0.007 (0.209); ( p^2=0.5148 )</td>
<td>✓</td>
</tr>
<tr>
<td>Luo et al. [68]</td>
<td>2010</td>
<td>Singapore</td>
<td>31</td>
<td>Both Well enough to complete surveys</td>
<td>Before and after self-comparison, 4 yrs</td>
<td>No intervention</td>
<td>PDQ-8 SI: B*: 17.74 (14.17); A*: 35.08 (17.43); ( p &lt; 0.0001 ). H&amp;Y: B*: 2.09 (0.38); A*: 2.40 (0.70); ( p = 0.0133 )</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D: B*: 0.76 (0.23); A*: 0.52 (0.33); ( p = 0.0014 )</td>
<td>✓</td>
</tr>
<tr>
<td>Noyes et al. [69]</td>
<td>2006</td>
<td>US</td>
<td>301</td>
<td>Early Age ≥ 30, duration with Parkinson’s ≥ 7 yrs, H&amp;Y 1-3, required dopaminergic anti-Parkinson’s therapy</td>
<td>RCT, 4 yrs; cost-utility analysis</td>
<td>I: pramipexole; C: levodopa</td>
<td>PDQUALIF: Diff over 4 yrs: 0.040; ( p = 0.45 ). Diff from yr 2 ~3: 0.015; ( p = 0.36 )</td>
<td>✓</td>
</tr>
<tr>
<td>Study</td>
<td>Publication year</td>
<td>Country</td>
<td>No. of participants</td>
<td>Stage of Parkinson’ s (Early or Advanced)</td>
<td>Intervention (I) and comparator (C) or; before (B) and after (A)</td>
<td>Other clinical characteristics</td>
<td>Study eligibility criteria</td>
<td>Study type and time horizon</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Nyholm et al.[70]</td>
<td>2005</td>
<td>Sweden</td>
<td>24</td>
<td>Advanced</td>
<td>Crossover RCT, 2 three wks trial plus 6 mths follow up; cost-utility analysis</td>
<td>Experiencing motor fluctuations and dyskinesia</td>
<td>2005</td>
<td>Sweden</td>
</tr>
<tr>
<td>Reuthe r et al. [71]</td>
<td>2007</td>
<td>Germany</td>
<td>145</td>
<td>Both</td>
<td>Prospective self-comparison non-intervention, 12 mths</td>
<td>Not reported</td>
<td>2007</td>
<td>Germany</td>
</tr>
<tr>
<td>Schröder et al. [72]</td>
<td>2012</td>
<td>Germany</td>
<td>161</td>
<td>Both</td>
<td>Cohort study, 8 mths</td>
<td>On anti-parkinsonian medication(s), age &gt;35, sufficient physical and cognitive ability to complete questionnaires without assistance</td>
<td>2012</td>
<td>Germany</td>
</tr>
<tr>
<td>Stocchi et al. [73]</td>
<td>2011</td>
<td>Bulgaria, Canada, Czech Republic, France, Hungary, Poland, Romania, Spain, UK.</td>
<td>177</td>
<td>Advanced</td>
<td>RCT, 24 wks</td>
<td>Age ≥30, H&amp;Y 2-4, not adequately controlled on L-dopa (3-12 hrs of daily awake time spent as ‘off’ time)</td>
<td>2011</td>
<td>Bulgaria, Canada, Czech Republic, France, Hungary, Poland, Romania, Spain, UK.</td>
</tr>
<tr>
<td>Study</td>
<td>Publication year</td>
<td>Country</td>
<td>No. of participants</td>
<td>Study eligibility criteria</td>
<td>Study type and time horizon</td>
<td>Intervention (I) and comparator (C) or: before (B) and after (A)</td>
<td>Evidence for responsiveness – change from baseline to primary endpoint: Mean change (standard deviation)</td>
<td>Responsiveness assessment</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
</tbody>
</table>
| Trend et al. [74]    | 2002             | UK               | 118                | Score of at least 7/10 on Hodkinson’s mini-mental test; no cognitive impairment | Before and after self-comparison                      | B and A: intensive multidisciplinary rehabilitation | HAD anxiety: B*: 5.51 (3.31); A*: 5.19 (3.43); p value not sig.  
HAD depression: B*: 6.06 (2.88); A*: 5.57 (2.80); p = 0.029.  
*p value of all of the other motor and non-motor scales achieved sig.  | EQ-5D:  
B*: 0.55 (0.24); A*: 0.63 (0.22); p = 0.001 | ✓ |
| Wade et al. [75]     | 2003             | UK               | 94                 | Without severe cognitive losses                                 | Crossover RCT, 24 wks; cost-consequence analysis      | I: multidisciplinary rehabilitation program; C: usual care     | PDQ-39:  
B: 25.5 (10.7); A: 26.0 (12.7); p = 0.687.  
SF-36 physical: B: 29.5 (11.1); A: 27.28 (10.9); p = 0.046.  
SF-36 mental: B: 51.0 (8.4); A: 50.5 (10.3); p = 0.655 | EQ-5D:  
B: 0.72 (0.22); A: 0.66 (0.21); p = 0.026 | 0 |
| Zhu et al. [76]      | 2014             | HK (China)       | 13                 | Disabling or troubling motor symptoms, dopa responsive, clear understanding risk of and realistic about surgery outcomes, age<70 | Prospective before and after self-comparison, 2 yrs; cost utility analysis (before-after) | B and A: deep brain stimulation surgery | PDQ-39:  
B*: 39 (13); A*: 27 (14); p = 0.019. | EQ-5D:  
B*: 0.504 (0.24); A*: 0.662 (0.13); p = 0.033 | ✓ |

ADL activities of daily living, H&Y Hoehn & Yahr scale, HAD Hospital Anxiety and Depression scale, UPDRS Unified Parkinson’s Disease Rating Scale, I intervention group, C control group, B before, A after, Diff difference of scores between the changes of the two comparative groups over the trial period, yrs years, mths months, hrs hours, sig significant  
* Reference measure could be either another PbQoL measure, Parkinson’s-specific QoL measure, or (if the former two not available) clinical measures.  
+ Difference between the intervention group and the control group at endpoint (no difference was found between two groups at baseline.)  
+ Assessment result for responsiveness: ‘✓’ evidence available to demonstrate that PbQoL measure and the reference measure were consistent; ‘o’ weak evidence available but uncertain; or the PbQoL measure and the reference measure were not always consistent; ‘’ the PbQoL measure and the reference measure were inconsistent.  
+ Hypothesis testing if the difference in change overtime between the intervention and the control group equals to zero  
* Score at either baseline or endpoint, instead of change overtime  
+ Hypothesis testing if the change within group overtime equals to zero
### Table 4. Measures used in the included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>PbQoL instruments</th>
<th>Non-preference based QoL instruments</th>
<th>Commonly used clinical measures of Parkinson's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EQ-5D</td>
<td>EQ-VAS</td>
<td>HUI-3</td>
</tr>
<tr>
<td>Studies for assessment of discriminant and convergent validity (n=10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benito-Leon et al. 2012 [54]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Garcia-Gordillo et al. 2013 [55]</td>
<td>✓ b</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Jones et al. 2009 [56]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Luo et al. 2009 [57]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Martinez-Martin et al. 2014 [58]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pohar et al. 2009 [59]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Rodriguez-Blazquez et al. 2010 [60]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Rodriguez-Blazquez et al. 2013 [61]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Siderowf et al. 2002 [62]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Swinn et al. 2003 [63]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Studies for assessment of responsiveness (n=13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daley et al. 2014 [64]</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ebersbach et al. 2010 [65]</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jarman et al. 2002 [66]</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larisch et al. 2011 [67]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Luo et al. 2010 [68]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Noyes et al. 2006 [68]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nyholm et al. 2005 [70]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Reuther et al. 2007 [71]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Schröder et al. 2012 [72]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stocchi et al. 2011 [73]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trend et al. [74]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Wade et al. 2003 [75]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zhu et al. 2014 [76]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

EQ-5D EuroQol EQ-5D-3L. EQ-VAS EuroQol Visual Analogue Scale. HUI-3 Health Utilities Index—Mark 3. HUI-2 Health Utilities Index—Mark 2. 15D 15 Dimensions. DDI Disability and Distress Index. PDQ-39 Parkinson’s Disease Questionnaire-39-item. PDQ-8 Parkinson’s Disease Questionnaire-8-item. PDQUALIF Parkinson’s Disease Quality of Life scale. SF-36 Short-Form 36-item. PDQL Parkinson’s Disease Quality of Life questionnaire. H&Y Hoehn and Yahr scale. UPDRS Unified Parkinson’s Disease Rating Scale.

* Movement disorder society - UPDRS

EQ-5D-5L
Table 5 Characteristics of the health-related QoL instruments in the included studies

<table>
<thead>
<tr>
<th>Name</th>
<th>Generic or Parkinson’s specific</th>
<th>Possible score range</th>
<th>Dimensions (D) / attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PbQoL measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EuroQoL EQ-5D [6]</td>
<td>Generic</td>
<td>-0.594 (worst) ~ 1 (full health)</td>
<td>5D: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression</td>
</tr>
<tr>
<td>HUI-2 (Health Utilities Index – Mark 2) [9]</td>
<td>Generic</td>
<td>-0.03 (worst) ~ 1 (full health)</td>
<td>6D: sensation, mobility, emotion, cognition, self-care, and pain</td>
</tr>
<tr>
<td>HUI-3 (Health Utilities index – Mark 3) [10]</td>
<td>Generic</td>
<td>-0.36 (worst) ~ 1 (full health)</td>
<td>8D: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain</td>
</tr>
<tr>
<td>15D (15 Dimensions) [80]</td>
<td>Generic</td>
<td>0 (being dead) ~ 1 (full health)</td>
<td>15D: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination (bladder and bowel function), usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity.</td>
</tr>
<tr>
<td>DDI (Disability and distress index, or Rosser Index) [78]</td>
<td>Generic</td>
<td>-1.486 (worst) ~ 1.0 (full health)</td>
<td>2D: disability and distress</td>
</tr>
<tr>
<td><strong>Non-preference based QoL measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 (Short-Form 36-item) [85]</td>
<td>Generic</td>
<td>Physical summary: 0 (worst) ~ 400 (full health) Mental summary: 0 (worst) ~ 400 (full health)</td>
<td>8D: physical functioning, role physical, bodily pain, general health perceptions, vitality, role emotional, social role functioning, and mental health</td>
</tr>
<tr>
<td>PDQ-39/8 (Parkinson’s Disease Questionnaire -39/8-item) [27]</td>
<td>Specific</td>
<td>0 (best) -100 (worst)</td>
<td>8D: mobility, activities of daily living, emotions, stigma, social support, cognition, communication, and bodily discomfort</td>
</tr>
<tr>
<td>PDQUALIF (Parkinson’s Disease QUAlity of LIFE scale) [96]</td>
<td>Specific</td>
<td>0 (best) -100 (worst)</td>
<td>7D: social/ role function, self-image/ sexuality/sleep, outlook, physical function, independence, urinary function and one global health-related quality of life item</td>
</tr>
<tr>
<td>PDQL (Parkinson’s Disease Quality of Life questionnaire) [97]</td>
<td>Specific</td>
<td>37 (worst) -185 (best)</td>
<td>4D: Parkinsonian symptoms, systemic symptoms, emotional functioning, and social functioning</td>
</tr>
</tbody>
</table>
Appendix I

Search strategy and No. of results in each database

Search first run: 26 November, 2013

Search update: 9 June 2015 (same search strategy, only limit the date, 01/01/2013- present (9 June 2015))

PUBMED

1st Result: 1196
2nd Result: 314

Search (((parkinsonian disorders[MeSH Terms]) OR parkinson*[Title/Abstract])) AND (((((cost effective*[Title/Abstract]) OR Cost-Benefit Analysis[Mesh]) OR Quality of Well-being[Title/Abstract]) OR Quality of Wellbeing[Title/Abstract]) OR QWB[Title/Abstract]) OR Health Utilities Index[Title/Abstract]) OR cost benefit*[Title/Abstract]) OR visual analogue scale[Title/Abstract]) OR time trade off) OR time tradeoff) OR standard gamble) OR discrete choice) OR dce) OR conjoint analysis) OR contingent valuation) OR preference*[Title/Abstract]) OR utility[Title/Abstract]) OR willingness to pay[Title/Abstract]) OR wtp[Title/Abstract]) OR QALY[Title/Abstract]) OR Quality-Adjusted Life Years[MeSH Terms]) OR QALE[Title/Abstract]) OR QALD[Title/Abstract]) OR Qtime[Title/Abstract]) OR quality adjusted life expectancy[Title/Abstract]) OR DALY[Title/Abstract]) OR Disability adjusted life[Title/Abstract]) OR HYE[Title/Abstract]) OR HYEs[Title/Abstract]) OR healthy year equivalent) OR SF-6D[Title/Abstract]) OR SF6D[Title/Abstract]) OR EuroQOL[Title/Abstract]) OR Euro qol[Title/Abstract]) OR EQ-5D[Title/Abstract]) OR HUI[Title/Abstract]) OR EQ5D[Title/Abstract]) OR HUI1[Title/Abstract]) OR HUI2[Title/Abstract]) OR HUI3[Title/Abstract]) Filters: English

Ovid MEDLINE® In-Process & Other Non-Indexed Citations and Ovid MEDLINE® 1946 to Present with Daily & Weekly Update

1st Result: 1202
2nd Result: 300 (01/01/2013-current)

Search Strategy:

1 exp "cost effective"
2 cost effective*.ab.
3 cost utility*.ab.
4  cost benefit*.ab.
5  cost benefit analysis.sh.
6  visual analogue scale.ab.
7  standard gamble.af.
8  time trade off.af.
9  time tradeoff.mp
10  dce.mp.
11  discrete choice.mp.
12  conjoint analysis.af.
13  willingness to pay.mp.
14  wtp.mp.
15  Patient Preference/ or preference.mp.
16  preference*.ab.
17  contingent valuation.mp.
18  QALY$.mp. or Quality-Adjusted Life Years/
19  QALE$.mp.
20  QALD$.mp.
21  Qtime$.mp.
22  quality adjusted life expectancy.mp.
23  quality adjusted life day$.mp.
24  DALY$.mp.
25  Disability adjusted life.mp.
26  HYE.mp.
27  HYE$.mp.
28  Health$ year$ equivalent$.mp.
29  SF-6D.mp.
30  SF6D.mp.
31  EuroQOL.mp.
32 Euro qol.mp.
33 EQ-5D.mp.
34 EQ5D.mp.
35 HUI.mp.
36 HUI1.mp.
37 HUI2.mp.
38 HUI3.mp.
39 Health Utilities Index.mp.
40 QWB.mp.
41 Quality of Wellbeing.mp.
42 Quality of Well-being.mp.
43 utilit$.mutiltp.
44 or/1-43
45 parkinson*.ab.
46 parkinsonian disorders.sh.
47 45 or 46
48 44 and 47
49 limit 48 to english language

Embase
1947 – Present, updated daily
1st Result: 1516
2nd Result: 553 (01/01/2013-current)

1. Parkinsonian Disorders/
2. parkinson*.ab.
3. Parkinsonian Disorders.sh.
4. 1 or 2 or 3  
5. cost effective*.ab.  
6. cost utility*.ab.  
7. cost benefit*.ab.  
8. cost benefit analysis.sh.  
9. visual analogue scale.ab.  
10. time trade off.af.  
11. standard gamble.af.  
12. Patient Preference/ or preference.mp.  
13. preference*.ab.  
14. discrete choice.mp.  
15. conjoint analysis.af.  
16. utilit$.mp.  
17. willingness to pay.mp.  
18. wtp.mp.  
19. dce.mp.  
20. contingent valuation.mp.  
21. QALY$.mp. or Quality-Adjusted Life Years/  
22. QALE$.mp.  
23. QALD$.mp.  
24. Qtime$.mp.  
25. quality adjusted life expectancy.mp.  
26. quality adjusted life day$.mp.  
27. DALY$.mp.  
29. HYE.mp.  
30. HYEs.mp.  
31. Health$ year$ equivalent$.mp.
32. SF-6D.mp.
33. SF6D.mp.
34. EuroQOL.mp.
35. Euro qol.mp.
36. EQ-5D.mp.
37. EQ5D.mp.
38. HUI.mp.
39. HUI1.mp.
40. HUI2.mp.
41. HUI3.mp.
42. Health Utilities Index.mp.
43. QWB.mp.
44. Quality of Wellbeing.mp.
45. Quality of Well-being.mp.
46. time tradeoff.mp.
47. exp "cost benefit analysis"/ or exp "cost effectiveness analysis"/
48. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47
49. 4 and 48
50. limit 49 to (human and english language)

CINAHL (EBSCO)

1st Result: 102

2nd Result: 25 (01/01/2013-present)

S1 MW parkinsonian disorders
S2 AB parkinson*
S3 S1 OR S2
S4  MW quality adjusted life years
S5  AB cost effective*
S6  AB cost utility
S7  MW cost benefit analysis
S8  AB cost benefit
S9  TX visual analogue scale
S10 TX time trade off
S11 TX standard gamble
S12 AB preference*
S13 TX discrete choice
S14 TX conjoint analysis
S15 TX willingness to pay
S16 TX time tradeoff
S17 TX dce
S18 AB utilit*
S19 TX wtp
S20 TX contingent valuation
S21 (MH "Quality-Adjusted Life Years") OR "QALY"
S22 TX hye
S23 TX hyes
S24 TX qaly
S25 TX qale
S26 TX Quality adjusted life expectancy
S27 TX QALD
S28 TX quality adjusted life days
S29 TX DALY
S30 TX disability adjusted life
S31 TX health* year* equivalent
S32 TX eq5d
S33 TX hui1
S34 TX hui2
S35 TX hui3
S36 TX eq-5d
S37 TX euroqol
S38 TX euro qol
S39 TX sf 6d
S40 TX sf6d
S41 TX health utilit* index
S42 TX qwb
S43 TX quality of wellbeing
S44 TX quality of well being
S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44
S46 S3 AND S45
PsycINFO(EBSCO)

1st Result: 440

2nd Result: 213 (01/01/2013-present)

S1 MM "Parkinsonism"
S2 AB parkinson*
S3 S1 AND S2
S4 MM "Quality of Life" OR MM "Quality of Work Life"
S5 AB cost effective*
S6 AB cost utility
S7 TX visual analogue scale
S8 TX time trade off
S9 TX standard gamble
S10 AB preference*
S11 TX discrete choice
S12 TX conjoint analysis
S13 TX willingness to pay
S14 TX time tradeoff
S15 TX dce
S16 TX wtp
S17 TX contingent valuation
S18 QALY
S19 TX qaly
S20 TX disability adjusted life
S21 TX quality adjusted life year*
S22 TX eq5d
S23 TX EQ-5D
S24 TX hui1
S25 TX hui2
S26 TX hui3
S27 TX euroqol
S28 TX utilit*
S29 TX sf 6d
S30 TX sf6d
S31 S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30
S32 S1 OR S2
S33 S31 AND S32
Applied Social Sciences Index and Abstracts (ASSIA) (Proquest)

1st Result: 30
2nd Result: 6 (01/01/2013-present)

(ab(parkinson*) OR su(parkinsonian disorders)) AND (su(cost effective*) OR ab(cost effective*) OR ab(cost utility) OR ab(cost benefit) OR su(cost - benefit analysis) OR (standard gamble) OR (time trade off) OR (visual analogue scale) OR (discrete choice) OR ab(preference*) OR (conjoint analysis) OR (willingness to pay) OR (contingent valuation) OR (time tradeoff) OR (dce) OR (wtp) OR su(Quality-Adjusted Life Years) OR (QALY) OR (QALE) OR (QALD) OR (Qtime) OR (quality adjusted life expectancy) OR (quality adjusted life day*) OR (DALY) OR (disability adjusted life) OR (hye) OR (hyes) OR (health* year* equivalent*) OR (sf6d) OR (sf 6d) OR (euroqol) OR (euro qol) OR (eq 5d) OR (eq5d) OR (hui) OR (hui1) OR (hui2) OR (hui3) OR (health utilities index) OR (qwb) OR (quality of wellbeing) OR (quality of well being) OR ab(utility*))

SOCIAL service abstract (SSA) (Proquest)

1st Result: 2
2nd Result: 1 (01/01/2013-present)

(ab(parkinson*) OR su(parkinsonian disorders)) AND (su(cost effective*) OR ab(cost effective*) OR ab(cost utility) OR ab(cost benefit) OR su(cost - benefit analysis) OR (standard gamble) OR (time trade off) OR (visual analogue scale) OR (discrete choice) OR ab(preference*) OR (conjoint analysis) OR (willingness to pay) OR (contingent valuation) OR (time tradeoff) OR (dce) OR (wtp) OR su(Quality-Adjusted Life Years) OR (QALY) OR (QALE) OR (QALD) OR (Qtime) OR (quality adjusted life expectancy) OR (quality adjusted life day*) OR (DALY) OR (disability adjusted life) OR (hye) OR (hyes) OR (health* year* equivalent*) OR (sf6d) OR (sf 6d) OR (euroqol) OR (euro qol) OR (eq 5d) OR (eq5d) OR (hui) OR (hui1) OR (hui2) OR (hui3) OR (health utilities index) OR (qwb) OR (quality of wellbeing) OR (quality of well being) OR ab(utility*))

AgelInfo open search

1st Result: 15
2nd Result: 1 (01/01/2013-present)

Parkinson* and quality of life
Database of Abstracts of Reviews of Effects (DARE) (CRD York)

1st Result: 51

2nd Result: 8 (01/01/2013-present)

(Parkinson*) AND (quality of life) OR (utility*) IN DARE

NHS Economic evaluation database (NHS EED)

1st Result: 26

2nd Result: 2 (01/01/2013-present)

(Parkinson*) AND (quality of life) OR (utility*) IN NHSEED