Assessing Health Related Quality of life and Distress in People with Parkinson's Disease

Bronagh Reynolds, MSc

Submitted in partial fulfilment of the requirements of the degree of Doctorate in Clinical Psychology

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

July 2020

Acknowledgements

Firstly, I would like to give my sincerest thanks to all the people who took part in my study. I am grateful you took the time to participate and make this project possible.

A very big thank you to Breda Cullen, your expertise and support over the past three years has been invaluable. The level of support you have given me has helped make this process as easy and enjoyable as it could be, so I really thank you for that.

I also want to thank Dr Watt for generously allowing us to recruit from the Movement Disorder Clinic. Thank you also to Susan O'Connell and Sharon Mulhern, you made sure this project got up and running and were always on hand to help. To Nick and Paula, thank you for your enthusiasm and dedication in supporting recruitment for this study.

Very special thanks to my mam and dad, your love and support means everything to me, I love you both. Mam, you have definitely earned this one! And thank you to the rest of my family and friends for the laughs and distractions from everything psychology related, I can't wait to celebrate and see you all again very soon!

Finally, to Fearghal, you are the kindest person I know. Thank you for being there every step of the way and always having confidence in me especially when I didn't. I am so happy to be moving on to the next chapter with you!

Foreword

Due to the outbreak of COVID-19, all recruitment for research projects in NHS Ayrshire and Arran was suspended on 16th March 2020. By this time, a portion of the required sample for the major research project in this thesis had already been recruited. The University later confirmed that projects could be presented for examination in which some usable data had been collected, but the nature or amount of the data differs from what was planned in the proposal. This major research project met these criteria and is therefore presented in this thesis.

Table of Contents

List of Figures

List of Tables

Chapter 1: Systematic Review

Chapter 2: Major Research Project

Chapter 1: Systematic Review

A Systematic Review of Health-Related Quality of Life Measurements for Parkinson's Disease across Cultures

Bronagh Reynolds, MSc,[1,2] Emma Mawdsley, MSc[1] and Breda Cullen, PhD[1]

[1] Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow.

[2] NHS Ayrshire and Arran

Correspondence to: Institute of Health and Wellbeing, College of Medical, Veterinary and Life Science, University of Glasgow, 1st floor, Administration Building, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow, G12 0XH

Bronagh Reynolds

Prepared in accordance with submission requirement for the Movement Disorder Journal (see Appendix A, page 64).

Word count including abstract references: 6,888

Abstract

Background: Health-related quality of life (HRQoL) is a key construct to assess in people with Parkinson's disease. Many HRQoL measures have been translated for use in different countries. Despite the proliferation in studies translating quality of life measures, there is no systematic review evaluating the psychometric properties of these translated measures.

Objective: The primary aim of this review was to evaluate the psychometric properties of nine HRQoL measures recommended for use in people with Parkinson's disease after they have been translated into a different language.

Method: Four databases were systematically searched. Studies which involved validation of one of the nine measures in a different language and/or adaptation of the measure were included for review. The Consensus-based Standards for Measurement Instruments guidance was used for data extraction and for rating risk of bias.

Results: After databases were searched, 5,451 papers were found and after screening a total of 33 papers were included for review representing six out of the nine measures. Internal consistency and construct validity were the measurement properties most commonly reported by the studies reviewed. Issues regarding unidimensionality of measures were noted along with limited transparency of the translation process.

Conclusion: Of the six measures reviewed, the PDQ-39 shows promise in Chinese and Filipino, along with the SCOPA-PS in French and potentially the SCOPA-PS in other languages reviewed, if evidence for unidimensionality is clarified and reliability is further evidenced.

Keywords: Parkinson's Disease; health-related quality of life; translation; validation; reliability

Introduction

Parkinson's disease (PD) is a neurodegenerative disease affecting both motor and non-motor functioning with symptoms fluctuating and progressing over time, causing significant changes in health-related quality of life (HRQoL). PD symptoms fluctuate and change over time meaning that it is important to be able to accurately assess and track HRQoL over time. For the purposes of this review, HRQoL is the subjective judgement of "the physical, emotional and social wellbeing and satisfaction related to health" (Martinez-Martin, Jeukens-Visser, Lyons, Rodriguez-Blazquez, et al., 2011, p. 2372).

In 2011, a Movement Disorder Society (MDS) Taskforce was commissioned to conduct a review of measures used to assess HRQoL in Parkinson's disease (Martinez-Martin, Jeukens-Visser, Lyons, Rodriguez-Blazquez, et al., 2011). A total of 17 measures were reviewed and graded according to predefined criteria and catalogued as either 'recommended', 'suggested' or 'listed'. Nine patient reported outcome measures (PROM) were graded as being 'recommended' for use.

These measures are therefore the most likely ones to be employed in different aspects of research and clinical care involving people with PD worldwide. The importance of appropriately translating and validating HRQoL measures to the applicable culture has been noted in the literature along with suggestions on how this may be done effectively. There are several key reasons why this is important for people with PD. These include areas of health care research and economics as well as policy and local clinical practice.

Firstly, research into potentially efficacious treatments is ongoing in PD. An intervention may be deemed beneficial if it leads to symptom reduction; however, it may be a distressing or unpleasant experience for the individual or result in side effects which greatly limit their independence. The subjective nature of HRQoL needs to be accurately captured to give a balanced picture of the impact of an intervention (NICE, 2017). The emergence of promising treatments will likely encourage larger scale international clinical trials encompassing different languages and cultures, making it vital that measures validated in these countries have comparable psychometric properties to the original measure.

Secondly, regarding health care economics, HRQoL instruments play a pivotal role, both in the UK and countries such including Germany and Spain, (Rios-Diaz et al., 2016) in

determining resource allocation for treatments. HRQoL, along with mortality rates, are combined to form quality adjusted life years (QALY) (Whitehead & Ali, 2010). The calculation of QALYs is central to ascertain the cost effectiveness of interventions, such as deep brain stimulation treatment for PD, and therefore inform health care policy and guidance.

Thirdly, at a more local, individual level, adaptation of HRQoL instruments is relevant as it is recognised that there is an increase in multiculturalism and linguistic diversity across Europe (Chriost & Thomas, 2008). Therefore, healthcare professionals require access to appropriate instruments to assess HRQoL, in the language used by the person to communicate. This is a crucial part of ensuring ethical and equitable practice as one aspect of HRQoL, psychological distress, is known to be commonly under-reported in this population (Chen & Marsh, 2014). Assessing distress in this way also allows clinicians to evidence their decision-making processes and informs the consideration of appropriate care pathways.

The aim of this systematic review was to examine the psychometric properties of MDS-recommended HRQoL measures for PD after they have been translated into a language other than that in which they were originally developed. A secondary objective was to examine whether any adaptations beyond translation into native language (e.g. alteration/removal of items) have taken place, as this has implications for the validity and reliability of the measure.

Methods

Review Framework and Protocol

This systematic review was conducted and reported in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidance (Mokkink, Prinsen, et al., 2018) and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (Shamseer et al., 2015). COSMIN are an international initiative who aim to improve the selection of health measurement instruments for research and in clinical practice (Mokkink, Prinsen, et al., 2018).

The guidance offered in its most recent manual aids in structuring and conducting systematic reviews on PROMs (see Appendix B, page 67). A modified approach was used for this review with not all steps outlined deemed necessary for the purposes of this review (see

Appendix C, page 69). Eight of the ten measurement properties listed by COSMIN were extracted for this review (structural validity; internal consistency; reliability; cross-cultural validity; responsiveness; measurement error; criterion validity and hypothesis testing for construct validity). For additional details on modifications made to COSMIN guidance see Appendix D, page 70.The protocol for this systematic review was registered on PROSPERO (CRD42019139161) and can be accessed at: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=139161

Eligibility Criteria

The nature of this review does not fit within widely used frameworks such as PICOS, but the criteria were developed systematically in consultation with the research supervisor and a specialist librarian. The following inclusion criteria were used to identify relevant articles: study population was people with the most commonly reported type of PD, idiopathic Parkinson's Disease (IPD); study focused on one of the nine MDS-recommended HRQoL measures; study involved translation of the measure; study involved validation of the measure.

Also, if the HRQoL measure used in a relevant study was translated and adapted in some form, the study was eligible to address the secondary review objective. If papers did not state whether PD participants had a specific diagnosis of IPD, this was assumed. The comparator used was the psychometric properties as set out by COSMIN. Only peer reviewed journal articles were included. Non-English language articles were eligible, but the title and abstract must be available in English, for screening purposes.

Information Sources and Search Strategy

Searches were performed in MEDLINE, Web of Science, CINAHL and PsycINFO databases, from inception up to October 19th, 2019. A combination of the following search terms was used: (Parkinson* disease questionnaire-39 or PDQ-39 or Parkinson* Disease Questionnaire-8 or PDQ-8 or Parkinson* Impact Scale or PIMS or Medical Outcomes study-short form 36 or SF-36 or scales of outcomes in Parkinson* disease psychosocial questionnaire or scopa-ps or euroqol or eq-5d or sickness impact profile or sip or nottingham health profile or nhp or parkinson* disease quality of life questionnaire or pdql) AND (cross-cultural or translat* or valid* or reliability or quality of life or health status) AND (Parkinson's Disease). See Appendix E on page 72 for a screenshot of search terms. To ensure literature saturation, the

'cited by' function available only in the Web of Science database was used. This found every article that has cited the original article in which each of the nine instruments was first published.

Study Selection

A two-phase screening approach was used. In phase 1, titles and abstracts were screened against the inclusion criteria by the first author (BR) with a second reviewer (EM) independently screening 50 results. In phase 2, articles that appeared relevant were obtained and read in full by the first author. EM independently reviewed 10 full text papers. Disagreements were resolved by discussion until consensus was reached.

Risk of Bias Assessment

The COSMIN Risk of Bias Checklist (Mokkink, de Vet, et al., 2018) was used in this review. The risk of bias for each of the eight measurement properties was assessed separately for each study by BR with three papers reviewed independently by EM, and consensus then reached by discussion. The risk of bias in relation to that measurement property was then rated either 'very good', 'adequate', 'doubtful', 'inadequate' or 'not applicable' according to predetermined criteria set out in the checklist, and using a 'worst counts score' approach to overall rating for each property. Details of this procedure are given in Appendix F page 73.

Data Extraction and Synthesis

Extraction of data from included papers was conducted by the primary reviewer and replicated for three papers by the secondary reviewer. Discrepancies were resolved through discussion until consensus was reached. The eight relevant measurement properties outlined by COSMIN (Mokkink, Prinsen, et al., 2018) were used to structure data extraction and synthesis for each study separately, with a rating of (+) sufficient, (-) insufficient or (?) indeterminable given. This rating convention is based on the one set out by Terwee et al (2007) and Prinsen et al (2016). In general, a sufficient (+) rating is achieved when the figure is greater than or equal to .70 which is the case for internal consistency, reliability and criterion validity. Template tables provided by COSMIN were used for this purpose. For the measurement property (hypotheses testing for construct validity), the following set of *Apriori* hypotheses (suggested by COSMIN) were formulated against which to evaluate the results of the studies. Correlations between measures evaluating similar constructs were expected to be ≥.50. Correlations were expected to be ≥.30 with measures evaluating related but dissimilar

constructs (e.g. specifically motor aspects of PD only). Correlations were expected to be <.30 with instruments measuring unrelated constructs. Number of participants and any adaptations to the content of the instrument were also extracted. Due to the expected heterogeneity amongst studies, it was decided *a priori* that data would be summarised qualitatively, with no meta-analysis conducted.

Results

A total of 42 studies were deemed eligible for inclusion; see PRISMA-P flow chart in Figure 1 below. Four studies conducted evaluations of multiple measures in multiple languages. This meant that the total number of data points would be 49. Given the time scale required for completion of the review, it would not have been possible to review each data point to the standard required. In addition, the review by Martinez-Martin et al (2011) posited their nine 'recommended' measures, so there was a high likelihood that these measures would be more frequently studied and employed in relevant translation studies after, rather than before this date. As such, studies conducted in the past 15 years were included which meant (2004-2019 inclusive) that no relevant measure was missed, and no language was missed. Several of the studies which were not included were replicated/subsumed by future work which was included e.g. Spanish version of the PDQ-39 was conducted in 1999 but this study was replicated again in 2005, and so is included in this review.

Therefore, studies conducted in the past 15 years (2004-2019 inclusive) were therefore included. This resulted in 33 studies being included, representing six of the nine instruments of interest (PDQ-39; PDQ-8; PIMS; EQ-5D; SCOPA-PS; PDQL); see Appendix G, page 76 for further details on each measure. The other three (NHP; SIP; SF-36) were not studied in any eligible paper from the past 15 years or earlier.

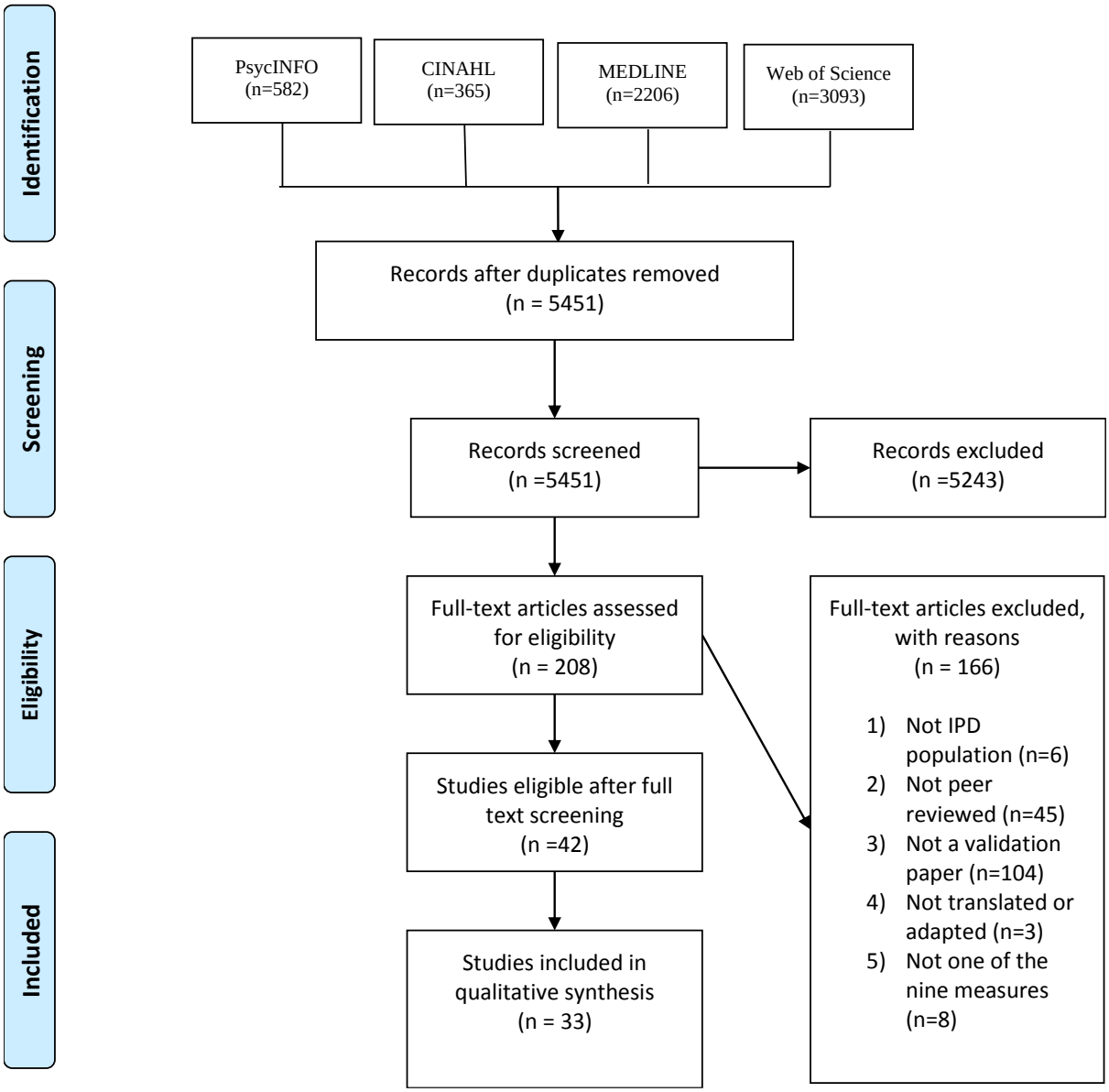


Figure 1.1: PRISMA flowchart displaying numbers for each stage of the review process. IPD = Idiopathic Parkinson's Disease.

There was a high level of agreement between reviewers in both phase 1 (98%) and 2 (97.5%). Four of the included studies conducted multiple evaluations of an instrument/language which resulted in a total of 40 sets of results which were reviewed. Fifteen sets of results were reviewed for the PDQ-39; nine for the PDQ-8; nine for the SCOPA-PS; four for the PDQL, two for the PIMS and one for the EQ-5D. A total of 17 languages were evaluated across the

six instruments. Spanish, Chinese and Portuguese were the most common language translations. For Spanish and Chinese languages, dialects varied across studies e.g. Ecuadorian Spanish, Singaporean Chinese or mainland Chinese. A breakdown of each instrument by language is given in Appendix H page 78.

Tables 1.1 and 2.1 below present the psychometric properties of all the instruments including risk of bias ratings for each. There was heterogeneity across studies regarding which measurement properties were reliably assessed and how these were subsequently reported; however, none of the studies evaluated cross-cultural validity or responsiveness and so they are not represented in the tables below. It was noted that construct validity (40 results) and internal consistency (39 results) were the most assessed, followed by reliability (20 results) and structural validity (16 results). Data extraction accuracy of reviewers BR and EM was high (97% agreement). Interpretation of the below results requires examination of both the risk of bias ratings and the related figures, or each measurement property assessed. Ideally, the risk of bias rating would be very good (V) or adequate (A) indicating a relatively low risk of bias for that property. Risk of bias ratings need to be considered alongside the measurement property figure as a low risk of bias may exist alongside a figure deemed insufficient (-) according to COSMIN. An example of this is Carod-Artal, Martinez-Martin, & Vargas (2007) who translated the PDQ-39 into Portuguese. They found that the reliability of the measure had a low risk of bias for reliability as well as an insufficient rating for the reliability figure. This indicates that the reliability of the PDQ-39 in this study was below the recommended level, and that this result is likely to be trustworthy as the methods showed low risk of bias. Overall, there is a pattern of a relatively high risk of bias throughout the measures presented in tables 1.1. and 1.2 below, with typical ratings of doubtful (D) and inadequate (I) being given for figures. This is commented on below in the risk of bias section. To interpret the results of this review, then, particular attention should be paid to the ratings of sufficient (+), insufficient (-) or indeterminable (?), as well as risk of bias ratings, as those figures which meet sufficient criteria, typically have a relatively high risk of bias in this review. In addition, some languages are represented more than others, and so this allows for sufficient ratings to be replicated across studies in the same language. This should also be taken into account when interpreting the results. For these reasons, the PDQ-39 in Chinese which was reviewed four times, shows promise, although sufficient results display a generally high risk of bias which is typical across all measures reviewed. Although in the

Chinese language, different dialects were used and thus readers interested in specific region of China may wish to pay special attention to the individual findings.

The PDQ-39 in Filipino shows promise as it is one of the only studies to find the internal consistency of both the subscales, and overall summary index figures to be sufficient. This is of interest and in contrast to, most of the other results for internal consistency and the PDQ-39, as the subscales were either insufficient, or not reported in favour of the summary index score.

The SCOPA-PS is not the most widely researched measure but when translated into French, many of the relevant measurement properties were examined and found to be sufficient. This needs to be interpreted in the context of a high risk of bias for the measurement properties. Given the heterogeneity of the properties reported and the generally high risk of bias for these, conclusions about which measures are the most valid and reliable are not straightforward, and so additional interpretations than the one made above are also possible. The presentation of the results in table format allows the reader to examine the results independently and facilitates closer inspection of a measure or language of interest as desired.

Structural validity

Structural validity of the PDQ-39, PDQ-8 and SCOPA-PS was assessed. Only two papers calculated a statistic deemed appropriate by COSMIN (20 & 32 in Table 1), for the PDQ-8 and SCOPA-PS respectively, although neither met the 'sufficient' criteria. There was heterogeneity across studies regarding whether a one or multiple factor model was appropriate, e.g. for the SCOPA-PS a one factor model was found by Soulas et al. (2016) but a two-factor model posited by (Martinez-Martin et al., 2009) (32 & 40 in Table 1).

Internal consistency, reliability and measurement error

Variations in reporting of subscale and summary index results were noted for both internal consistency and reliability measurement properties, with only six studies reporting on both the subscale and summary index. Of these, only two studies found 'sufficient' internal consistency for both subscales and the summary index, for the Chinese and Filipino versions of the PDQ-39 (4 and 14 in Table 1). Only two results were available for measurement error (32 & 40 in Table 1), for the SCOPA-PS French and Spanish versions, although the methods used were deemed inadequate according to COSMIN as the standard error of the mean was calculated from one time point only.

Criterion validity and construct validity

Three studies assessed criterion validity with all meeting ‘adequate’ criteria. Results for convergent validity were mixed. Overall, there was a pattern of correlations above .50 for measures which were hypothesised to be related although this varied between subscales and summary index results.

For discriminant validity, the most used comparator was the Hoehn and Yahr scale, based on the hypothesis that groups at different disease stages would differ in their HRQoL ratings. However, some studies reported p values from between-group hypothesis tests, without interpreting the magnitude or direction of differences between groups. This resulted in an ‘inadequate’ rating according to COSMIN criteria.

Risk of bias

For many studies, several factors were determined to result in a high risk of bias: small sample size for structural validity analysis; inappropriate time interval for test re-test reliability, and evidence of multidimensionality of measures. For more details see Appendix I page 86 . Risk of bias rating concordance for reviewers BR and EM was high with 94.4% agreement.

Adaptations to content

Although in many studies cultural adaptions were stated as having been made, the details of these changes were not reported. In their translation of the PDQ-39 into Estonian, Krikmann, Taba, Lai, and Asser (2008) changed the distance conversion system so that metric units were used in the translated version. This was the only change made that was explicitly detailed by any study included in this review.

Table 1.1: Psychometric Properties of the Instruments (Structural Validity; Internal Consistency; Reliability; Measurement Error).

	Author (Year)	PROM	Country (language in which the questionnaire was evaluated)	Structural validity			Internal consistency				Reliability				Measurement Error		
				No.	RoB	Result (Rating)	No.	RoB	Subscale rating	SI rating	No.	RoB	Subscale rating	SI rating	No.	RoB	Rating
1.	(N. Luo, Low, Lau, Au, & Tan, 2009)	EQ-5D	Singapore(Chinese)		N	(?)		N	(?)	(?)		N	(?)	(?)		N	(?)
2.	(Martinez-Martin, Serrano-Duenas, & Vaca-Baquero, 2005)	PDQ-39	Ecuador(Ecuadorian Spanish)		N	(?)	137	D	.48-.92(-)	(?)		N	(?)	(?)		N	(?)
3.	(Nojomi, Mostafavian, Shahidi, & Jenkinson, 2010)	PDQ-39	Iran(Persian)		N	(?)	200	D	.60-.91(-)	.93(+)	200	I	.47-.90(-)	.80(+)		N	(?)
4.	(W. Luo et al., 2010)	PDQ-39	China(Chinese)		N	(?)	71	D	.86-.88(+)	.84(+)	71	I	.56-.82(-)	.82(+)		N	(?)
5.	(Žiropađa, Stefanova, Potrebić, & Kostić, 2009)	PDQ-39	Serbia(Serbian)	102	I	PCA(?)	102	D	.59-.91(-)	(?)		N	(?)	(?)		N	(?)
6.	(Zhang & Chan, 2012)	PDQ-39	China(Chinese)		N	(?)	126	D	.54-.91(-)	(?)		N	(?)	(?)		N	(?)
7.	(Park, Sohng, & Kim, 2014)	PDQ-39	Korea(Korean)	93	I	EFA(?)	93	D	.70-.97(+)	(?)		N	(?)	(?)		N	(?)
8.	(Krikmann et al., 2008)	PDQ-39	Estonia(Estonian)		N	(?)	137	D	.81-.86(+)	(?)	78	D	(?)	(?)		N	(?)

9.	(N. Luo, Tan, Li, Soh, & Thumboo, 2005)	PDQ-39	Singapore(Chinese)	63	I	PCA(?)	63	D	.64-.90(-)	.74(+)	36	I	.66-.86(-)	(?)		N	(?)
10.	(Ma, Hwang, & Chen-Sea, 2005)	PDQ-39	China(Chinese)		N	(?)	73	D	.58-.96(-)	(?)	22	I	.71-.95(+)	(?)		N	(?)
11.	(Carod-Artal, Martinez-Martin, & Vargas, 2007)	PDQ-39	Brazil(Portuguese)		N	(?)	144	D	.61-.85(-)	(?)	144	V	.52-.80(-)	(?)		N	(?)
12.	(Marinus, Visser, Jenkinson, & Stiggelbout, 2007)	PDQ-39	Holland(Dutch)		I	(?)	177	D	.59-.91(-)	(?)		N	(?)	(?)		N	(?)
13.	(Kwon et al., 2013)	PDQ-39	Korea(Korean)		N	(?)	102	D	.58-.80(-)	(?)	101	I	(?)	(?)		N	(?)
14.	(Suratos, Saranza, Sumalapao, & Jamora, 2018)	PDQ-39	Philippines(Filipino)		N	(?)	30	D	.86-.88(+)	.85(+)		N	(?)	(?)		N	(?)
15.	(Galeoto et al., 2018)	PDQ-39	Italy(Italian)		N	(?)	104	D	.69-.92(-)	(?)	35	I	.85-.96(+)	(?)		N	(?)
16.	(Jesus-Ribeiro, Vieira, Ferreira, Januário, & Freire, 2017)	PDQ-39	Portugal(Portuguese)		N	(?)	100	D	.66-.98(-)	(?)	13	I	.49-.96(-)	(?)		N	(?)
17.	(Katsarou et al., 2004)	PDQ-8	Greece(Greek)		N	(?)	228	D	(?)	.72(+)	91	I	(?)	.72(+)		N	(?)
18.	(Tan, Lau, Au, & Luo, 2007)	PDQ-8	China(Chinese)	79	I	PCA(?)	79	D	(?)	.87(+)	79	N	(?)	(?)		N	(?)

19.	(Kahraman et al., 2018)	PDQ-8	Turkey(Turkish)		N	(?)	83	D	(?)	.78(+)	24	I	(?)	.97(+)		N	(?)
20.	(Huang, Hsu, Wang, & Chen, 2011)	PDQ-8	Taiwan(Chinese)	100	V	CFI .95(+) RMSEA .08(-)	100	D	(?)	.81(+)		N	(?)	(?)		N	(?)
21.	(Franchignoni, Giordano, & Ferriero, 2008)	PDQ-8 (data pooled with PDQ 8/39)	Italy(Italian)	100	D	(?)		D	(?)	.72(+)		N	(?)	(?)		N	(?)
22.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	Japan(Japanese)	96	A	PCA(?)	96	D	(?)	.87(+)		N	(?)	(?)		N	(?)
23.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	Italy(Italian)	195	A	PCA(?)	195	D	(?)	.79(+)		N	(?)	(?)		N	(?)
24.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	Spain(Spanish)	194	A	PCA(?)	194	D	(?)	.73(+)		N	(?)	(?)		N	(?)
25.	(K. Chen et al., 2017)	PDQ-8/39	China(Chinese)		A	PCA(?)		D	(?)	.80(+)	283 Baseline	D	(?)	.96(+)		N	(?)
											101 Year one	D	(?)	.96(+)			
											81 Year two	D	(?)	.98(+)			
26.	(Dereli et al., 2015)	PDQL	Turkey(Turkish)	89	N	(?)		D	(?)	.97(+)	46/51	I	(?)	.82(+)		N	(?)
27.	(Campos et al., 2011)	PDQL	Brazil(Portuguese)		N		52	D	.65-.83(-)	.93(+)	21	I	(?)	(?)		N	(?)

28.	(Serrano-Dueñas, Martínez-Martín, & Vaca-Baquero, 2004)	PDQL	Spain(Ecuadorian Spanish)		N		137	D	.69-.85(-)			N	(?)	(?)		N	(?)
29.	(Jesus-Ribeiro et al., 2017)	PDQL	Portugal(Portuguese)		N	(?)	100	D	.78-.98(+)	(?)	13	I	.65-.96(-)	(?)		N	(?)
30.	(Serrano-Dueñas & Serrano, 2007)	PIMS	Ecuador(Ecuadorian Spanish)		N	(?)	131	D	.68-.87(-)	.88(+)	131	A	.91-.97(+)	.98(+)		N	(?)
31.	(Todorova & Stambolieva, 2007)	PIMS	Bulgaria(Bulgarian)		N	(?)	40	D	(?)	.82(+)	40	I	(?)	.71(+)		N	(?)
32.	(Soulas et al., 2016)	SCOPA-PS	France(French)	154	D	CFI.98(+) RMSEA .09(-)	73	D	(?)	.86(+)	73	I	(?)	.88(+)		I	SEM calculated using single time point(?)
33.	(Fereshtehnejad et al., 2014)	SCOPA-PS	Iran(Persian)		N	(?)	110	D	(?)	.87(+)	110	N	(?)	(?)		N	(?)
34.	(Carod-Artal et al., 2007)	SCOPA-PS	Brazil(Portuguese)		N	(?)	144	D	(?)	.84(+)	144	V	(?)	.71(+)		N	(?)
35.	(Virués-Ortega et al., 2009)	SCOPA-PS	Argentina(Local Spanish)		D	(?)	61	D	(?)	.92(+)		N	(?)	(?)		N	(?)
36.	(Virués-Ortega et al., 2009)	SCOPA-PS	Brazil(Portuguese)		D	(?)	127	D	(?)	.86(+)		N	(?)	(?)		N	(?)
37.	(Virués-Ortega et al., 2009)	SCOPA-PS	Ecuador(Local Spanish)		D	(?)	75	D	(?)	.87(+)		N	(?)	(?)		N	(?)

38.	(Virués-Ortega et al., 2009)	SCOPA-PS	Paraguay(Guarani)		D	(?)	68	D	(?)	.86(+)		N	(?)	(?)		N	(?)
39.	(Knudsen, Martinez-Martin, & Deuschl, 2007)	SCOPA-PS	Germany(German)		N	(?)	105	D	(?)	.90(+)	54	A	(?)	.77(+)		N	(?)
40.	(Martinez-Martin et al., 2009)	SCOPA-PS	Spain(Spanish)	387	A	EFA(?)	387	D	(?)	.85(+)		N	(?)	(?)	387	I	SEM calculated(single time point)(?)

CFI = Comparative fit index; EFA = Exploratory factor analysis; PCA = Principal components analysis; RMSEA = Root mean square error of approximation; SEM = Standard Error of the Mean; RoB = Risk of bias rating; V = Very good; A = Adequate; D = Doubtful; I = Inadequate; N = Not applicable; SI = summary index; (+) = Sufficient; (-)= insufficient; (?) = Indeterminable.

Table 1.2: Psychometric Properties of Instruments Continued (Construct Validity and Criterion Validity).

	Author (Year)	PROM	Hypothesis testing for construct validity									Criterion validity
			No.	RoB Convergent validity	RoB Discriminant validity	Results (Rating)				No.	RoB	Result (Rating)
1.	(N. Luo et al., 2009)	EQ-5D	71	V	A	EQ-5D subscales and H&Y -.05-.43(-) SI -.28(-) EQ-VAS -.37(+)	UPDRS Motor -.07 to .22 (-) SI.04(-) EQ-VAS.19(-)	Schwab and England activities of daily living -.10 to -.60(-) SI .36(+) EQ-VAS .27(-)	Chinese MMSE .03 to -.20(-) SI -.01(-) EQ-VAS .11(-)		N	(?)
2.	(Martinez-Martin et al., 2005)	PDQ-39	137	V	A	PDQL EV SI .91 (+)	H&Y stage& PDQ-39 SI .60 (+) H&Y stage PDQ-39 subscales - .22 -.74(-)	HADS – A SI .53(+) HADS-A and subscales .19-.52(-) HADS-D SI .67(+) HADS – D& Subscales .10-.57(-)	UPDRS I .50(+) UPDRS II .71(+) UPDRS III.46(+)		N	(?)
3.	(Nojomi et al., 2010)	PDQ-39	200	V	A	SF-36 -.55 to -.61(+)					N	(?)
4.	(W. Luo et al., 2010)	PDQ-39	71	V	A	SF-36 -.46 to -.69(-)					N	(?)
5.	(Žiropađa et al., 2009)	PDQ-39	102	V	A	SF-36 .80-.81(+)					N	(?)
6.	(Zhang & Chan, 2012)	PDQ-39	126	V	A	SF-36 .12-.76(-)					N	(?)
7.	(Park et al., 2014)	PDQ-39	93	I	A	H&Y .04-.88(-) stigma				93	I	.10-65 (-)
8.	(Krikmann et al., 2008)	PDQ-39	81	N	A						N	(?)
9.	(N. Luo et al., 2005)	PDQ-39	63	V	N	EQ-5D .38-.76(-)					N	(?)
10.	(Ma et al., 2005)	PDQ-39	73	V	A	UPDRS .42-.86(+)	*SF-36* -.29 to -.93(-)				N	(?)
11.	(Carod-Artal et al., 2007)	PDQ-39	144	V	A	SF-36 Physical component summary -.26 to -.59(-)	SF-36 Mental component summary -.18 to -.51(-)				N	(?)
12.	(Marinus et al., 2007)	PDQ-39	177	V	D	SCOPA-PS .82 (+)	HADS-A .69(+) HADS-D .65(+)	EQ-5D .63(+) EQ-VAS -.54(+)			N	(?)
13.	(Kwon et al., 2013)	PDQ-39	102	V	N	Korean Montgomery Asberg Depression Scale .47 to .70(-) SI .66(+)	ESS-.24to.43 (-) SI .40(-)	H&Y stage .14-.35(-) PDQ-39 SI .36(+)	NMSS .19-.66 (-)		N	(?)

14.	(Suratos et al., 2018)	PDQ-39	100	V	A	HADS A .32 to .58(-) SI. 61(+) HADS D .32-.48 (-) SI .52(+)	H&Y stage & PDQ-39 subscales .20 -.64(-) PDQ-39 SI .46(+)	UPDRS-I .19-.42(-) SI .40(+) UPDRS II .08 to .56(-) UPDRS III -.14-.50(-) SI .51(+)	PDQ-39 subscales & NMSS.12 to .56 (-) NMSS & PDQ-39 SI . 56(+)		N	(?)
15.	(Galeoto et al., 2018)	PDQ-39	104	V	N	*SF-36* -.50(+)					N	(?)
16.	(Jesus-Ribeiro et al., 2017)	PDQ-39	100	V	A	*SF-36* .03 to -.77(-)					N	(?)
17.	(Katsarou et al., 2004)	PDQ-8	81	V	V	SF-36 Correlations for physical disability -.46(-); pain -.42(-); energy/vitality -.45(-)	BDI .58(+)				V	(?)
18.	(Tan et al., 2007)	PDQ-8	79	V	A	H&Y .29(-)					N	(?)
19.	(Kahraman et al., 2018)	PDQ-8	83	V	D	SF-36 Physical -.52(+)	SF-36 Mental -.64(+)	H&Y stage .56(+)			N	(?)
20.	(Huang et al., 2011)	PDQ-8	100	V	V	H&Y .09 to .59(-) SI .53(+)	Pittsburgh Sleep Quality Index .09-.39(-)	Schwab and England activities of daily living scale -.13 to -.66(-) SI -.65(+)	Taiwanese Depression Questionnaire .32 to 74(+) SI .71(+)		N	(?)
21.	(Franchignoni et al., 2008)	PDQ-8 (data pooled with PDQ 8/39)[3]	100	V	N	IPA-I .47(+)	HY .38(+)	UPDRS-AUL .44(+)			N	
22.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	96	V	N	H&Y stage .47(+)					V	0.92 (+)
23.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	195	V	N	H&Y stage .47(+)				195	V	0.96 (+)
24.	(Jenkinson & Fitzpatrick, 2006)	PDQ-8/39	194	V	N	H&Y Stage .61(+)				194	V	0.93 (+)
25.	(K. Chen et al., 2017)	PDQ-8/39	283 Baseline	V	D	H&Y stage.48(+)	UPDRS-III.47(+)	BDI .64(+)	)	283	V	.93 (+)
			101 Year one			H&Y stage .29(-)	UPDRS-III .34(+)	BDI .69(+)			I	(?)
			81 Year two			H&Y stage .32 (+)	UPDRS-III .43 (+)	BDI .69(+)			N	(?)
26.	(Dereli et al., 2015)	PDQL	89	V	D	H&Y SI -.55 to -.73(+) H&Y -.64(+)	UPDRS 1 -.60(+) UPDRS II -. 64(+) UPDRS III -.64(+) UPDRS total -.73(+)				N	(?)
27.	(Campos et al., 2011)	PDQL	52	V	D	Emotional domain of the PDQL-BR &	Emotional domain of the				N	(?)

						UPDRS I 59(+) UPDRS III .78(+)	PDQL-BR & BDI .57(+)					
28.	(Serrano-Dueñas et al., 2004)	PDQL	137	V	D	HADS-A -.34 to -.70(-) SI-.55(+) HADS-D -.45 to -.74(-), SI -.67(+)	UPDRS .32to.70(+)	*PDQ-39* -0.91(-)	Schwab and England scale .30 to .67 SI .65		N	(?)
29.	(Jesus-Ribeiro et al., 2017)	PDQL	100	V	A	*SF-36* .32 -.70(-)						(?)
30.	(Serrano-Dueñas & Serrano, 2007)	PIMS	131	V	V	*PDQ-39* .80(+)	PDQL -.87(+)	HADS Anxiety .64(+) HADS Depression .76(+)	SCOPA/SPES A.72(+) B.78(+) C.58(+) Total .77(+)		N	(?)
31.	(Todorova & Stambolieva, 2007)	PIMS	40	V	V	*PDQ-39* .40(-)	PDQL .27(-)	UPDRS II .31(+) UPDRS III .44(+)	DV assessed using ANOVA(?)		N	(?)
32.	(Soulas et al., 2016)	SCOPA-PS	154	V	V	*PDQ-39* .35-.73(-) PDQ SI .83(+)	BDI II .69(+)	UPDRS III .49(+)	STAI-Y .56(+)		N	(?)
33.	(Fereshtehnejad et al., 2014)	SCOPA-PS	110	V	N	*PDQ-39* .43-.72(-) PDQ-39 SI .82 (+)	HADS A .64(+) HADS D .72(+)	H&Y.34(+)			N	(?)
34.	(Carod-Artal et al., 2007)	SCOPA-PS	144	V	D	SF-36 physical component summary-.42(-)	SF-36 Mental component summary -.41(-)				N	(?)
35.	(Virués-Ortega et al., 2009)	SCOPA-PS	61	V	D	*PDQ-39* .46 to .73 (-) *PDQ-39 SI* . 85 (+)	SCOPA-Motor .43(-)	CISI-G.52(+)	HADS-A .63(+) HADS-D .62(+)		N	(?)
36.	(Virués-Ortega et al., 2009)	SCOPA-PS	127	V	D	*PDQ-39* .39 to 65(-) SI .80(+)	SCOPA-Motor .51(+)	CISI-G 0.54(+)	HADS-A .62(+) HADS-D .69		N	(?)
37.	(Virués-Ortega et al., 2009)	SCOPA-PS	75	V	D	*PDQ-39* .53 to .74(-) SI .78(+)	SCOPA-Motor .59(+)	CISI-G 0.62(+)	HADS-A .58(+) HADS-D .63(+)		N	(?)
38.	(Virués-Ortega et al., 2009)	SCOPA-PS	68	V	D	*PDQ-39* .17 to .76(-) SI .83(+)	SCOPA-Motor .42(-)	CISI-G 0.52(+)	HADS-A .57(+) HADS-D .58(+)		N	(?)
39.	(Knudsen et al., 2007)	SCOPA-PS	54	V	N	*PDQ-39* .86(+)	EQ-5D -.73(+) EQ-VAS .61 (+)	HADS A .76(+) HADS D .76(+)	CISI-PD .57(+)		N	(?)
40.	(Martinez-Martin et al., 2009)	SCOPA-PS	387	V	V	CISI-G .18(-)	EQ-5D -.61(+)	HADS A .58(+) HADS D .67(+)	SCOPA-AUT .56(+) SCOPA- Motor .52(+)		N	(?)

ANOVA = Analysis of Variance; BDI = Beck's Depression Inventory; CISI-G = Clinical Impression of Severity Index-Geriatric; DV = Discriminant validity

H&Y = Hoehn and Yahr staging scale; HADS = Hospital Anxiety and Depression Scale; IPA-I = Impact on Participation and Autonomy (perceived limitations in participation and autonomy subscale; SCOPA-AUT = SCales for Outcomes in Parkinson's disease - Autonomic Dysfunction**;** SF-36 = The 36-Item Short

Form Health Survey; UPDRS = Unified Parkinson's Disease Rating Scale. RoB = Risk of bias rating; V = Very good; A = Adequate; D = Doubtful; I = Inadequate; N = Not applicable; (+) = Sufficient; (-)= insufficient; (?) = Indeterminable.

Discussion

This review presented the psychometric properties of 33 studies, across 17 languages, which had translated and validated one or more of six MDS-recommended HRQoL instruments in PD. The findings indicate the wide variety in reporting of measurement properties which impacts how well the utility of an instrument can be compared and contrasted across cultures. This is congruent with the variable reporting of measurement properties that was highlighted in a previous systematic review on cross-cultural translations and adaptations of the Multidimensional Perceived Social Support Scale (Dambi et al., 2018).

There was also variability in instruments used as comparators to examine convergent validity, resulting in mixed results for this measurement property across studies. As reported in Table 1.1, the dimensionality of instruments was not routinely assessed despite evidence that all instruments in this review may be multidimensional; see supplementary material in Martinez-Martin, Jeukens-Visser, Lyons, Roiguez-Blazquez, et al. (2011). This issue directly affects internal consistency and reliability as if an instrument is not unidimensional, a summary index result is not useful and subscale scores should be calculated instead. Given the variations in reporting subscale and/or summary index results in this review, adequate internal consistency or reliability cannot be reliably assumed for many of the instruments. This finding may prompt re-evaluation of the dimensionality of these measures, taking into account the concept of HRQoL itself is multidimensional (Lin, Lin, & Fan, 2013) and therefore a multidimensional instrument may be more appropriate.

Regarding the secondary question posited, adaptations to instruments were generally not well defined. Transparency in reporting the translation process is essential, as it has been highlighted that issues such as different cultural interpretations of the word ‘family’ (as perhaps including only first degree or wider for example) may inadvertently lower the validity of the instrument (Sperber, 2004). Despite the translation process being outlined generally by all studies (i.e. forwards and backwards translation and review by experts), the more detailed process of how translation discrepancies were resolved and what these discrepancies were, was not described.

Studies utilised differing translation methods e.g. MAPI Research Trust protocol; consulting relevant literature or simply citing methods used more broadly. This finding is similar to that

which has been reported in the literature e.g. Danielsen et al. (2015) who found that the process of translating instruments measuring HRQoL is not standardised across studies.

Strengths and Limitations

A strength of this review is the wide range of different languages which were included and therefore it allowed for multiple languages to be included and relevant measurement properties extracted. Another strength of the present review is the use of the detailed COSMIN framework to extract relevant measurement properties, conduct risk of bias ratings and synthesise findings. This framework allows for a standardisation of information presented regarding PROMs.

A limitation of the current review is that the date range included was limited to the past 15 years. Including all relevant studies would have resulted in comparisons in risk of bias across early translation i.e. from the 1990s, to present day to be considered. It would also have enhanced the data set so that patterns in the results of certain measures may have been elucidated. Despite this, covering earlier years would not have captured any of the missing three PROMS, and coverage went back several years prior to when the MDS taskforce recommendations were originally published in 2011. A second limitation is that although the topic is adaptations into other languages, the search was performed using English-language phrases and therefore results indexed in non-English language databases may have been missed. Nevertheless, articles were found from a very wide range of journals from around the world which are representative of numerous countries and languages.

Overall, standardisation when assessing and reporting psychometric properties would enhance comparison across instruments for individual clinicians/researchers and future reviews with the COMSIN guidance being a potentially useful framework for this.

Conclusions

This review presents psychometric properties of six HRQoL instruments and shows that they are promising in assessing HRQoL across cultures. Standardised reporting of psychometric properties, statistical methodology and transparent translation processes are needed to better interpret validity and reliability of these instruments. The evidence in this review suggests that the PDQ-39 shows promise in Chinese and Filipino, along with the SCOPA-PS in French and potentially in other languages reviewed, if procedures are put in place to reduce risk of

bias i.e. by following COSMIN guidance, evidence for unidimensionality is clarified and reliability is further evidenced.

References

Campos, M., Rezende, C. H. A. d., Farnese, V. d. C., da Silva, C. H. M., Morales, N. M. d. O., & Pinto, R. d. M. C. (2011). Translation, Cross-Cultural Adaptation, and Validation of the Parkinson's Disease Quality of Life Questionnaire (PDQL), the "PDQL-BR", into Brazilian Portuguese. *ISRN Neurology, 2011*, 954787-954785. doi:10.5402/2011/954787

Carod-Artal, F. J., Martinez-Martin, P., & Vargas, A. P. (2007). Independent validation of SCOPA–psychosocial and metric properties of the PDQ-39 Brazilian version. *Movement Disorders, 22*(1), 91-98. doi:10.1002/mds.21216

Chen, J. J., & Marsh, L. (2014). Anxiety in Parkinson's disease: identification and management. *Therapeutic Advances in Neurological Disorders, 7*(1), 52-59. doi:10.1177/1756285613495723

Chen, K., Yang, Y. J., Liu, F. T., Li, D. K., Bu, L. L., Yang, K., . . . Wu, J. J. (2017). Evaluation of PDQ-8 and its relationship with PDQ-39 in China: a three-year longitudinal study. *Health and Quality of Life Outcomes, 15*(1), 170-177. doi:10.1186/s12955-017-0742-5

Chriost, D. M. G., & Thomas, H. (2008). Linguistic Diversity and the City: Some Reflections, and a Research Agenda. *International Planning Studies, 13*(1), 1-11. doi:10.1080/13563470801969624

Dambi, J. M., Corten, L., Chiwaridzo, M., Jack, H., Mlambo, T., & Jelsma, J. (2018). A systematic review of the psychometric properties of the cross-cultural translations and adaptations of the Multidimensional Perceived Social Support Scale (MSPSS). *Health and Quality of Life Outcomes, 16*(1), 1-19. doi:10.1186/s12955-018-0912-0

Danielsen, A. K., Pommergaard, H. C., Burcharth, J., Angenete, E., Rosenberg, J., Sahlgrenska, a., . . . Sahlgrenska, A. (2015). Translation of Questionnaires Measuring

Health Related Quality of Life Is Not Standardized: A Literature Based Research Study. *PLoS ONE, 10*(5), e0127050. doi:10.1371/journal.pone.0127050

Dereli, E. E., Yaliman, A., Kuru Colaka, T., Cakmak, A., Razak Ozdincler, A., & Badilli Demirbas, S. (2015). Turkish Version Study of "Parkinson's Disease Quality of Life Questionnaire" (PDQL). *Noropsikiyatri Arsivi-Archives of Neuropsychiatry, 52*(2), 128-132. doi:10.5152/npa.2015.7359

Fereshtehnejad, S.-M., Farhadi, F., Hadizadeh, H., Shahidi, G. A., Delbari, A., & Lökk, J. (2014). Cross-cultural validity, reliability, and psychometric properties of the persian version of the scales for outcomes in Parkinson's disease-psychosocial questionnaire. *Neurology Research International, 2014*, 260684. doi:10.1155/2014/260684

Franchignoni, F., Giordano, A., & Ferriero, G. (2008). Rasch Analysis of the Short Form 8-Item Parkinson's Disease Questionnaire (PDQ-8). *Quality of Life Research, 17*(4), 541-548. doi:10.1007/s11136-008-9341-6

Galeoto, G., Colalelli, F., Massai, P., Berardi, A., Tofani, M., Pierantozzi, M., . . . Fabbrini, G. (2018). Quality of life in Parkinson's disease: Italian validation of the Parkinson's Disease Questionnaire (PDQ-39-IT). *Neurological Sciences, 39*(11), 1903-1909. doi:10.1007/s10072-018-3524-x

Huang, T.-T., Hsu, H.-Y., Wang, B.-H., & Chen, K.-H. (2011). Quality of life in Parkinson's disease patients: validation of the Short-Form Eight-item Parkinson's Disease Questionnaire (PDQ-8) in Taiwan. *Quality of Life Research, 20*(4), 499-505. doi:10.1007/s11136-010-9777-3

Jenkinson, C., & Fitzpatrick, R. (2006). Cross-cultural evaluation of the short form 8-item Parkinson's Disease Questionnaire (PDQ-8): Results from America, Canada, Japan, Italy and Spain. *Parkinsonism and Related Disorders, 13*(1), 22-28. doi:10.1016/j.parkreldis.2006.06.006

Jesus-Ribeiro, J., Vieira, E., Ferreira, P., Januário, C., & Freire, A. (2017). Reliability and Validity of 39-Item Parkinson's Disease Questionnaire and Parkinson's Disease Quality of Life Questionnaire. *Acta Medica Portuguesa 30*(5), 395. doi:10.20344/amp.8202

Kahraman, T., Genc, A., Soke, F., Goz, E., Donmez Colakoglu, B., & Keskinoglu, P. (2018). Validity and Reliability of the Turkish Version of the 8-Item Parkinson's Disease Questionnaire. *Noropsikiyatri Arsivi-Archives of Neuropsychiatry, 55*(4), 337-340. doi:10.5152/npa.2017.19343

Katsarou, Z., Bostantjopoulou, S., Peto, V., Kafantari, A., Apostolidou, E., & Peitsidou, E. (2004). Assessing quality of life in Parkinson's disease: Can a short-form questionnaire be useful? *Movement Disorders, 19*(3), 308-312. doi:10.1002/mds.10678

Knudsen, K., Martinez-Martin, P., & Deuschl, G. (2007). Evaluation of the reliability and validity of the German version of a questionnaire on the psychosocial impairment of patients with Parkinson's Disease (SCOPA-PS scales for outcome of Parkinson's Disease, a short psychosocial questionnaire for patients with Parkinson's Disease). *Aktuelle Neurologie, 34*(5), 267-271. doi:10.1055/s-2007-970843

Krikmann, U., Taba, P., Lai, T., & Asser, T. (2008). Validation of an Estonian version of the Parkinson's disease Questionnaire (PDQ-39). *Health and Quality of Life Outcomes, 6*(1), 23-23. doi:10.1186/1477-7525-6-23

Kwon, D. Y., Kim, J. W., Ma, H. I., Ahn, T. B., Cho, J., Lee, P. H., . . . Koh, S. B. (2013). Translation and Validation of the Korean Version of the 39-Item Parkinson's Disease Questionnaire. *Journal of Clinical Neurology, 9*(1), 26-31. doi:10.3988/jcn.2013.9.1.26

Lin, X.-J., Lin, I. M., & Fan, S.-Y. (2013). Methodological issues in measuring health-related quality of life. *Tzu Chi Medical Journal, 25*(1), 8-12. doi:10.1016/j.tcmj.2012.09.002

Luo, N., Low, S., Lau, P. N., Au, W. L., & Tan, L. C. S. (2009). Is EQ-5D a Valid Quality of Life Instrument in Patients With Parkinson's Disease? A Study in Singapore. *Annals Academy of Medicine Singapore, 38*(6), 521-528. Retrieved from https://pdfs.semanticscholar.org/303f/1a81903b6865c88c8fc63273edb1fe9915be.pdf?_ga=2.9492469.1535524533.1593692297-1460634777.1593692297

Luo, N., Tan, L. C. S., Li, S. C., Soh, L. K., & Thumboo, J. (2005). Validity and Reliability of the Chinese (Singapore) Version of the Parkinson's Disease Questionnaire (PDQ-39). *Quality of Life Research, 14*(1), 273-279. doi:10.1007/s11136-004-2654-1

Luo, W., Gui, X. H., Wang, B., Zhang, W. Y., Ouyang, Z. Y., Guo, Y., . . . Ding, M. P. (2010). Validity and reliability testing of the Chinese (mainland) version of the 39-item Parkinson's Disease Questionnaire (PDQ-39). *Journal of Zhejiang University-Science B, 11*(7), 531-538. doi:10.1631/jzus.B0900380

Ma, H.-I., Hwang, W.-J., & Chen-Sea, M.-J. (2005). Reliability and Validity Testing of a Chinese-Translated Version of the 39-Item Parkinson's Disease Questionnaire (PDQ-39). *Quality of Life Research, 14*(2), 565-569. doi:10.1007/s11136-004-0687-0

Marinus, J., Visser, M., Jenkinson, C., & Stiggelbout, A. M. (2007). Evaluation of the Dutch version of the Parkinson's Disease Questionnaire 39. *Parkinsonism and Related Disorders, 14*(1), 24-27. doi:10.1016/j.parkreldis.2007.05.005

Martinez-Martin, P., Carroza-Garcia, E., Frades-Payo, B., Rodriguez-Blazquez, C., Forjaz, M. J., de Pedro-Cuesta, J., . . . Grupo, E. (2009). Psychometric attributes of the sclaes for outcomes in Parkinson's Disease - Psychosocial (SCOPA-PS): Validation in Spain and review. *Revista de Neurologia, 49*(1), 1-7. doi:10.33588/rn.4901.2009014

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. *Mov Disord, 26*(13), 2371-2380. doi:10.1002/mds.23834

Martinez-Martin, P., Jeukens-Visser, M., Lyons, K. E., Roiguez-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

Martinez-Martin, P., Serrano-Duenas, M., & Vaca-Baquero, V. (2005). Psychometric characteristics of the Parkinson's disease questionnaire (PDQ-39)—Ecuadorian version. *Parkinsonism and Related Disorders, 11*(5), 297-304. doi:10.1016/j.parkreldis.2005.02.003

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research, 27*(5), 1171-1179. doi:10.1007/s11136-017-1765-4

Mokkink, L. B., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H., . . . Mokkink, L. (2018). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual, 78*, 1.

Nojomi, M., Mostafavian, Z., Shahidi, G. A., & Jenkinson, C. (2010). Quality of life in patients with Parkinson’s disease: Translation and psychometric evaluation of the Iranian version of PDQ-39. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences, 15*(2), 63-69. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3082793/

Park, H. J., Sohng, K. Y., & Kim, S. (2014). Validation of the Korean version of the 39-Item Parkinson's Disease Questionnaire (PDQ-39). *Asian Nursing Research, 8*(1), 67-74. doi:10.1016/j.anr.2014.02.004

Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select

outcome measurement instruments for outcomes included in a "Core Outcome Set" - a

practical guideline. Trials. 2016;17(1):449.

Rios-Diaz, A. J., Lam, J., Ramos, M. S., Moscoso, A. V., Vaughn, P., Zogg, C. K., & Caterson, E. J. (2016). Global Patterns of QALY and DALY Use in Surgical Cost-Utility Analyses: A Systematic Review. *PLoS ONE, 11*(2), e0148304. doi:10.1371/journal.pone.0148304

Serrano-Dueñas, M., Martínez-Martín, P., & Vaca-Baquero, V. (2004). Validation and cross-cultural adjustment of PDQL-questionnaire, Spanish version (Ecuador) (PDQL-EV). *Parkinsonism and Related Disorders, 10*(7), 433-437. doi:10.1016/j.parkreldis.2004.05.002

Serrano-Dueñas, M., & Serrano, S. (2007). Psychometric characteristics of PIMS—Compared to PDQ-39 and PDQL—To evaluate quality of life in Parkinson's disease patients: Validation in Spanish (Ecuadorian style). *Parkinsonism and Related Disorders, 14*(2), 126-132. doi:10.1016/j.parkreldis.2007.07.006

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . the, P.-P. G. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ : British Medical Journal, 349*(jan02 1), g7647-g7647. doi:10.1136/bmj.g7647

Soulas, T., Storme, M., Martínez-Martín, P., Pichlak, M., Gurruchaga, J. M., Palfi, S., & Fénelon, G. (2016). Assessing health-related quality of life with the SCOPA-PS in French individuals with Parkinson's disease having undergone DBS-STN: A validation study. *Revue Neurologique, 172*(4-5), 281-288. doi:10.1016/j.neurol.2015.10.010

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology, 126*(1), S124-S128. doi:10.1053/j.gastro.2003.10.016

Suratos, C. T. R., Saranza, G. R. M., Sumalapao, D. E. P., & Jamora, R. D. G. (2018). Quality of life and Parkinson's disease: Philippine translation and validation of the Parkinson's disease questionnaire. *Journal of Clinical Neuroscience, 54*, 156-160. doi:10.1016/j.jocn.2018.06.013

Tan, L. C. S., Lau, P.-N., Au, W.-L., & Luo, N. (2007). Validation of PDQ-8 as an independent instrument in English and Chinese. *Journal of the Neurological Sciences, 255*(1), 77-80. doi:10.1016/j.jns.2007.01.072

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, *60*(1), 34-42.

Todorova, A., & Stambolieva, K. (2007). Validation of the Bulgarian version of Parkinson's Impact Scale (PIMS) as a measure of quality of life for patients with Parkinson's disease. *Acta Medica Bulgarica, 34*(2), 35-40. Retrieved from https://www.researchgate.net/publication/289802755_Validation_of_the_Bulgarian_version_of_Parkinson's_Impact_Scale_PIMS_as_a_measure_of_quality_of_life_for_patients_with_Parkinson's_disease

Virués-Ortega, J. P., Carod-Artal, F. J. M. D. P., Serrano-Dueñas, M. M. D., Ruiz-Galeano, G. M. D., Meza-Rojas, G. M. D., Velázquez, C. M. D., . . . Martínez-Martín, P. M. D. P. (2009). Cross-Cultural Validation of the Scales for Outcomes in Parkinson's Disease-Psychosocial Questionnaire (SCOPA-PS) in Four Latin American Countries. *Value in Health, 12*(2), 385-391. doi:10.1111/j.1524-4733.2008.00436.x

Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British medical bulletin, 96*(1), 5-21. doi:10.1093/bmb/ldq033

Zhang, J.-L., & Chan, P. (2012). Reliability and validity of PDQ-39: a quality-of-life measure for patients with PD in China. *Quality of Life Research, 21*(7), 1217-1221. doi:10.1007/s11136-011-0026-1

Žiropađa, L., Stefanova, E., Potrebić, A., & Kostić, V. S. (2009). Quality of Life in Serbian Patients with Parkinson's Disease. *Quality of Life Research, 18*(7), 833-839. doi:10.1007/s11136-009-9500-4

Chapter 2: Major Research Project

Validation of the Distress Thermometer with People with Parkinson's Disease

Bronagh Reynolds, MSc,[1,2] Susan O'Connell, DClinPsy[2] and Breda Cullen PhD[1]

[1]University of Glasgow/NHS Scotland Clinical Psychology Training Programme, Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, 1st floor, Administration Building, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow, G12 0XH

[2] NHS Ayrshire and Arran

Correspondence to:

Institute of Health and Wellbeing
College of Medical, Veterinary and Life Sciences

University of Glasgow
1st floor, Administration Building
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow G12 0XH

Bronagh Reynolds email:

Prepared in accordance with submission requirement for the Parkinson's Disease Journal (see Appendix J, page 87).

Word count including abstract and references: 5,957

Plain English Summary

Is the Distress Thermometer a valid measure to screen for distress in people with Parkinson's disease?

Background

The Distress Thermometer (DT) is a very brief screening tool, originally created for use with people with cancer (Roth et al., 1998). Firstly, you are required to rate your distress from 0-10 with 10 indicating higher distress, and then you are asked to tick relevant problems from a problem list. As it is brief, it has become a popular screening tool in healthcare settings and has been found to be valid in other groups of people such as those with irritable bowel syndrome and respiratory conditions. Parkinson's disease (PD) is a progressive condition for which there are treatments but no cure. As the disease progresses, people's ability to do day to day tasks such as driving, shopping and managing finances can reduce. This understandably has a huge effect on their quality of life and can cause considerable distress for people. Symptoms and their progression vary from person to person, and so it is important to be able to identify those who are distressed and may benefit from further healthcare, or other, support.

Aims and Questions

As the DT has not been formally evaluated with people with PD previously, we aimed to find out if it is a useful tool to screen for distress in people with PD. To do this, we compared the DT to an existing measure which has been shown to be valid for use with people with PD, the Hospital Anxiety and Depression Scale (HADS). Our question was:

- Is the Distress Thermometer a valid screening tool in people with PD?

Method

People with a diagnosis of PD were recruited by PD Nurses from a movement disorder clinic in Ayrshire and Arran. People were excluded if they had other neurodegenerative conditions (apart from dementia), and all had to be aged 18 or older and speak English. Unfortunately, due to the outbreak of COVID-19 and subsequent lockdown restrictions, only 40 people (out of our target of 102) were recruited for this study. The 40 people who took part were asked to fill out the DT and HADS after a routine appointment with their PD Nurse. Responses for both measures were then analysed using statistical software to see how correlated (associated) the two were. The more closely correlated they are, the more likely it is that the DT is valid to use with people with PD.

Results and conclusions

Analysis revealed a large correlation between the DT and HADS, especially for the HADS Anxiety score. The results seem to support the use of the DT in people with PD, but further research with more participants is needed to say this for sure. Results of the study will be presented to the clinical team.

References

Roth, A. J., Kornblith, A. B., Batel-Copel, L., Peabody, E., Scher, H. I., & Holland, J. C. (1998). Rapid screening for psychologic distress in men with prostate carcinoma: A pilot study. *Cancer*, *82*(10), 1904-1908. doi:10.1002/(SICI)1097-0142(19980515)82:10<1904::AID-CNCR13>3.0.CO;2-X

Abstract

Background: The Distress Thermometer (DT) was developed as a tool to detect distress in cancer patients and has since been validated across various medical conditions and age groups. Idiopathic Parkinson's disease (IPD) is a progressive neurological condition that causes a variety of disabling motor symptoms (such as tremor, rigidity and slowness of movement) and non-motor symptoms including cognitive and mood disorders, and sleep disturbances. To date, no studies have assessed the performance of the DT for detecting distress in this population.

Objective: The primary aim of this study was to determine if the DT is a valid measure to detect distress in patients with IPD.

Methods: This was a prospective observational study. Participants with IPD were recruited from the Movement Disorder Clinic in NHS Ayrshire & Arran. The DT was administered along with the Hospital Anxiety and Depression Scale (HADS). To assess test-retest reliability, the DT was administered before and after the clinic appointment.

Results: Forty participants took part. Large correlations were found between the DT and HADS anxiety (rho = 0.68, 98% confidence interval (CI) 0.38 to 0.85) and depression (rho = 0.58, 98% CI 0.24 to 0.79). Test retest reliability was excellent (rho = 0.98, 95% CI 0.96 to 0.99). The most frequently endorsed problems were related to difficulties sleeping and walking.

Conclusions: Despite limitations of the current study, primarily the modest sample size, the DT may be a promising measure to assess distress in people with IPD. Further research with a larger IPD sample is needed.

Keywords: Parkinson Disease; Psychological distress; validation study**;** surveys and questionnaires.

Introduction

"Distress is a multi-faceted concept pertaining to a negative or unpleasant experience which may be psychological (i.e. cognitive, behavioural, emotional), social and/or spiritual in nature…" (National Comprehensive Cancer Network, 2002, p.6). Distress has been widely documented in the literature as negatively impacting on quality of life and can act as a barrier to seeking appropriate treatment (Ransom, Jacobsen, & Booth-Jones, 2006). Distress, in conjunction with other factors, can lower adherence to treatment which can lead to further distress and treatment complications, as well as increased health care costs (Straka, Minar, Gazova, Valkovic, & Kyselovic, 2018). Therefore, it is essential that healthcare professionals are aware of the impact of distress and can screen for this in medical settings.

There are many useful measures which assess distress such as the Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983), the Brief Symptom Inventory (Derogatis & Melisaratos, 1983) and the Nottingham Health Profile Index of Distress (Wann-Hansson, Klevsgard, & Hagell, 2008). One widely used, brief measure to screen for distress is the Distress Thermometer (DT), originally developed for use with cancer patients (Roth et al., 1998). It is a visual analogue scale ranging from 0-10 with individuals required to circle the number which best represents their levels of distress in the past week, with higher scores indicating higher levels of distress. They are then asked to indicate the source of this distress under domains including practical, family and/or emotional problems, by completing the accompanying problem list (PL). The application of the DT has been expanded from adult cancer patients to other clinical populations, such as childhood cancer survivors (Geest, Dorp, Pluijm, & Heuvel-Eibrink, 2018), adolescents with schizophrenia (Bai et al., 2020), and people with irritable bowel syndrome (Canaletti et al., 2020).

Another clinical population for which distress screening is important is Parkinson's disease. Idiopathic Parkinson's disease (IPD) is a progressive neurological condition that causes a variety of disabling motor and non-motor symptoms. Psychological distress, including anxiety and depression, has been widely shown to be under-reported in people with PD and to have a negative impact on quality of life (Chen & Marsh, 2014). As the disease progresses medication effectiveness decreases and the risk of experiencing an array of symptoms such as cognitive impairment, psychosis, pain and autonomic dysfunction increases (Brown &

Fernie, 2014). Over time these impairments can lead to reductions in independence such as the ability to carry out instrumental activities of daily living (Foster, 2014).

Awareness of these issues has motivated clinical services to screen for distress in people with IPD. This is consistent with National Institute for Health and Care Excellence (NICE) guidelines for chronic health conditions, which state that screening tools should be used when depression is suspected, in order to inform stepped care interventions, with the DT mentioned as being a useful screening tool (National Institute for Health and Care Excellence, 2009). The DT is recommended for use in other chronic conditions such as epilepsy (Scottish Intercollegiate Guidelines Network, 2018) and neurodegenerative diseases such as dementia (Brechin, Codner, James & Murphy, 2020). It can be of particular use when an individual has motor or literacy difficulties.

Based on NICE guidance and its established usage in chronic conditions and neurodegenerative conditions, the Movement Disorders Clinic (MDC) in NHS Ayrshire & Arran introduced the DT with an adapted PL, the movement disorder -problem list (MD-PL) in 2017. Prior to the introduction of the DT, distress was not routinely enquired about at the MDC, with the onus on the patient to instigate disclosure of distress and related issues. Introducing the DT led to a more systematic approach to identifying distress and informing onward referral to appropriate services.

Despite the acceptability and apparent usefulness of this process in clinical practice; however, a key issue is that the validity of the DT for identifying distress in people with IPD has not been formally examined in the research literature. This represents an important gap in the evidence base. The design of such a study in IPD can be informed by previous validation research in other clinical populations. These have used various comparators to assess validity against the DT. Although there are examples of studies using both diagnostic schedules and other screening tools as comparators – e.g. the Composite International Diagnostic Interview (CIDI) (Patel, Sharpe, Thewes, Bell, & Clarke, 2010) the use of the HADS (Zigmond and Snaith, 1983) for comparator analysis against the DT is most common in the literature (Campbell et al., 2009; Geest et al., 2018; Zwahlen, Hagenbuch, Carley, Recklitis, & Buchi, 2008).

The optimal cut-off range for detecting clinical levels of distress using the DT has typically been reported in the literature to be between 3-5 on the 10-point scale (Geest et al., 2018; Patel et al., 2010; Wiener, Battles, Zadeh, Widemann, & Pao, 2017). It is not yet known

whether a similar score threshold would be suitable in IPD. The PL which accompanies the DT has been adapted successfully in order to make it more relevant to specific clinical populations (San Giorgi et al., 2017). In people with recurrent respiratory papillomatosis, the adapted PL was found to have excellent overall internal consistency, with several of the adapted items predictive of distress in this population e.g. inclusion of speech difficulties. It would be of interest to investigate whether a similar relationship is found in IPD.

Investigating the number and type of problems endorsed on the PL would also be beneficial to the service. It would inform and provide justification for the running of particular groups or interventions. For example, if those who rate themselves as being highly distressed also tended to endorse a particular type of problem, this may be a potential area to investigate further either through gathering qualitative information from people or completion of an audit.

A further consideration is test-retest reliability. It has been suggested in the literature that clinician reassurance and empathy may moderate patients' distress levels (Derksen, Bensing, & Lagro-Janssen, 2013). Lower DT ratings after a medical appointment may indicate that the DT registers transient distress levels which are likely to resolve after routine consultation with a medical professional, and so would speak to its clinical utility in the IPD population for streamlining onward referrals to relevant specialist services. The DT has not as yet been evidenced to have reliability in this population and so this is an important consideration for the current study. This is of particular note as the DT is currently only used prior to the person's consultation which occurs every six to twelve months, and so clinicians need to be assured that the DT score is measuring a stable construct.

Aims

This study aimed to assess the validity of the DT and associated PL in detecting distress in people with IPD. Test retest reliability was also evaluated to assess the stability of the DT before and after the medical consultation. Lastly, this study examined the PL to identify its usefulness as a potential explanation of participants' DT scores. The specific research questions were as follows:

Primary research question:

Is the DT a valid measure to detect distress in patients with IPD?

Secondary research questions:

What is the optimal cut-off point on the DT to accurately classify presence of distress?

Does the DT have appropriate test retest reliability in IPD patients when measured prior to and after a medical appointment relating to their IPD care?

What are the most common problems endorsed on the MD-PL in this population?

What is the correlation between the DT score and total number of problems overall?

What is the correlation between the DT score and number of problems in each separate domain?

Methods

Participants

Participants were recruited in NHS Ayrshire and Arran at the Biggart Hospital and Douglas Grant Rehabilitation Centre. Inclusion criteria were diagnosis of IPD; registered with a Movement Disorder Consultant in the MDC; able to give informed consent; able to understand and respond to the study questionnaires (carer/staff may assist with writing if necessary); aged 18 years or older. Exclusion criteria: other comorbid neurodegenerative diseases (except dementia). Recruitment began on 25th October 2019. As of 16th March 2020, all research recruitment was suspended due to the outbreak of COVID-19 and subsequent lockdown restrictions.

Ethical Approval

This study was approved by the Research and Development Department in NHS Ayrshire and Arran (reference number CM/KLB/NM R&D2019AA030; see Appendix K on page 89) and NHS Grampian Research Ethics Committee (reference number 19/NS/0112; see Appendix L on page 91). All participants gave written informed consent and relevant data protection regulations were followed for the secure storage of study data.

Measures

As previously outlined, the DT is an 11-point scale with an associated MD-PL tailored for movement disorders (see Appendix M page 94). The HADS (Zigmond & Snaith, 1983) is a

self-report measure to assess mood state, which is commonly used in IPD patients and is recommended by Scottish Intercollegiate Guideline Network (SIGN) guidelines (2010) as a screen for depression in patients with IPD as it has shown to be a reliable measure in people with PD (Schrag et al., 2007). It consists of 14 items (7 items on depression and 7 on anxiety) with each item rated from 0 to 3 with higher scores indicating elevated distress. Subscale totals for depression and anxiety are calculated separately, with an overall range between 0 and 21 for each subscale. Scores between 0-7 indicate minimal depression/anxiety, scores between 8–10 indicate borderline depression/anxiety and scores of 11 and above meet criteria for 'caseness', indicating moderate to severe anxiety and/or depression is present.

A demographic questionnaire (see Appendix N, page 95) was also administered. This gathered information on participants' gender, age, employment and relationship status, as well as asking about whether their DT score had changed since their appointment. Participants were asked what year they received their diagnosis of PD.

Procedure

A PD nurse screened clinic appointment schedules to identify eligible patients and sent out the participant information sheet several weeks in advance of their appointment. The DT was completed by the patient in the MDC just before their appointment, as is current routine practice within the MDC. As part of their consultation, the healthcare professional asked eligible patients if they were willing to talk to the researcher about the study. If they expressed interest in speaking to the researcher, they were brought to a nearby clinic room where any questions were answered. If they consented to take part and signed the consent form, they then completed the study questionnaires, with the total time estimated to be 25 minutes. This entailed completing the DT again (timepoint two), along with the HADS and the demographic questionnaire. GP information was obtained and a summary of the research, DT and HADS scores was sent to the relevant medical practice.

Sample size calculation

A priori power analysis was conducted using G*power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) to determine the required sample size. The minimum correlation level was set at 0.3 based on the estimated correlation between the DT and HADS subscales: 0.51 to 0.56 for the anxiety subscale, and 0.36 to 0.69 for the depression subscale (San Giorgi et al., 2017; Zwahlen et al., 2008). As the primary research question involved two statistical tests

(separately for the depression and anxiety subscales), a Bonferroni correction was applied to reduce the likelihood of type 1 error (p-value =0.025). The parameters were: $r = 0.3$ (medium), $p = 0.025$ (two-tailed), power (1-β) = 0.80 (Cohen, 1992). The required sample size was determined to be 102 participants. For additional details on expected recruitment rates see Appendix O, page 98.

Statistical Analysis

Data were analysed using SPSS version 26. Descriptive statistics were used to characterise the study sample. To investigate the relationship between the DT and the HADS, Spearman correlation analysis was used. Due to the Bonferroni-corrected significance threshold for these analyses, the 98% confidence interval (CI) is presented with these correlation estimates instead of the usual 95% CI.

Although this study was not a diagnostic accuracy study, in that distress itself is not a diagnosis, a receiver operating characteristic (ROC) curve can be used to identify the optimal cut-off point on the DT for capturing clinically significant distress, in line with methods outlined by the Standards for reporting of Diagnostic Accuracy Studies (STARD) (Bossuyt et al., 2003). Therefore, to answer the second research question, ROC curve analysis was planned to assess the classification performance (specificity and sensitivity) for each score on the DT against the HADS (score ≥11 versus <11). The graphic display of the curve facilitates the selection of an optimal cut-off score for distress for IPD patients. As there is no consensus on what constitutes appropriate specificity and sensitivity, Youden's index was considered to establish the cut-off point that maximises both of these domains.

To determine the test retest reliability of the DT, Spearman correlation coefficient analysis was carried out for pre and post appointment scores. Descriptive statistics were reported for the number and type of problems endorsed on the MD-PL, pre- and post-appointment. Spearman correlations were examined between the DT score and number of problems identified by participants, both overall and within each of the seven problem domains.

Due to the large number of correlations calculated for this research question, Bonferroni correction was deemed to be too restrictive, and the false discovery rate (FDR) correction was applied instead, to ensure that the proportion of false positives was controlled at 0.05 across these MD-PL analyses (Benjamini & Hochberg, 1995). The FDR corrections were performed using an online calculator (https://www.sdmproject.com/utilities/?show=FDR). As

the false positive rate is already controlled in these adjusted p values, the significance threshold for each of these results was 0.05 and the 95% CI is reported with the correlation estimates.

Results

A total of 40 people with a diagnosis of IPD participated in this study; see Appendix P, page 99, for a flowchart of recruitment. Demographic information for participants is presented below in Table 2.1

Table 2.1: Demographic characteristics of the sample (N =40).

	N	%
Age in years		
55-64	7	17.5
65-74	14	35
75 or older	19	47.5
Gender		
Male	23	57.5
Female	17	42.5
Living arrangements		
Alone	10	25
With family or friend	30	75
Relationship status		
Single	4	10
Married or have partner	29	72.5
Widowed	6	15
Other	1	2.5
Employment status		
Employed/self-employed	3	7.5
Retired	37	92.5

Descriptive statistics for the DT pre and post appointment and HADS subscales are presented in Table 2.2. The HADS anxiety subscale was found to be normally distributed after inspection of a scatterplot; however, the HADS depression subscale was non-normally distributed. For consistency, the median and quartiles are presented for both.

Table 2.2: Descriptive statistics for clinical measures.

	Median	Percentile	
		25th	75th
Duration of Parkinson's disease (years)	3.5	1	6
Distress thermometer pre appt	3.5	1	6
Distress thermometer post appt	4	1	6
HADS anxiety	5	3	8
HADS depression	6	4	10
	N	**%**	
HADS anxiety caseness	4	10	
HADS depression caseness	2	5	

HADS = Hospital Anxiety and Depression Scale

Is the DT a valid measure to detect distress in patients with IPD?

Spearman's correlation coefficient was calculated between the DT at time point 2 (post appointment) and the HADS subscales. The HADS anxiety subscale demonstrated a large, significant correlation of 0.68 with the DT, p = <0.001, 98% CI = 0.38 to 0.85. The HADS depression subscale also demonstrated a large, significant correlation of 0.58 with the DT, p = <0.001, 98% CI = 0.24 to 0.79.

However, the sample size recruited for this study was below the amount determined by *a priori* analysis to reliably detect an effect size of 0.30 or higher. Although the estimated correlation sizes for the DT with both the HADS anxiety and depression are above .30 in the current study, the CIs show that the true correlations could be as low as .38 or .24 respectively.

Apriori sensitivity power analysis was conducted using GPower*3 to determine the minimum correlation size a sample of 40 would be able to detect reliably. This determined that a correlation of 0.46 or higher could reliably be detected by the current study. Given the lower bounds of the CIs reported above, it is possible that the true correlations are below 0.46 and therefore not reliably detectable by the current study.

What is the optimal cut-off point on the DT to accurately classify presence of distress?

Unfortunately, *Apriori* sensitivity power analysis using pROC software in R revealed that to detect a reasonable balance between sensitivity and specificity i.e. area under the curve =.75, the current study was underpowered. This was due to the limited number of participants meeting caseness for anxiety (N=4) or depression (N=2) according to the HADS, giving an estimated statistical power of only 0.29 for anxiety and 0.15 for depression. For this reason, the ROC curve analyses were not undertaken.

Does the DT have appropriate test retest reliability in IPD patients when measured prior to and after a medical appointment relating to their IPD care?

Spearman's correlation analysis revealed a very large, significant correlation of 0.98, p <0.001, 95% CI = 0.96 to 0.99. The DT scores changed post appointment for eight participants. Two participants' scores decreased by one point whilst the other six increased by one to two points. For the two participants whose scores had decreased, both explained that they had felt reassured talking about their diagnosis with their PD Nurse.

For those whose DT score increased post appointment, varied reasons were given. One participant noted that seeing all the potential problems written down increased their distress, whilst another reported that they were under time pressure after their appointment to get to another appointment and so this slightly increased their distress. Another participant reported that since filling out the pre-DT, they had thought more about the problems listed and identified more, leading to a higher reported level of overall distress. Other reasons for an increase in DT scores were most relevant to the emotional domain of the DT problem list and involved other issues arising during their appointment regarding limitations of treatment.

What are the most common problems endorsed on the MD-PL in this population?

The number of problems endorsed was explored. Median and quartiles for the number of problems identified per domain pre and post appointment are displayed below in Table 2.3. As can be seen from the table, most problems reported were in the physical and motor domains.

Table 2.3: Descriptive statistics for Movement Disorder Problem List pre and post appointment.

		Number of problems endorsed			Number of problems endorsed		
Domain	**No. problems listed**	**Pre**	**Percentile**		**Post**	**Percentile**	
		Median	**25th**	**75th**	**Median**	**25th**	**75th**
Physical	22	5	3	6.75	5	3	7
Motor	8	3	1	4	2.5	1.25	3.75
Cognitive	6	1	0	2	1	0	2
Practical	4	0	0	0	0	0	0
Family	4	0	0	0	0	0	0
Emotional	7	1	0	2	1	0	2
Spiritual	3	0	0	0	0	0	0
Total	54	9.5	6	13	9.5	6	13

Table 2.4 below displays the type of problems endorsed. In the physical domain, the most frequently reported problems both pre and post appointment were sleep problems and fatigue; sleep problems were endorsed by 65% both pre and post appointment, and fatigue was endorsed by 57.5% pre and 55% post.

In the motor domain the most reported problem was walking, followed by stiffness. There were six problems not reported by any of the participants in this study: odd/bizarre behaviour (cognitive domain), housing problems (practical domain), problems with relatives/friends (family domain), and none of the three problems were endorsed in the spiritual domain. Other problems not listed were backache, senses, symptom change and speech impairment.

Table 2.4: Types of problems endorsed from the Movement Disorder Problem List.

Problem List	Pre		Post	
	N	%	N	%
Physical				
Your Appearance	2	5%	3	7.5%
Bathing or dressing	8	20%	7	17.5%
Dribbling saliva	11	27.5%	11	27.5%
Swallowing problems	4	10%	4	10%
Eating/Appetite	4	10%	7	17.5%
Change in weight	6	15%	5	12.5%
Sore/dry mouth	9	22.5%	9	22.5%
Eating/Appetite	3	7.5%	3	7.5%
Nausea/Vomiting	3	7.5%	3	7.5%
Urinary problems	17	42.5%	17	42.5%
Bowel problems	12	30%	14	35%
Sleep problems	26	65%	26	65%
Nightmares	2	5%	3	7.5%
Acting out in sleep	3	7.5%	3	7.5%
Need to move legs at night	10	25%	10	25%
Day time sleepiness	21	52.5%	16	40%
Fatigue or tiredness	23	57.5%	22	55%
Swollen legs	8	20%	9	22.5%
Pain	13	32.5%	13	32.5%
Sweats	5	12.5%	4	10%

Problem List	Pre		Post	
	N	%	N	%
Sexual concerns	2	5%	2	5%
Taking medication	3	7.5%	4	10%
Motor				
Tremor	19	47.5%	17	42.5%
Fine motor control	6	15%	6	15.00%
Walking	24	60%	22	55.00%
Stiffness	18	45%	19	47.50%
Weakness	15	37%	16	40.00%
Freezing	9	22.5%	10	25.00%
Bed/Chair mobility	8	20%	8	20.00%
Falls	6	15%	4	10.00%
Cognitive				
Memory	13	32.5%	13	32.50%
Speed of thinking	11	27.5%	12	30.00%
Concentration and attention	8	20%	7	17.50%
Judging distance/Space	4	10%	5	12.50%
Odd/Bizarre behaviour	0	0%	0	0.00%
Impulsive	4	10%	4	10.00%
Practical				
Caring responsibilities	2	5.00%	2	5.00%
Finances, work	2	5.00%	2	5.00%
Housing	0	0.00%	1	2.50%

Problem List	Pre		Post	
	N	%	N	%
Transport/Driving	3	7.50%	3	7.50%
Family				
Relationship with children	1	2.50%	1	2.50%
Relationship with partner	1	2.50%	1	2.50%
Relationship with relatives/Friends	0	0.00%	0	0.00%
Burden (on family, friends etc)	4	10.00%	4	10.00%
Emotional				
Sadness or depression	10	25%	11	27.50%
Loneliness or isolation	5	12.5%	4	10.00%
Hopelessness	2	5%	2	5.00%
Worry, fear or anxiety	14	35%	14	35.00%
Loss of control or freedom	7	17.5%	7	17.50%
Anger or frustration	9	22.5%	10	25.00%
Seeing/Hearing things not there	5	12.5%	5	12.50%
Spiritual				
Spiritual concerns	0	0%	0	0%
Religious concerns	0	0%	0	0%
Other spiritual concerns	0	0%	0	0%

What is the correlation between pre and post appointment DT scores with number of problems endorsed on the pre and post PL?

Correlations were calculated between the DT score and the number of problems (overall total, and total per domain), both pre and post appointment. Results are displayed below in Table

2.5. No problems were endorsed in the spiritual domain and so no correlation could be calculated.

Table 2.5: Correlations between the Distress Thermometer score and number of problems endorsed.

	Pre			**Post**		
	Correlation (*rho*)	**FDR corrected p-value**	**95% confidence interval**	**Correlation (*rho*)**	**FDR corrected p-value**	**95% confidence interval**
Physical	0.62	<0.001	0.36 to 0.79	0.61	<0.001	0.34 to 0.79
Motor	0.64	<0.001	0.38 to 0.80	0.52	<0.001	0.23 to 0.73
Cognitive	0.21	0.21	-0.11 to 0.49	0.22	0.18	-0.10 to 0.50
Practical	0.21	0.21	-0.11 to 0.49	0.31	0.07	-0.01 to 0.57
Family	0.45	<0.001	0.15 to 0.68	0.35	0.04	0.03 to 0.60
Emotional	0.68	<0.001	0.44 to 0.83	0.57	<0.001	0.29 to 0.76
Total	0.76	<0.001	0.56 to 0.88	0.71	<0.001	0.48 to 0.85

FDR = False Discovery Rate

A similar pattern of correlations was observed both pre and post appointment. The emotional domain demonstrated the highest correlation with the DT pre appointment, while the cognitive domain showed the lowest correlation with the DT both pre and post appointment. These results indicate that interventions targeted at reducing emotional distress for example, may be worthwhile as those tending to score higher on the DT also tended to tick problems in the emotional domain.

Similarly, descriptive statistics has already established that the problems most frequently endorsed were in the physical and motor domains. The large correlations shown in table 2.5 indicate that interventions in these areas may be the most promising.

Discussion

This study investigated whether the distress thermometer was a valid screening tool for distress in an IPD population. The results indicate the DT may be a valid measure to screen for distress in this population. The correlation sizes found in this study are at the larger end of those that typically have been found in the literature (San Giorgi et al., 2017; Zwahlen et al., 2008). This may be due to an over-inflation of the correlation due to the relatively small sample size in the current study (discussed further below).

The present results are in line with the literature in other respects, as the HADS anxiety subscale tends to be more highly correlated with the DT than the depression subscale (Geest et al., 2018; Gil et al., 2005; Testoni et al., 2018). Although promising, it is important to note that the use of the DT as a screening tool should enhance, not replace, clinician judgement (Mitchell, 2007). The DT may aid in the identification of distress which may not otherwise have been discussed, but it is the clinician who facilitates referral to the appropriate pathway of care for the person as required, utilising skilled questioning techniques and decision-making within the wider multi-disciplinary team.

As explained above, ROC curve analysis was not performed and so the optimal cut-off point to enhance sensitivity and specificity in detecting distress in this study is not known. Acceptable test retest reliability results in this study indicate that the DT scores remain relatively stable after an appointment related to IPD care, which supports the assumption that the DT measures stable, not transient, levels of distress. This is congruent with another study in cancer patients, which found a test retest reliability coefficient for the DT of 0.80 after 7-10 days (Tang, Zhang, Pang, Zhang, & Song, 2011). The larger coefficient found in the present study may, again, be due to the modest sample size, or the short duration between the two measurements.

It is important to highlight that three participants stated that their level of distress post appointment was itself influenced by being asked to rate their distress.[1] This may indicate that the DT can have an unintended adverse effect of increasing distress, instead of serving

[1]Distress was raised by one point for each person; this was discussed with all participants and they were aware they could contact their GP/MDC clinician if their distress continued. All stated they understood and felt okay to leave either alone or with partner/friend.

the intended purpose of identifying existing distress in order to provide help. This effect has been acknowledged in another UK based study where 5% of 171 participants reported that filling out the DT had upset them in some way (Gessler et al., 2008). In another UK study, seven out of 598 respondents to a telephone survey using the DT were described as being too tearful or upset to answer the question about their level of distress, although it was not stated whether this was related to being asked about their distress, or if the individual had been upset prior to this (Hughes, Sargeant, & Hawkes, 2011).

The next research question focused on the problem list pre and post appointment and its correlation with the DT score pre and post appointment. The number of problems in the physical, motor and emotional domains were highly correlated with the DT. This is a similar pattern to what has been found in the literature. A validation study in a sample of people with cancer found that the DT score was most highly correlated with the physical ($r = 0.64$) and emotional domains ($r = 0.61$), while other correlations were moderate such as the practical ($r = 0.39$), family ($r = 0.31$) and spiritual domains ($r = 0.26$) (Tuinman, Gazendam-Donofrio, & Hoekstra-Weebers, 2008). In the present sample only four participants added in problems not represented in the MD-PL, so it is likely that the seven PL domains accurately captured the main sources of participant distress.

Strengths and limitations

This is the first study to validate the use of the DT in the IPD population. A strength of the current study is that it explored the PL as well as the overall DT rating of distress, as this is not typically explored in similar studies. This makes it a more comprehensive investigation into the usefulness of the DT and associated PL in people with IPD. However, the possibility of type 1 error in this study is increased as the sample size was underpowered to reliably detect a correlation below 0.46. This is a significant limitation of the current study and so all results should be interpreted with caution as they may be over inflated and not truly representative of the relationships between the variables studied.

Another limitation is the largely cross-sectional nature of the study (repeated measures conducted on the same day), as the sensitivity of the DT in tracking distress over time could not be determined by the present study. It could be argued that the use of the HADS as the comparator in this study was not ideal, as the HADS is itself a self-report screen. However, using more detailed diagnostic schedules such as the CIDI or DSM criteria to assess construct validity would not be appropriate here, as this study did not set out to identify clinical

disorders, and this is not what the DT is used for in routine clinical practice in this setting. A third limitation is the design of the study for assessing test retest reliability. DTs were completed over a short space of time, before and after their appointment. This design can only demonstrate whether or not the DT is stable pre and post appointment, not over a longer period of time. Other studies validating the DT typically assesses distress over a longer time period, such as week. This design would have allowed for an increased level of confidence in the reliability of the DT in people with IPD, over time. As the DT measures distress over the past week, this design is appropriate. Assessing distress using the DT immediately post appointment and one week post appointment would allow for stability over time to be assessed, in the absence of clinical influence. Although it was not feasible given the time constraints of the current study, this is an area for future research to consider a longer period between timepoints.

Future directions

The results of this study provide a basis for future research to replicate these findings using a larger sample. This would allow for more robust conclusions to be drawn. This is an important area because, as previously mentioned, the NICE guidelines recommend the DT to be used to screen for distress in chronic conditions (NICE, 2009). Further research in this population would therefore usefully inform these guidelines. The validity of the DT in other Parkinsonian disorders such as progressive supranuclear palsy, and other movement disorders should also be evaluated.

Conclusion

The DT shows promise and relevance as a screening tool for distress in individuals with IPD; however, replication of these results is needed in a larger sample to determine reliability and generalisability beyond that of the current study.

References

Bai, X., Wang, A., Cross, W., Lam, L., Plummer, V., Guan, Z., . . . Tang, S. (2020). Validation of the distress thermometer for caregivers of children and adolescents with schizophrenia. *Journal of Advanced Nursing, 76*(2), 687-698. doi:10.1111/jan.14233

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Grp, S. S. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal, 326*(7379), 41-44. doi:10.1136/bmj.326.7379.41

Brown, R. G., & Fernie, B. A. (2014). Metacognitions, anxiety, and distress related to motor fluctuations in Parkinson's disease. *Journal of Psychosomatic Research, 78*(2), 143-148. doi:10.1016/j.jpsychores.2014.09.021

Campbell, A., Steginga, S. K., Ferguson, M., Beeden, A., Walls, M., Cairns, W., & Dunn, J. (2009). Measuring distress in cancer patients: the Distress Thermometer in an Australian sample. *Progress in Palliative Care, 17*(2), 61-68. doi:10.1179/096992609X392259

Canaletti, C., Colombo, F., Dessì, A., Geccherle, E., Tongiorgi, A., Cai, P., . . . Monica, F. (2020). P106 Depressive symptoms and anxiety and their screening in IBD patients: performance of the distress thermometer in Italian IBD patients. *Journal of Crohn's and colitis, 14*(Supplement_1), S188-S189. doi:10.1093/ecco-jcc/jjz203.235

Chen, J. J., & Marsh, L. (2014). Anxiety in Parkinson's disease: identification and management. *Therapeutic Advances in Neurological Disorders, 7*(1), 52-59. doi:10.1177/1756285613495723

Cohen, J. (1992). A Power Primer. *Psychological bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Derksen, F., Bensing, J., & Lagro-Janssen, A. L. M. (2013). Effectiveness of empathy in general practice: a systematic review. *British Journal of General Practice, 63*(606), 76-84. doi:10.3399/bjgp13X660814

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: an introductory report. *Psychological Medicine, 13*(3), 595-605. doi:10.1017/S0033291700048017

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using GPower 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160. doi:10.3758/BRM.41.4.1149

Foster, E. R. (2014). Instrumental Activities of Daily Living Performance Among People With Parkinson's Disease Without Dementia. *American Journal of Occupational Therapy, 68*(3), 353-362. doi:10.5014/ajot.2014.010330

Geest, I. M. M., Dorp, W., Pluijm, S. M. F., & Heuvel‐Eibrink, M. M. (2018). The distress thermometer provides a simple screening tool for selecting distressed childhood cancer survivors. *Acta paediatrica, 107*(5), 871-874. doi:10.1111/apa.14251

Gessler, S., Low, J., Daniells, E., Williams, R., Brough, V., Tookman, A., & Jones, L. (2008). Screening for distress in cancer patients: is the distress thermometer a valid measure in the UK and does it measure change over time? A prospective validation study. *Psycho‐oncology, 17*(6), 538-547. doi:10.1002/pon.1273

Gil, F., Grassi, L., Travado, L., Tomamichel, M., Gonzalez, J. R., Grp, S., . . . The, S. G. (2005). Use of distress and depression thermometers to measure psychosocial morbidity among southern European cancer patients. *Supportive Care in Cancer, 13*(8), 600-606. doi:10.1007/s00520-005-0780-0

Hughes, K. L., Sargeant, H., & Hawkes, A. L. (2011). Acceptability of the Distress Thermometer and Problem List to community-based telephone cancer helpline operators, and to cancer patients and carers. *Bmc Cancer, 11*(1), 46-46. doi:10.1186/1471-2407-11-46

Mitchell, A. J. (2007). Pooled Results From 38 Analyses of the Accuracy of Distress Thermometer and Other Ultra-Short Methods of Detecting Cancer-Related Mood

Disorders. *Journal of Clinical Oncology, 25*(29), 4670-4681. doi:10.1200/JCO.2006.10.0438

Patel, D., Sharpe, L., Thewes, B., Bell, M. L., & Clarke, S. (2010). Using the distress thermometer and hospital anxiety and depression scale to screen for psychosocial morbidity in patients diagnosed with colorectal cancer. *Journal of Affective Disorders, 131*(1), 412-416. doi:10.1016/j.jad.2010.11.014

Ransom, S., Jacobsen, P. B., & Booth‐Jones, M. (2006). Validation of the Distress Thermometer with bone marrow transplant patients. *Psycho‐oncology, 15*(7), 604-612. doi:10.1002/pon.993

Roth, A. J., Kornblith, A. B., Batel-Copel, L., Peabody, E., Scher, H. I., & Holland, J. C. (1998). Rapid screening for psychologic distress in men with prostate carcinoma: A pilot study. *Cancer, 82*(10), 1904-1908. doi:10.1002/(SICI)1097-0142(19980515)82:10<1904::AID-CNCR13>3.0.CO;2-X

San Giorgi, M. R., Aaltonen, L. M., Rihkanen, H., Pian, R., van der Laan, B., Hoekstra-Weebers, J., & Dikkers, F. G. (2017). Validation of the Distress Thermometer and Problem List in Patients with Recurrent Respiratory Papillomatosis. *Otolaryngology-Head and Neck Surgery, 156*(1), 180-188. doi:10.1177/0194599816668307

Schrag, A., Barone, P., Brown, R. G., Leentjens, A. F. G., McDonald, W. M., Starkstein, S., . . . Goetz, C. G. (2007). Depression rating scales in Parkinson's disease: Critique and recommendations. *Movement Disorders, 22*(8), 1077-1092. doi:10.1002/mds.21333

Straka, I., Minar, M., Gazova, A., Valkovic, P., & Kyselovic, J. (2018). Clinical aspects of adherence to pharmacotherapy in Parkinson disease A PRISMA-compliant systematic review. *Medicine, 97*(23), e10962. doi:10.1097/MD.0000000000010962

Tang, L.-l., Zhang, Y.-n., Pang, Y., Zhang, H.-w., & Song, L.-l. (2011). Validation and reliability of distress thermometer in Chinese cancer patients. *Chinese Journal of Cancer Research, 23*(1), 54-58. doi:10.1007/s11670-011-0054-y

Testoni, I., Sansonetto, G., Ronconi, L., Rodelli, M., Baracco, G., & Grassi, L. (2018). Meaning of life, representation of death, and their association with psychological

distress. *Palliative & Supportive Care, 16*(5), 511-519. doi:10.1017/S1478951517000669

Tuinman, M. A., Gazendam-Donofrio, S. M., & Hoekstra-Weebers, J. E. (2008). Screening and referral for psychosocial distress in oncologic practice: Use of the distress thermometer. *Cancer, 113*(4), 870-878. doi:10.1002/cncr.23622

Wann-Hansson, C., Klevsgard, R., & Hagell, P. (2008). Cross-diagnostic validity of the Nottingham health profile index of distress (NHPD). *Health and Quality of Life Outcomes, 6*(1), 47-47. doi:10.1186/1477-7525-6-47

Wiener, L., Battles, H., Zadeh, S., Widemann, B. C., & Pao, M. (2017). Validity, specificity, feasibility and acceptability of a brief pediatric distress thermometer in outpatient clinics. *Psycho‐oncology, 26*(4), 461-468. doi:10.1002/pon.4038

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta psychiatrica scandinavica, 67*(6), 361-370. doi:10.1111/j.1600-0447.1983.tb09716.x

Zwahlen, D., Hagenbuch, N., Carley, M. I., Recklitis, C. J., & Buchi, S. (2008). Screening cancer patients' families with the distress thermometer (DT): a validation study. *Psycho‐oncology, 17*(10), 959-966. doi:10.1002/pon.1320

Chapter 3: Appendices

Appendix A. Author guidelines for submission to Movement Disorders Journal

Form of Manuscripts

Pages should be numbered in succession, the title page being number one.

The text of the manuscript should be in the following sequence:

(1) Title page:

The opening page of each manuscript should include:

(a) article title (no abbreviations/acronyms).Titles should be short, specific and clear. They should not exceed 100 characters. Do not use abbreviations/acronyms in the title;

(b) authors' names, degrees, and affiliations (indicate the specific affiliation of each author by superscript, Arabic numerals);

(c) name, address, telephone and email address of the corresponding author;

(d) word count;

(e) a running title not exceeding 45 letters and spaces;

(f) Key words – up to 5;

(g) Financial Disclosure/Conflict of Interest concerning the research related to the manuscript: All information on support and financial issues from all authors relative to the research covered in the submitted manuscript must be disclosed regardless of date. Other financial information unrelated to the current research covering the past year will be documented at the end of the manuscript (see below).

(h) Funding sources for study.

(2) Abstract

Structured Abstract: We require that authors submit structured abstracts. The page following the title page of Full-Length Articles should include an abstract of up to 250 words. The abstract should be structured. The page following the title page of a Brief Report should include a structured abstract of up to 150 words. Reviews should include an unstructured abstract. Viewpoints do not need any abstract.

(3) Introduction

Give a brief description of the background and relevance of the scientific contribution.

(4) Methods

Describe the methodology of the study. For experimental investigation of human or animal subjects, please state in this section that an appropriate institutional review board approved the project. For those investigators who do not have formal ethics review committees, the principles outlined in the “Declaration of Helsinki” should be followed. For investigations in human subjects, state in this section the manner in which informed consent was obtained from the subjects. A letter of consent must accompany all photographs, patient descriptions, and pedigrees in which a possibility of identification exists. The authors are responsible for ensuring anonymity.

(5) Results

No specific regulations.

(6) Discussion

No specific regulations.

(7) Acknowledgment

No specific regulations. These may be published online at the discretion of the editor.

(8) Authors' Roles

List all authors along with their specific roles in the project and preparation of the manuscript.

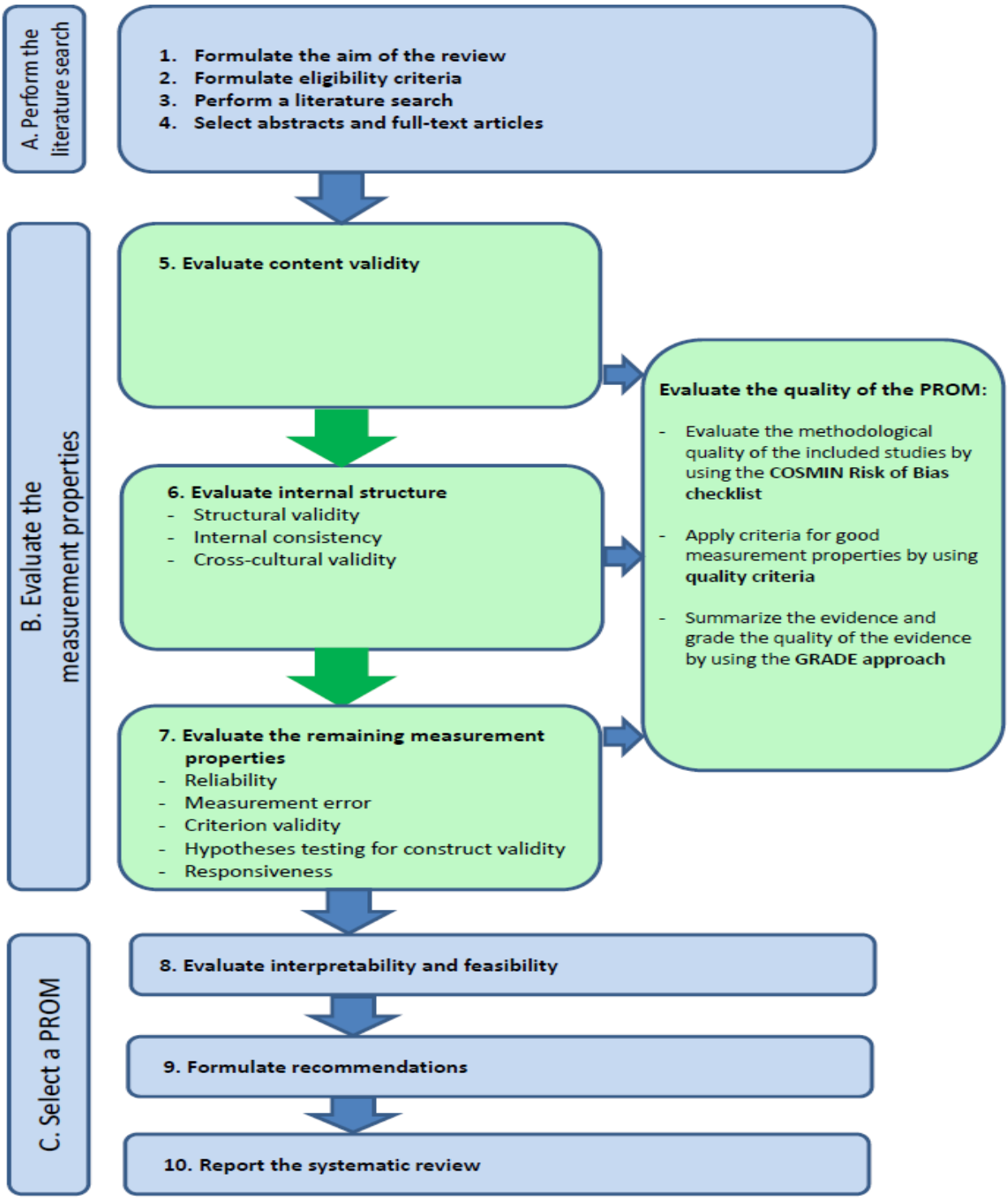
These may include but are not restricted to:

1) Research project: A. Conception, B. Organization, C. Execution;

2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique;

3) Manuscript: A. Writing of the first draft, B. Review and Critique.

Appendix B. Ten steps for conducting a systematic review of PROMS (Reprinted from with permission from the COSMIN manual (Mokkink et al., 2018).

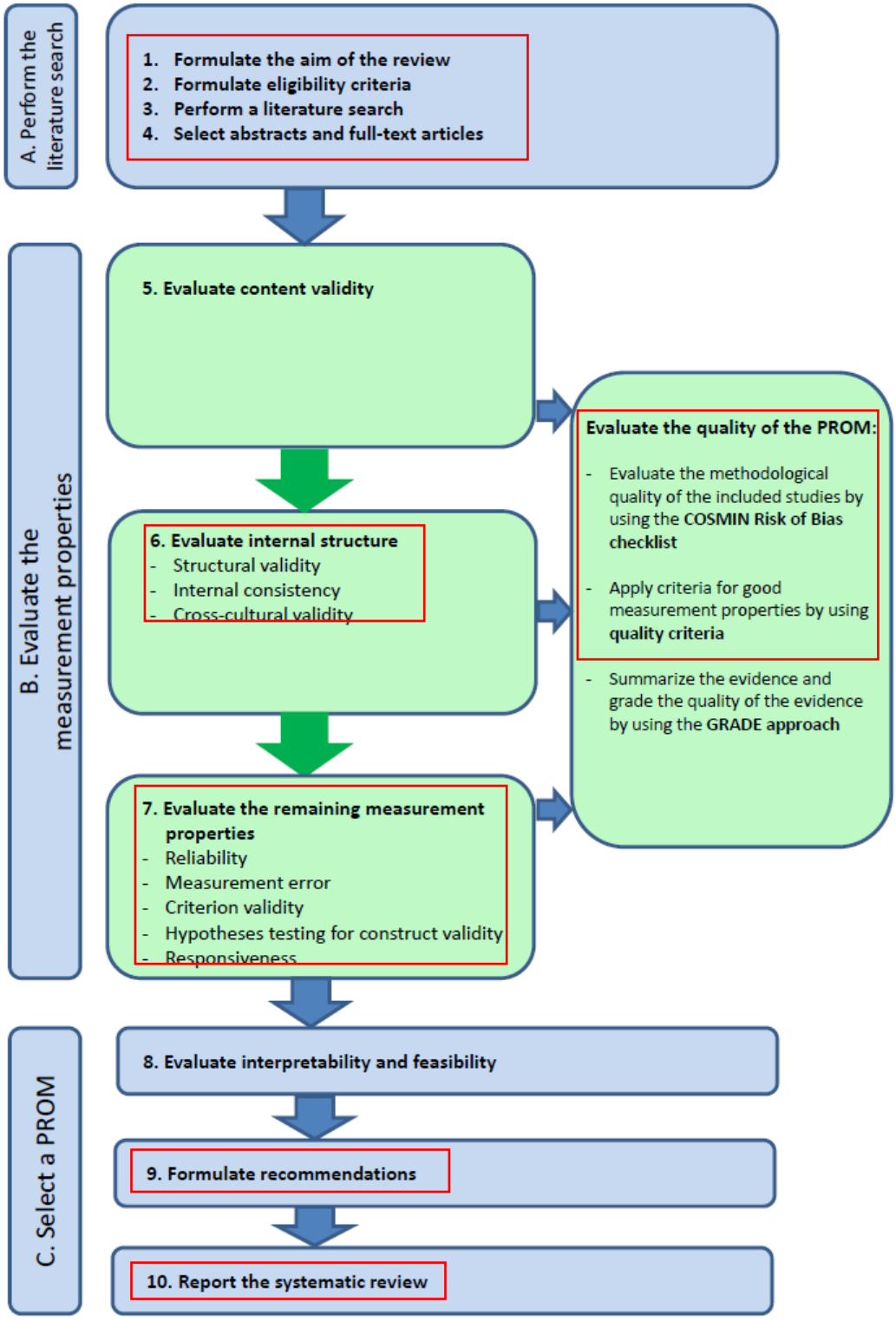


References

Mokkink, L. B., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H., . . . Mokkink, L. (2018). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual, 78*, 1.

Appendix C. Modifications made to COSMIN guidance

Steps followed are in red.



Appendix D Additional modifications made to COSMIN guidance

- COSMIN do not prescribe a column to report both subscales and summary index metrics for internal consistency and reliability; however, they recommend that if there are doubts about the potential multidimensionality of the instrument, the summary index number should not be used. As there are queries in the literature about the unidimensionality of all instruments included in this review, apart from the Parkinson's Impact Scale, which was designed to be multidimensional, it was decided that displaying both subscale and summary index figures would aid transparency when considering the psychometric properties of instruments.

- COSMIN guidance states that for construct validity, it is enough to report whether hypotheses for construct validity have been met. In results table 2, this was expanded upon to include the name of the instrument that was used as comparator and whether discriminant or convergent validity had been assessed.

- Regarding discriminant validity, the results tables templates provided by COSMIN do not include this; however, the risk of bias ratings do provide a specific column for discriminant validity. Therefore, it was decided to include this information in the results table to enable a more comprehensive review of each instrument. To balance transparency with following COSMIN guidance, the risk of bias rating for discriminant validity are displayed in table 2 but the specific metrics are not (e.g. correlations or means and standard deviations). Most studies included for review used p-value statistics to calculate discriminant validity. COSMIN guidance states that p-values are not an appropriate method of assessing this construct as it is not important

to know whether the figures are significantly different from zero (Mokkink et al., 2018). To prevent unfair or misleading risk of bias ratings; however, it was decided that if studies reported the figures i.e. the mean and standard deviations as well as the p-value, the risk of bias rating would be A to indicate 'adequate'.

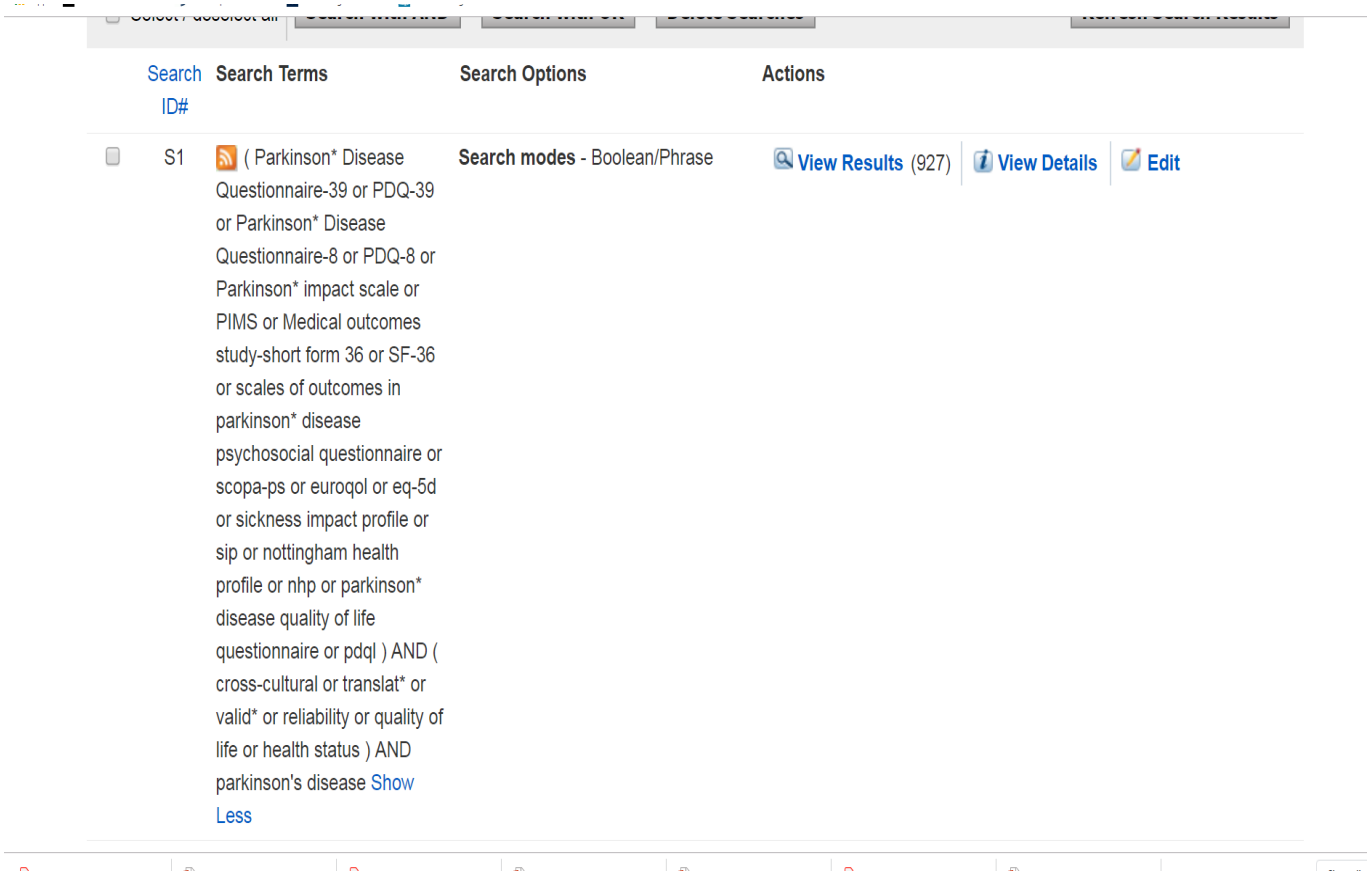
- COSMIN incorporates the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach to synthesise the evidence across studies. This was deemed unnecessary for the purposes of the current review as studies included were not predicted to be homogeneous given the varied languages and cultures across studies. It would therefore be unsuitable to collate these studies and assign an overall quality rating per measurement property.

- Two characteristics noted by COSMIN (PROM development and content validity) were not necessary to evaluate for the present review, as these apply to newly developed instruments. Content validity has been established in all instruments reviewed.

References

Mokkink, L. B., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H., . . . Mokkink, L. (2018). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual, 78*, 1.

Appendix E. Screenshot of search terms in EBSCOhost database.



Search ID#	Search Terms	Search Options	Actions
S1	(Parkinson* Disease Questionnaire-39 or PDQ-39 or Parkinson* Disease Questionnaire-8 or PDQ-8 or Parkinson* impact scale or PIMS or Medical outcomes study-short form 36 or SF-36 or scales of outcomes in parkinson* disease psychosocial questionnaire or scopa-ps or euroqol or eq-5d or sickness impact profile or sip or nottingham health profile or nhp or parkinson* disease quality of life questionnaire or pdql) AND (cross-cultural or translat* or valid* or reliability or quality of life or health status) AND parkinson's disease Show Less	**Search modes** - Boolean/Phrase	View Results (927) | View Details | Edit

Appendix F. Risk of bias procedure.

Reprinted with permission from COMSIN (Mokkink et al., 2018).

3.1.1 Evaluating the methodological quality of studies

To evaluate the methodological quality of the included studies using the COSMIN Risk of Bias checklist, it should first be determined which measurement properties are assessed in each article. The methods used in each study to produce the results for a given measurement property were evaluated to identify risk of bias.

Determine which measurement properties are assessed

Often multiple studies on different measurement properties are described in one article (i.e. one study for each measurement property, e.g. a study on internal consistency, construct validity, and reliability, each with its own specific design requirements). The quality of each study is separately evaluated using the corresponding COSMIN box (Table 3). The COSMIN Risk of Bias checklist should be used as a modular tool; only those boxes should be completed for the measurement properties that are evaluated in the article.

Table 3. Boxes of the COSMIN Risk of Bias checklist

Mark the measurement properties that have been evaluated in the article*.

Content validity
Box 1. PROM development
Box 2. Content validity
Internal structure
Box 3. Structural validity
Box 4. Internal consistency
Box 5. Cross-cultural validity\measurement invariance

Remaining measurement properties
Box 6. Reliability
Box 7. Measurement error
Box 8. Criterion validity
Box 9. Hypotheses testing for construct validity
Box 10. Responsiveness

* If a box needs to be completed more than once two or more marks can be placed.

Sometimes the same measurement property is reported for multiple (sub)groups in one article. For example, when an instrument is validated in two different countries and the measurement properties are reported for both countries separately. In that case, the same box may need to be completed multiple times if the design of the study was different among countries. The review team should decide which boxes should be completed (and how many times).An example of one question asked to determine risk of bias for structural validity is given below:

Statistical methods 1 For CTT: Was exploratory or confirmatory factor analysis performed?				
Very good	*Adequate*	*Doubtful*	*Inadequate*	*N/A*
Confirmatory factor analysis performed	Exploratory factor analysis performed		No exploratory or confirmatory analysis performed	Not applicable

References

Mokkink, L. B., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H., . . . Mokkink, L. (2018). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual, 78*, 1.

Appendix G. Descriptions of each measure included for review

EuroQol/EQ-5D

The EQ-5D is a self-report measure of generic HRQoL originally developed in 1990 (Brooks & EuroQol, 1996). It consists of five dimensions—mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels which are rated by the individual (no problem; some problem; extreme problem). These levels have since been expanded to five levels to prevent against ceiling effects for ratings (EQ-5D-5L). In addition, individuals are also asked to rate their health status on a 100-point visual analogue scale, with higher scores indicating higher HRQoL.

Parkinson's Disease Questionnaire-39 (PDQ-39)

The PDQ-39 is a 39-item self-report questionnaire which assess the HRQoL of people with PD. It was originally developed in 1995 and has eight separate subscales/domains including: Mobility, Activities of Daily Living, Emotional Well-Being, Stigma, Social Support, Cognitions, Communication and Bodily Discomfort (Jenkinson, Fitzpatrick, Peto, Greenhall, & Hyman, 1997). People are asked to respond to these 39 questions in relation to who they have felt over the past four weeks. Individual items are rated on a four-point Likert scale with higher scores representing worse HRQoL.

Parkinson's Disease Questionnaire-8 (PDQ-8)

The PDQ-8 was developed as a briefer version of the PDQ-39 and also asks for ratings of HRQoL over the past four weeks (Jenkinson et al., 1997). Individuals answer eight questions, one from each of the domains mentioned above for the PDQ-39. Several studies in this review utilised the PDQ-8/39 (nested in the PDQ-39). The PDQ-39 is given to participants

with eight questions highlighted to answer, rather than the PDQ-8 being presented as a separate questionnaire.

Parkinson's Disease Quality of Life Scale (PDQL)

The PDQL is a Parkinson's Disease specific HRQoL measure consisting of four subscales : parkinsonian symptoms (14 items), systemic symptoms (seven items), social function (seven items), and emotional function (nine items) (de Boer, Wijker, Speelman, & de Haes, 1996). This scale has a total of 37 items with higher scores indicating better HRQoL. It was originally developed in the Dutch language. All studies included in this review utilised the English version of the PDQL as a template for further translation.

Parkinson's Impact Measure Scale (PIMS)

The PIMS is a multi-dimensional self-report PD specify HRQoL measure which consists of 10 items (Calne et al., 1996). It was originally developed for use with individuals with a diagnosis of idiopathic PD. The items in the PIMS reflect broader domains of emotional, social and economic issues for the individual. Higher scores reflect lower HRQoL. Individuals whose PD symptoms are stable only score each item once, whereas those with fluctuations in symptoms judge the negative impact for both 'on' and 'off' periods. Broadly speaking 'on' refers to times when they are functioning at their best and 'off' when they are functioning at their worst. The scale contains two optional items (sexuality and financial security).

Scales for Outcomes - Psychosocial (SCOPA-PS)

The SCOPA-PS is an 11-item self-report instrument which assesses psychosocial functioning over the previous month (Marinus, Visser, Martínez-Martín, van Hilten, & Stiggelbout, 2003). Each item has four-point response rating scale (0-3). Higher scores indicate worse

psychosocial functioning. It focuses on areas of social activities, emotions and contact with others It was originally developed in the Dutch language. All studies included in this review used the English language version of the SCOPA-PS as a template.

References

Brooks, R., & EuroQol, G. (1996). EuroQol: the current state of play. *Health Policy, 37*(1), 53-72. doi:10.1016/0168-8510(96)00822-6

Calne, S., Schulzer, M., Mak, E., Guyette, C., Rohs, G., Hatchard, S., . . . Pegler, S. (1996). Validating a quality of life rating scale for idiopathic parkinsonism: Parkinson's Impact Scale (PIMS). *Parkinsonism and Related Disorders, 2*(2), 55-61. doi:10.1016/1353-8020(95)00026-7

de Boer, A. G., Wijker, W., Speelman, J. D., & de Haes, J. C. (1996). Quality of life in patients with Parkinson's disease: development of a questionnaire. *Journal of Neurology, Neurosurgery & Psychiatry, 61*(1), 70-74. doi:10.1136/jnnp.61.1.70

Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R., & Hyman, N. (1997). The PDQ-8: Development and validation of a short-form parkinson's disease questionnaire. *Psychology & Health, 12*(6), 805-814. doi:10.1080/08870449708406741

Marinus, J., Visser, M., Martínez-Martín, P., van Hilten, J. J., & Stiggelbout, A. M. (2003). A short psychosocial questionnaire for patients with Parkinson's disease: the SCOPA-PS. *Journal of Clinical Epidemiology, 56*(1), 61-67. doi:10.1016/S0895-4356(02)00569-3

Appendix H. Breakdown of instruments by language

Language breakdown for each instrument		
Language	**Instrument**	**Reference**
Spanish		
	PDQ-39	(Martinez-Martin, Serrano-Duenas, & Vaca-Baquero, 2005)
	PDQ-8/39	(Jenkinson & Fitzpatrick, 2006)
	PDQL	(Serrano-Dueñas, Martínez-Martín, & Vaca-Baquero, 2004)
	PIMS	(Serrano-Dueñas & Serrano, 2007)
(Local Spanish of Argentina and Ecuador)	SCOPA-PS	(Virués-Ortega et al., 2009)
		(Martinez-Martin et al., 2009)
Persian		
	PDQ-39	(Nojomi, Mostafavian, Shahidi, & Jenkinson, 2010)
	SCOPA-PS	(Fereshtehnejad et al., 2014)
Chinese		
	PDQ-39	(Luo, Tan, Li, Soh, & Thumboo, 2005)
(Taiwan)		(Ma, Hwang, & Chen-Sea, 2005)
		(Luo et al., 2010)
		(Zhang & Chan, 2012)
	PDQ-8	(Huang, Hsu, Wang, & Chen, 2011)
	EQ-5D	(Luo, Low, Lau, Au, & Tan, 2009)
	PDQ-8	(Tan, Lau, Au, & Luo, 2007)
	PDQ-8/39	(Chen et al., 2017)
Serbian		
	PDQ-39	(Žiropađa, Stefanova, Potrebić, & Kostić, 2009)
Korean		

	PDQ-39	(Kwon et al., 2013)
		(Park, Sohng, & Kim, 2014)
Filipino		
	PDQ-39	(Suratos, Saranza, Sumalapao, & Jamora, 2018)
Italian		
	PDQ-39	(Galeoto et al., 2018)
	PDQ-8 (data pooled with PDQ 8/39)	(Franchignoni, Giordano, & Ferriero, 2008)
	PDQ-8/39	(Jenkinson & Fitzpatrick, 2006)
Greek		
	PDQ-8	(Katsarou et al., 2004)
Turkish		
	PDQ-8	(Kahraman et al., 2018)
	PDQL	(Dereli et al., 2015)
Japanese		
	PDQ-8/39	(Jenkinson & Fitzpatrick, 2006)
Portuguese		
	PDQ-39	(Carod-Artal, Martinez-Martin, & Vargas, 2007)
		(Jesus-Ribeiro, Vieira, Ferreira, Januário, & Freire, 2017)
	PDQL	(Campos et al., 2011)
		(Jesus-Ribeiro et al., 2017)
	SCOPA-PS	(Carod-Artal et al., 2007)
		(Virués-Ortega et al., 2009)
Guarani		
	SCOPA-PS	(Virués-Ortega et al., 2009)
German		
	SCOPA-PS	(Knudsen, Martinez-Martin, & Deuschl, 2007)
French		
	SCOPA-PS	(Soulas et al., 2016)
Dutch		

	PDQ-39	(Marinus, Visser, Jenkinson, & Stiggelbout, 2007)
Bulgarian		
	PIMS	
		(Todorova & Stambolieva, 2007)
Estonian		
	PDQ-39	(Krikmann, Taba, Lai, & Asser, 2008)

References

Campos, M., Rezende, C. H. A. d., Farnese, V. d. C., da Silva, C. H. M., Morales, N. M. d. O., & Pinto, R. d. M. C. (2011). Translation, Cross-Cultural Adaptation, and Validation of the Parkinson's Disease Quality of Life Questionnaire (PDQL), the "PDQL-BR", into Brazilian Portuguese. *ISRN Neurology, 2011*, 954787-954785. doi:10.5402/2011/954787

Carod-Artal, F. J., Martinez-Martin, P., & Vargas, A. P. (2007). Independent validation of SCOPA–psychosocial and metric properties of the PDQ-39 Brazilian version. *Movement Disorders, 22*(1), 91-98. doi:10.1002/mds.21216

Chen, K., Yang, Y. J., Liu, F. T., Li, D. K., Bu, L. L., Yang, K., . . . Wu, J. J. (2017). Evaluation of PDQ-8 and its relationship with PDQ-39 in China: a three-year longitudinal study. *Health and Quality of Life Outcomes, 15*(1), 170-177. doi:10.1186/s12955-017-0742-5

Dereli, E. E., Yaliman, A., Kuru Colaka, T., Cakmak, A., Razak Ozdincler, A., & Badilli Demirbas, S. (2015). Turkish Version Study of "Parkinson's Disease Quality of Life Questionnaire" (PDQL). *Noropsikiyatri Arsivi-Archives of Neuropsychiatry, 52*(2), 128-132. doi:10.5152/npa.2015.7359

Fereshtehnejad, S.-M., Farhadi, F., Hadizadeh, H., Shahidi, G. A., Delbari, A., & Lökk, J. (2014). Cross-cultural validity, reliability, and psychometric properties of the persian version of the scales for outcomes in Parkinson's disease-psychosocial questionnaire. *Neurology Research International, 2014*, 260684. doi:10.1155/2014/260684

Franchignoni, F., Giordano, A., & Ferriero, G. (2008). Rasch Analysis of the Short Form 8-Item Parkinson's Disease Questionnaire (PDQ-8). *Quality of Life Research, 17*(4), 541-548. doi:10.1007/s11136-008-9341-6

Galeoto, G., Colalelli, F., Massai, P., Berardi, A., Tofani, M., Pierantozzi, M., . . . Fabbrini, G. (2018). Quality of life in Parkinson's disease: Italian validation of the Parkinson's

Disease Questionnaire (PDQ-39-IT). *Neurological Sciences, 39*(11), 1903-1909. doi:10.1007/s10072-018-3524-x

Huang, T.-T., Hsu, H.-Y., Wang, B.-H., & Chen, K.-H. (2011). Quality of life in Parkinson's disease patients: validation of the Short-Form Eight-item Parkinson's Disease Questionnaire (PDQ-8) in Taiwan. *Quality of Life Research, 20*(4), 499-505. doi:10.1007/s11136-010-9777-3

Jenkinson, C., & Fitzpatrick, R. (2006). Cross-cultural evaluation of the short form 8-item Parkinson's Disease Questionnaire (PDQ-8): Results from America, Canada, Japan, Italy and Spain. *Parkinsonism and Related Disorders, 13*(1), 22-28. doi:10.1016/j.parkreldis.2006.06.006

Jesus-Ribeiro, J., Vieira, E., Ferreira, P., Januário, C., & Freire, A. (2017). Reliability and Validity of 39-Item Parkinson's Disease Questionnaire and Parkinson's Disease Quality of Life Questionnaire. *Acta Medica Portuguesa 30*(5), 395. doi:10.20344/amp.8202

Kahraman, T., Genc, A., Soke, F., Goz, E., Donmez Colakoglu, B., & Keskinoglu, P. (2018). Validity and Reliability of the Turkish Version of the 8-Item Parkinson's Disease Questionnaire. *Noropsikiyatri Arsivi-Archives of Neuropsychiatry, 55*(4), 337-340. doi:10.5152/npa.2017.19343

Katsarou, Z., Bostantjopoulou, S., Peto, V., Kafantari, A., Apostolidou, E., & Peitsidou, E. (2004). Assessing quality of life in Parkinson's disease: Can a short-form questionnaire be useful? *Movement Disorders, 19*(3), 308-312. doi:10.1002/mds.10678

Knudsen, K., Martinez-Martin, P., & Deuschl, G. (2007). Evaluation of the reliability and validity of the German version of a questionnaire on the psychosocial impairment of patients with Parkinson's Disease (SCOPA-PS scales for outcome of Parkinson's Disease, a short psychosocial questionnaire for patients with Parkinson's Disease). *Aktuelle Neurologie, 34*(5), 267-271. doi:10.1055/s-2007-970843

Krikmann, U., Taba, P., Lai, T., & Asser, T. (2008). Validation of an Estonian version of the Parkinson's disease Questionnaire (PDQ-39). *Health and Quality of Life Outcomes, 6*(1), 23-23. doi:10.1186/1477-7525-6-23

Kwon, D. Y., Kim, J. W., Ma, H. I., Ahn, T. B., Cho, J., Lee, P. H., . . . Koh, S. B. (2013). Translation and Validation of the Korean Version of the 39-Item Parkinson's Disease Questionnaire. *Journal of Clinical Neurology, 9*(1), 26-31. doi:10.3988/jcn.2013.9.1.26

Luo, N., Low, S., Lau, P. N., Au, W. L., & Tan, L. C. S. (2009). Is EQ-5D a Valid Quality of Life Instrument in Patients With Parkinson's Disease? A Study in Singapore. *Annals Academy of Medicine Singapore, 38*(6), 521-528. Retrieved from https://pdfs.semanticscholar.org/303f/1a81903b6865c88c8fc63273edb1fe9915be.pdf?_ga=2.9492469.1535524533.1593692297-1460634777.1593692297

Luo, N., Tan, L. C. S., Li, S. C., Soh, L. K., & Thumboo, J. (2005). Validity and Reliability of the Chinese (Singapore) Version of the Parkinson's Disease Questionnaire (PDQ-39). *Quality of Life Research, 14*(1), 273-279. doi:10.1007/s11136-004-2654-1

Luo, W., Gui, X. H., Wang, B., Zhang, W. Y., Ouyang, Z. Y., Guo, Y., . . . Ding, M. P. (2010). Validity and reliability testing of the Chinese (mainland) version of the 39-item Parkinson's Disease Questionnaire (PDQ-39). *Journal of Zhejiang University-Science B, 11*(7), 531-538. doi:10.1631/jzus.B0900380

Ma, H.-I., Hwang, W.-J., & Chen-Sea, M.-J. (2005). Reliability and Validity Testing of a Chinese-Translated Version of the 39-Item Parkinson's Disease Questionnaire (PDQ-39). *Quality of Life Research, 14*(2), 565-569. doi:10.1007/s11136-004-0687-0

Marinus, J., Visser, M., Jenkinson, C., & Stiggelbout, A. M. (2007). Evaluation of the Dutch version of the Parkinson's Disease Questionnaire 39. *Parkinsonism and Related Disorders, 14*(1), 24-27. doi:10.1016/j.parkreldis.2007.05.005

Martinez-Martin, P., Carroza-Garcia, E., Frades-Payo, B., Rodriguez-Blazquez, C., Forjaz, M. J., de Pedro-Cuesta, J., . . . Grupo, E. (2009). Psychometric attributes of the sclaes for outcomes in Parkinson's Disease - Psychosocial (SCOPA-PS): Validation in Spain and review. *Revista de Neurologia, 49*(1), 1-7. doi:10.33588/rn.4901.2009014

Martinez-Martin, P., Serrano-Duenas, M., & Vaca-Baquero, V. (2005). Psychometric characteristics of the Parkinson's disease questionnaire (PDQ-39)—Ecuadorian version. *Parkinsonism and Related Disorders, 11*(5), 297-304. doi:10.1016/j.parkreldis.2005.02.003

Nojomi, M., Mostafavian, Z., Shahidi, G. A., & Jenkinson, C. (2010). Quality of life in patients with Parkinson's disease: Translation and psychometric evaluation of the Iranian version of PDQ-39. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences, 15*(2), 63-69. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3082793/

Park, H. J., Sohng, K. Y., & Kim, S. (2014). Validation of the Korean version of the 39-Item Parkinson's Disease Questionnaire (PDQ-39). *Asian Nursing Research, 8*(1), 67-74. doi:10.1016/j.anr.2014.02.004

Serrano-Dueñas, M., Martínez-Martín, P., & Vaca-Baquero, V. (2004). Validation and cross-cultural adjustment of PDQL-questionnaire, Spanish version (Ecuador) (PDQL-EV). *Parkinsonism and Related Disorders, 10*(7), 433-437. doi:10.1016/j.parkreldis.2004.05.002

Serrano-Dueñas, M., & Serrano, S. (2007). Psychometric characteristics of PIMS—Compared to PDQ-39 and PDQL—To evaluate quality of life in Parkinson's disease patients: Validation in Spanish (Ecuadorian style). *Parkinsonism and Related Disorders, 14*(2), 126-132. doi:10.1016/j.parkreldis.2007.07.006

Soulas, T., Storme, M., Martínez-Martín, P., Pichlak, M., Gurruchaga, J. M., Palfi, S., & Fénelon, G. (2016). Assessing health-related quality of life with the SCOPA-PS in French individuals with Parkinson's disease having undergone DBS-STN: A

validation study. *Revue Neurologique, 172*(4-5), 281-288. doi:10.1016/j.neurol.2015.10.010

Suratos, C. T. R., Saranza, G. R. M., Sumalapao, D. E. P., & Jamora, R. D. G. (2018). Quality of life and Parkinson's disease: Philippine translation and validation of the Parkinson's disease questionnaire. *Journal of Clinical Neuroscience, 54*, 156-160. doi:10.1016/j.jocn.2018.06.013

Tan, L. C. S., Lau, P.-N., Au, W.-L., & Luo, N. (2007). Validation of PDQ-8 as an independent instrument in English and Chinese. *Journal of the Neurological Sciences, 255*(1), 77-80. doi:10.1016/j.jns.2007.01.072

Todorova, A., & Stambolieva, K. (2007). Validation of the Bulgarian version of Parkinson's Impact Scale (PIMS) as a measure of quality of life for patients with Parkinson's disease. *Acta Medica Bulgarica, 34*(2), 35-40. Retrieved from https://www.researchgate.net/publication/289802755_Validation_of_the_Bulgarian_version_of_Parkinson's_Impact_Scale_PIMS_as_a_measure_of_quality_of_life_for_patients_with_Parkinson's_disease

Virués-Ortega, J. P., Carod-Artal, F. J. M. D. P., Serrano-Dueñas, M. M. D., Ruiz-Galeano, G. M. D., Meza-Rojas, G. M. D., Velázquez, C. M. D., . . . Martínez-Martín, P. M. D. P. (2009). Cross-Cultural Validation of the Scales for Outcomes in Parkinson's Disease-Psychosocial Questionnaire (SCOPA-PS) in Four Latin American Countries. *Value in Health, 12*(2), 385-391. doi:10.1111/j.1524-4733.2008.00436.x

Zhang, J.-L., & Chan, P. (2012). Reliability and validity of PDQ-39: a quality-of-life measure for patients with PD in China. *Quality of Life Research, 21*(7), 1217-1221. doi:10.1007/s11136-011-0026-1

Žiropađa, L., Stefanova, E., Potrebić, A., & Kostić, V. S. (2009). Quality of Life in Serbian Patients with Parkinson's Disease. *Quality of Life Research, 18*(7), 833-839. doi:10.1007/s11136-009-9500-4

Appendix I. Risk of bias additional information by measurement property

Structural validity: Five studies were deemed inadequate based on sample size, i.e. they did not have seven times the number of items in the instrument as specified by COSMIN. Six were rated as doubtful due to uncertainties about sample size included for factor analysis in the paper, and lack of clarity about which rotation method was used.

Reliability: COSMIN guidance of at least 2 weeks/14 days between first and second administration of an instrument was followed. In addition, many HRQoL measures such as the PDQ-39 were designed to measure HRQoL in the past month. Therefore, the time limit for this review deemed 'very good' was between 2-4 weeks.

There were 12 data points for reliability which were rated as 'inadequate'. Reasons for this were a change of environment between administrations, typically this was reported when an instrument was self-administered at home as opposed to the clinic/hospital. This may therefore result in increased missing data and different responses than the original questionnaire as unlike in the clinic/hospital setting, any participant queries cannot be clarified by a health care professional. Another reason was the length of time between administrations which typically was reported as being less than 14 days for some papers.

Appendix J. Author guidelines for submission to the Parkinson's Disease Journal

Research Reports

Organization and style of presentation

Manuscripts must be written in US English. Authors whose native language is not English are recommended to seek the advice of a native English speaker or English language service before submitting their manuscripts. A language or editing service that we recommend is PeerWith.

Manuscripts should be double spaced throughout with wide margins (2.5cm or 1in), including the abstract and references. Every page of the manuscript, including the title page, references, tables, etc., should include a page number centered at the bottom.

Manuscripts should be organized in the following order with headings and subheadings typed on a separate line, without indentation.

Title Page

1. Title (should be clear, descriptive and concise).
2. Full name(s) of author(s).
3. Full affiliation(s). Delineate affiliations with lowercase letters.
4. Present address of author(s), if different from affiliation.
5. Running title (45 characters or less, including spaces).
6. Complete correspondence address, including telephone number, fax number and e-mail address.

Leave the author information blank if double-blind peer review is wished for, but do include the information in the submission letter to the editor.

Abstract and Keywords

The abstract for research papers should follow the "structured abstract" format:

BACKGROUND:

OBJECTIVE:

METHODS:

RESULTS:

CONCLUSIONS:

The abstract should try to be no longer than 250 words.

For other papers such as Reviews, the abstract should be clear, descriptive, and self-explanatory, and no longer than 250 words.

Include a list of 4-10 keywords. These keywords should be terms from the MeSH database.

Introduction

Materials and Methods

There is no word limit to the materials and methods section, as the journal's policy is that methodological rigour and reproducibility is of great importance.

Results

Discussion

Acknowledgments including sources of support

Conflict of Interest

If there is no conflict of interest to declare, do still include this section and insert "The authors have no conflict of interest to report".

Appendix K. Copy of R&D approval

Amended 15/10/19 – additional investigators added

NHS Ayrshire & Arran

Research & Development
56a Lister Street
University Hospital Crosshouse
Kilmarnock
KA2 0BB

Dr Breda Cullen
Lecturer
University of Glasgow
Mental Health and Wellbeing
Admin Building
Gartnavel Royal Hospital
Glasgow
G12 0XH

Date	13 September 2019 (amended 15 October 2019)
Your Ref	
Our Ref	CM/KLB/NM R&D 2019AA030
Enquiries to	Karen Bell
Extension	25850
Direct line	01563 825850
Fax	01563 825806
Email	Karen.Bell2@aapct.scot.nhs.uk

Dear Dr Cullen

Validation of Distress Thermometer with People with Parkinson's

I confirm that NHS Ayrshire and Arran have reviewed the undernoted documents and grant R&D Management approval for the above study.

Documents received:

Document	Version	Date
OID	1.2	13/09/19
Protocol	3.0	27/05/19
R&D Form	5.12	13/06/19
Additional Questionnaire	2.0	10/06/19
Distress Thermometer	1.0	10/06/19
Consent Form	5.0	04/07/19
PIS	6.0	04/07/19
HADS	No version	No date
Information Letter for Clinical Team	1.0	01/07/19

The terms of approval state that the investigator authorised to undertake this study within NHS Ayrshire & Arran is: -

Paula Hewat, Parkinsons Disease Specialist, NHS Ayrshire & Arran

With additional investigators:-
- Miss Bronagh Reynolds, Trainee Clinical Psychologist, NHS Ayrshire & Arran / University of Glasgow
- Nicholas Bryden, Parkinson's Nurse Specialist, NHS Ayrshire & Arran
- Amie Walker, Assistant Psychologist, NHS Ayrshire & Arran

The sponsors for this study are University of Glasgow.

Email from R&D dated 20.02.20

Dear Bronagh

2019AA030 Validation of the Distress Thermometer with People with Parkinson's Disease

I have received the documents below relating to a minor amendment to the above study:

- Notification of Non Substantial Amendment
- Sponsor Confirmation
- Samantha Ross CV

I can confirm that the above amendment has been acknowledged and given continued R&D permission.

I can also confirm that Samantha Ross has now been added as an investigator for the study.

A letter is no longer issued for non-substantial amendments. Please retain this email for your records.

Kind regards

Colin Irving
R&D Assistant
NHS Ayrshire & Arran
Research & Development
56A Lister Street
University Hospital Crosshouse
Kilmarnock
KA2 0BE

Appendix L. Ethics approval letter

North of Scotland Research Ethics Committee
Summerfield House
2 Eday Road
Aberdeen
AB15 6RE

Telephone: 01224 558458
Facsimile: 01224 558609
Email: nosres@nhs.net

NHS Grampian

02 July 2019

Dr Breda Cullen
Gartnavel Royal Hospital
Administration Building, 1st floor
1055 Great Western Road
GLASGOW
G12 0XH

Dear Dr Cullen

Study title:	**Validation of the Distress Thermometer with People with Parkinson's Disease**
REC reference:	**19/NS/0112**
Protocol number:	**262666**
IRAS project ID:	**262666**

The Proportionate Review Sub-committee of the North of Scotland Research Ethics Committee 1 reviewed the above application on 02 July 2019.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Number	Conditions
1	The Participant Information Sheet should include a copy of the reply slip with options to tick Yes or No to the statement "I am interested in discussing this study with the researcher at the clinic".
2	In order to make the process clearer to the participant, the Participant Information Sheet should include the following text.

	"Although we do not expect that participating in this study will cause you any distress, if you express distress when speaking to the researcher or through your responses on the study questionnaires, we will help you to access appropriate support if needed. This would involve passing on some information to the Parkinson's disease team, in line with usual procedures."

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

North of Scotland Research Ethics Committee
Summerfield House
2 Eday Road
Aberdeen
AB15 6RE

Telephone: 01224 558458
Email: nosres@nhs.net

NHS Grampian

17 July 2019

Miss Bronagh Reynolds
Trainee Clinical Psychologist
NHS Ayrshire and Arran
Institute of Health and Wellbeing, Gartnavel Royal Hospital
Administration Building, 1st floor
1055 Great Western Road
G12 0XH

Dear Miss Reynolds

Study title:	**Validation of the Distress Thermometer with People with Parkinson's Disease**
REC reference:	**19/NS/0112**
Protocol number:	**262666**
IRAS project ID:	**262666**

Thank you for your letter of 16 July 2019. I can confirm the REC has received the documents listed below and that these comply with the approval conditions detailed in our letter dated 02 July 2019

Documents received

The documents received were as follows:

Document	Version	Date
Movement Disorder Clinic Letter	1	01 July 2019
Participant consent form	5	04 July 2019
Participant information sheet	6	04 July 2019

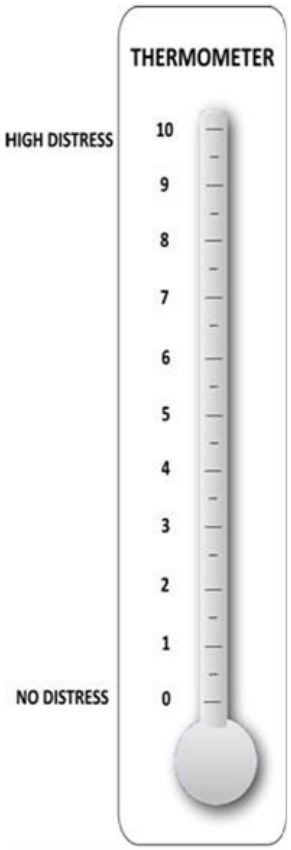
Approved documents

The final list of approved documentation for the study is therefore as follows:

Appendix M. Distress Thermometer and problem list.

IRAS Project ID 262666 V1 10.06.19
Participant ID: Pre/Post:
Date completed:

1) PLEASE CIRCLE THE NUMBER (0-10) that best describes in general how much distress you have been experiencing over the past week, including today.



2) If any of the following has been a problem *for you over the past week*, including today, please tick the box next to it. Leave it blank if it does not apply to you.

Physical Problems:
- ☐ Your Appearance
- ☐ Bathing or Dressing
- ☐ Dribbling Saliva
- ☐ Swallow problems
- ☐ Eating/Appetite
- ☐ Change in weight
- ☐ Sore/Dry mouth
- ☐ Eating or Appetite
- ☐ Nausea/Vomiting
- ☐ Urinary problems
- ☐ Bowel problems
- ☐ Sleep problems
- ☐ Nightmares
- ☐ Acting out in sleep
- ☐ Need to move legs at night
- ☐ Day time sleepiness
- ☐ Fatigue or Tiredness
- ☐ Swollen legs
- ☐ Pain
- ☐ Sweats
- ☐ Sexual concerns
- ☐ Taking medication

Motor issues:
- ☐ Tremor
- ☐ Fine motor control
- ☐ Walking
- ☐ Stiffness
- ☐ Weakness
- ☐ Freezing
- ☐ Bed/Chair mobility
- ☐ Falls

Cognitive Issues:
- ☐ Memory
- ☐ Speed of thinking
- ☐ Concentration& attention
- ☐ Judging distance/Space
- ☐ Odd bizarre behaviour
- ☐ Impulsive

Practical Problems:
- ☐ Caring responsibilities
- ☐ Finances, work
- ☐ Housing
- ☐ Transport/Driving

Family Problems:
- ☐ Relationship with children
- ☐ Relationship with partner
- ☐ Relationship with relatives / friends
- ☐ Burden (on family, friends etc)

Emotional Problems:
- ☐ Sadness or depression
- ☐ Loneliness or isolation
- ☐ Hopelessness
- ☐ Worry, fear or anxiety
- ☐ Loss of control or freedom
- ☐ Anger or frustration
- ☐ Seeing/Hearing things not there

Spiritual Problems:
- ☐ Spiritual concerns
- ☐ Religious concerns
- ☐ Anything else

Any other problems not mentioned above?
_____

3) If you ticked any boxes above, please choose *three main* problems *you are having*

1._____2._____3._____

Appendix N. Additional Questionnaire

All questions are optional. Not answering these questions will not affect your participation in the study.

Q1. Age

1. 18-34 years old ☐
2. 35-44 years old ☐
3. 45-54 years old ☐
4. 55-64 years old ☐
5. 65-74 years old ☐
6. 75 years or older ☐

Q2. Gender

1. Male ☐
2. Female ☐
3. Other ☐

PLEASE TURN OVER

Q3. Relationship status

1. Single ☐
2. Married/partner/cohabiting ☐
3. Widowed ☐
4. Other ☐

Q4. Living arrangements

1. Alone ☐
2. With family/friend ☐
3. Supported/sheltered accommodation ☐
4. Nursing/care home ☐

Q5. Employment status

1. Employed/self-employed ☐
2. Retired ☐

3. Unemployed ☐

PLEASE TURN OVER

Q6a. Is your Distress Thermometer score after the clinic appointment different from your Distress Thermometer score before the clinic appointment?

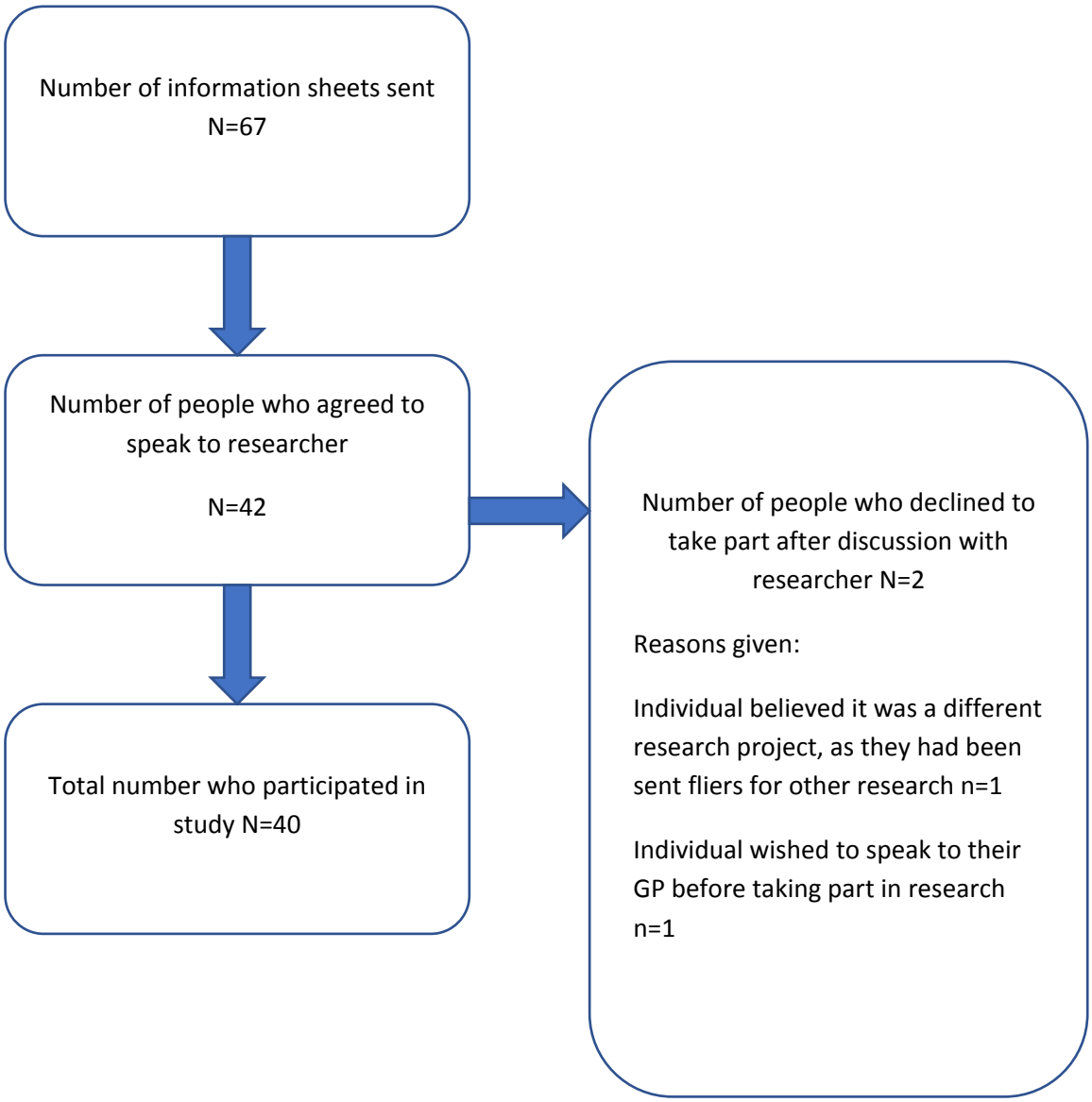
1. Yes ☐
2. No ☐
3. Not sure ☐

Q6b. If yes, can you give us some information about why the score has changed? Please write your answer in the box.

Appendix O. Additional information regarding recruitment.

A discussion with staff including Psychiatry at the Movement Disorder Clinic revealed that there are approximately 600 people are diagnosed with PD in Ayrshire and Arran with 208 patients on regular review in South Ayrshire MDC. The majority of these diagnoses were IPD and would therefore meet the study eligibility criteria. It was estimated by MDC that during the planned 12-month recruitment period, 208 patients would have scheduled appointments. Given the low DNA (12%) rate recorded at the clinic, the required sample size was therefore thought to be feasible for this study.

Appendix P. Flowchart of recruitment



Appendix Q. Major Research Project Protocol

Validation of the Distress Thermometer with People with Parkinson's Disease

27th May 2019

Version: V3

Abstract

Background. The Distress Thermometer (DT) was developed as a tool to detect distress in prostate cancer patients and has since been validated in various medical conditions and age groups. Idiopathic Parkinson's disease (IPD) is a progressive neurological condition that causes a variety of motor symptoms (such as tremor, rigidity, and slowness of movement) and non-motor symptoms including cognitive and mood disorders, nerve pain, and sleep disturbances. To date, no studies have assessed the success of the DT at detecting distress in this population. ***Aims.*** The primary aim of the proposed study is to determine if the DT is a valid measure to detect distress in patients with IPD. ***Method.*** This is a prospective observational study. Participants with IPD will be recruited from the Movement Disorder Clinic in NHS Ayrshire & Arran. The DT will be administered with the Hospital Anxiety and Depression Scale used as a comparator. To assess test-retest reliability of the DT, this will be measured before and after clinic appointment. ***Applications.*** If the DT is found to be a valid measure to detect distress in people with IPD it may encourage clinicians to use DT to assess distress in this population in a busy clinical setting.

Introduction

Distress is a multi-faceted concept pertaining to a negative or unpleasant experience which may be psychological (i.e. cognitive, behavioural, emotional), social, and/or spiritual in nature (Holland & Bultz, 2007). It is documented in the literature as impacting negatively on

medical patients' quality of life and can act as a barrier to seeking appropriate treatment (Ransom, Jacobson & Booth-Jones, 2006).

The importance of screening for distress in medical populations is also documented in the literature (Donovan et al., 2014) and is currently assessed for as part of routine clinical practice for many physical health conditions.

One commonly used measure is the Distress Thermometer (DT), developed by Roth et al. (1998). The DT is a visual analogue scale ranging from 0-10 with individuals asked to indicate which number best represents their levels of distress in the past week, with 0 indicating no distress and 10 indicating extreme distress. After rating their overall distress level, patients are then asked to indicate the source of this distress under domains including practical, family and/or emotional problems, by completing the accompanying problem list (PL). Although the DT has been assessed as a valuable tool in various clinical populations, it is used in populations which have not yet validated the DT as being a suitable screening tool.

Idiopathic Parkinson's disease (IPD) is a progressive neurological condition that causes a variety of specific motor symptoms and non-motor symptoms (Park & Stacy 2009). In people with IPD, psychological distress including anxiety and depression has been shown to be under reported in this population and to impact negatively on quality of life (e. g. Chen & Marsh, 2014). Impairments in instrumental activities of daily living such as driving and ability to manage finances are also associated with IPD (Martin et al., 2013) and may result in elevated distress levels.

The Movement Disorders Clinic (MDC) in NHS Ayrshire & Arran (NHSA&A) introduced the DT with an adapted problem list (MD-PL) (Appendix A) in 2017. Prior to the introduction of the DT, distress was not routinely enquired about in MDCs, with the onus on the patient to divulge issues related to distress. The DT has been deemed to be a useful brief

screening tool for use in busy clinical settings (Snowden et al., 2011) for a variety of cancers and clinical conditions across the lifespan; however, in an IPD population which has a different genesis and trajectory, its utility for identifying distress has not yet been established.

Validation studies of the DT in other clinical populations have used various comparators to assess validity. Although there are examples of studies using a screening tool as the comparator, along with a diagnostic tool e.g. the Composite International Diagnostic Interview (CIDI) used by Patel et al (2011); using the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983) alone for comparator analysis against the DT is common in the literature (Zwahlen et al., 2008; San Giorgi et al., 2016; van der Geest et al., 2018; Campbell et al., 2009). There are also studies using the HADS or alternative screening tools such as The Patient Health Questionnaire–9, to assess for affective syndromes specifically (e.g. Hegel et al., 2008); however, the HADS is typically used as comparator of the DT to compare it as a useful tool to detect psychosocial and physical distress more generally.

Using a tool such as the CIDI or DSM criteria to assess construct validity would not be appropriate in the current study as the primary aim is not to identify clinical disorders using the DT in an IPD population. However, it may be useful to examine classification accuracy against the caseness threshold on the HADS, as a secondary objective. Although this study is not a diagnostic accuracy study, in that distress itself is not a diagnosis, a receiver operating characteristics (ROC) curve can be used to identify the optimal cut-off point on the DT for capturing clinically significant distress (HADS caseness), maximising both sensitivity and specificity, in line with methods outlined by the Standards for reporting of Diagnostic Accuracy Studies (STARD) (Bossuyt et al., 2003).

Currently, patients are asked to complete the DT before their consultation appointment at the MDC. After their consultation, it is possible that some patients may be reassured about their difficulties and this may be reflected in repeated DT scores. This possibility will be investigated in the proposed study by administering the DT both before and after the patients' appointment at the MDC. If distress levels are lowered after the appointment, this may indicate that the DT registers transient distress levels which are likely to resolve after routine consultation with a medical professional.

Lastly, the proposed study will examine the problem list to identify its usefulness as a potential explanation of participants' DT scores.

Aims and Hypotheses

The main aim of this study is to examine whether the ultra-short screen, the DT, is a valid measure to assess distress in an IPD population against another screening tool, the HADS.

Primary research question:

- Is the distress thermometer a valid measure to detect distress in patients with Idiopathic Parkinson's Disease?

Secondary research questions:

- What is the optimal cut-off point on the DT to accurately classify presence of distress?
- Does the DT have appropriate test re-test reliability in IPD patients when measured prior to and after a medical appointment relating to their IPD care?
- What are the most common problems endorsed on the MD-PL in this population?
- What is the correlation between the DT score and total number of problems overall?

- What is the correlation between the DT score and number of problems in each separate domain?

Hypotheses for primary research question

In line with other studies the primary research hypothesis will be based on the correlation between the DT and the HADS.

H0: The DT will not demonstrate appropriate construct validity to detect distress in IPD patients when the HADS is used as comparator. Specifically, the correlation coefficient will not be reliably different from zero, so that the 95% confidence interval will include zero.

H1: The DT will demonstrate appropriate construct validity to detect distress in IPD patients when the HADS is used a comparator. Specifically, the DT score will have a positive correlation with either the HADS anxiety or depression score. Furthermore, a correlation coefficient of 0.3 or higher will be interpreted as clinically meaningful.

Plan of Investigation

Participants

Patients with IPD will be recruited from the NHSA&A MDC.

Inclusion and Exclusion criteria

Inclusion criteria: Diagnosis of Idiopathic Parkinson's Disease; registered with a Movement Disorder Consultant in the Movement Disorders Clinic; able to give informed consent; able to understand and respond to the study questionnaires (carer/staff may assist with writing if necessary); aged 18 years or older.

Exclusion criteria: Other comorbid neurodegenerative diseases.

Recruitment Procedures

See Research Procedures section below.

Measures

Distress Thermometer & Movement Disorder Problem List

See appendix A. The adapted PL was developed in consultation with the MD Consultant, Community Nurse Specialists, Pharmacy and Clinical Psychologists.

Hospital Anxiety and Depression Scale

The DT thermometer will be validated using a licensed version of the HADS. The HADS is a commonly used self-report measure and is recommended by SIGN guidelines (2010) as a screen for depression in patients with IPD and has demonstrated good internal reliability and validity (Bjelland et al., 2002). The HADS consists of 14 items (7 Depression and 7 Anxiety) each rated from 0 to 3 with higher scores indicating increased severity. Subscale totals are calculated, and an individual can score between 0 and 21 for either subscale. Scores of 11 and above indicate depression/anxiety is present.

Additional Information

A brief questionnaire asking for information on employment status, marital status, and feedback on their DT responses. Consent will also be obtained to gather other routine clinical information such as age and clinical history from medical records.

Design

Prospective observational study.

Research Procedures

Relevant professionals will be approached to be involved in the research to identify potential participants. A participant information sheet with details of the study will be sent out to potential participants several weeks in advance of their appointment. The DT (time point one) will be filled out in the reception area of the MDC prior to their appointment, as is current practice within the MDC.

As part of their consultation the medical professional will again inform them of the study and ask if they are willing to participate. Prior to the introduction of the DT in the MDC, distress was not routinely enquired about, and so all eligible participants will be asked to take part in the study regardless of DT and PL responses or clinical discussions during their appointments. If they wish to take part, the researcher will take them to a pre-booked room in the clinic to complete questionnaires with the total time estimated to be 25 minutes. Once participants have read through the participant information sheet again if they wish to do so, and signed the consent form, they will then be asked to fill out the DT again (time point two), and the HADS . They will then be asked to complete a brief questionnaire to collect additional information.

If participants are unable to complete the questionnaires immediately after their appointment, the consent form will be signed in clinic and participants will be given the measures to complete in a pre-paid postage envelope. An arrangement will be made to phone them within two working days to check if they have any queries about completing them. Due to the timescale required for assessing test retest reliability of the DT, for those whom a phone call is not feasible due to speech impairment/do not have a telephone in their home, a home visit within two working days will be arranged. If participants complete the DT at a later date, for example, at home after the clinic, this will be addressed as a limitation of the study in assessing test retest reliability of the DT. This is due to the possibility that other variables

have interacted with the participant's distress levels apart from clinician assurance, with the risk of this increasing if participants do not complete time point two DT at the clinic.

Data from these measures will be collated onto an SPSS database on the University network. Information on this database will only be available to the Trainee Clinical Psychologist and Chief Investigator conducting the research and will be retained until the study is completion and submission to a scientific publication has been approved. No identifiable participant information will be entered onto this database. Participants will be identified using an ID number, which will be stored separately from research data in a locked filing cabinet in NHS Ayrshire and Arran. These will only be accessible to the Trainee who is conducting the research and will be stored until the study is completion and summaries of the results have been sent to those who opted in to receive this.

Data Analysis

Descriptive statistics will be used to describe the study sample. Primary aim: To investigate the relationship between DT and the HADS (both completed after the medical consultation), Spearman correlation coefficient analysis will be used.

Secondary aims: Receiver operating characteristic (ROC) curve analysis will be conducted to assess the classification performance (specificity and sensitivity) for each score on the DT with the HADS (score ≥11 versus <11) used for comparison. The graphic display of the curve will facilitate the selection of an optimal cut-off score for distress for IPD patients. There is no consensus in the literature as to what constitutes 'good' specificity and sensitivity, as this depends on the clinical purpose of the screen. Youden's index is used to find the cut-off that maximises both sensitivity and specificity, and so this is how the optimum cut-off point will be identified (Akobeng, 2007).

To determine the test retest reliability of the DT, correlation coefficient analysis will be carried out between time point one (before appointment) and time point two (after appointment).

Descriptive statistics for the number and type of problems endorsed on the MD-PL, both before and after the appointment will be reported. Correlations between DT score and number of problems identified by participants will be examined. Due to the large number of correlations , False Discovery Rate (FDR) corrections will be applied to the p values, to ensure that the proportion of false positives is controlled at 0.05 within the analyses for each separate research question (Benjamini & Hochberg, 1995).

Exploratory multiple regression models between problems and DT score, adjusting for demographic and clinical covariates, will be conducted if the final sample size permits this.

Justification of sample size

Apriori power analysis was conducted using G*power 3.1.9.2 (Faul et al.,2009) to determine the required sample size. Analysis was based on the estimated correlation between the DT and HADS measures. Parameters: $r = 0.3$ (medium), $p = 0.025$ (two-tailed; Bonferroni adjustment because HADS anxiety and depression will be analysed separately), power $(1-\beta) = 0.80$ (Cohen, 1992). The rationale for choosing 0.3 as the estimated effect size is based on results from previous studies which found correlations for anxiety subscales ranging from 0.51 to 0.56, and depression subscales from 0.36 to 0.69 (San Giorgi et al., 2017; Zwahlen et al., 2008). Furthermore, a correlation of less than 0.3 would likely be clinically negligible.

Sample size was determined to be 102 participants. Approximately 600 people are diagnosed with PD in Ayrshire and Arran with 208 patients on regular review in South Ayrshire MDC.

It is estimated that during the 12-month recruitment period, 208 patients will be attending the MDC.

Currently, the two largest MDC clinics take place on Monday morning and Friday mornings which, given the primary researchers' University currently allocated timetable, will allow for data collection on these days. The non- attendance (DNA) rate at the clinics was estimated to be 12% by clinician's who are part of the MDC and who are facilitating the proposed research. Given the low DNA rate recorded at the clinic, the required sample size was therefore deemed to be feasible for the proposed study.

At three months into recruitment, participant numbers thus far will be evaluated. If this falls significantly below what would be expected at this time point (i.e. 102 x 0.25 = 26), the primary researcher will liaise with their clinical practice placement at that time to facilitate their attendance at more clinics in order to increase recruitment numbers.

Settings and Equipment

The MDC will be the primary setting for the proposed research. A room will be booked out to allow for completion of measures.

Dissemination of results

Results will be disseminated through scientific publication, conference presentations and the dissertation arising from this research. As noted on the consent form, participants can opt-in to receive a summary of the results from the study if they wish to do so.

Health and Safety Issues

Researcher safety issues

Home visits will be conducted as necessary and both the Lone Working Policy provided by NHS A&A will be followed and the Lone Study Procedure provided by the University of Glasgow.

Patient safety issues

Patients will be informed in the participant information sheet sent out prior to their appointment that they are not required to take part in this study and that not doing so will have no effect on their current or future treatment. Participants who express distress will be followed up in line with the current processes in place in the clinic. Their DT and HADS scores will be communicated to the Parkinson ’s disease team as part of their care pathway. If concerns are raised the clinical team will initiate their routine clinical care and direct referrals if required including accessing neuropsychology and clinical health psychology as part of standard practice.

Ethical Issues

An NHS Research Ethics Committee and the Research and Development Department in NHS A&A will be contacted for approval of the proposed research.

Financial Issues

Item	**Unit Cost (£)**	**Number required**	**Total (£)**
Postage of Information Sheet and DT prior to appointment	0.62	150*	93.00
Postage of pre-paid envelopes	0.62	55** participants (***estimating half the sample posting back their questionnaires)	34.10
Printing Costs	0.05	Information sheet (2 pages) x 150* + DT x 150* (time point 1) + Consent form (2 copies, one for primary researcher and one for participant) x 110** + DT x 110** (time point 2) = 780	39.00
Photocopy costs	0.05	HADS x 110 + Additional Information Questionnaire x 110 = 220	11.00
À4 envelopes box of 250	9.01	1 box	9.01
Total			**186.11**

*At least 102 participants are required for the study. However, if we assume a 70% uptake rate, then approx. 150 people will need to be asked to take part in the study.

**Assuming 110 people take part in the study

*** Due to the increased likelihood if mobility issues, carer/family/support staff waiting for the individual after their clinic appointment and having other appointments/activities to

attend, and the increased likelihood that transportation by ambulance will be required for participants, the choice of completing questionnaires at a later date may be preferred by a large portion of participants.

Proposed Timetable

Application to ethics and R&D: February 2019

Recruitment commencement: April/May 2019

Data analysis: May 2020

Writing-up: June-July 2020

Practical Applications

If the DT is found to be a valid measure to detect distress in people with IPD it may encourage clinicians to use DT to assess distress in this population in a busy clinical setting.

References

Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatrica, 96*(5), 644-647. https://doi.org/10.1111/j.1651-2227.2006.00178.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of Psychosomatic Research, 52*(2), 69-77.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., ... & Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine*, *138*(1), W1-12.

Campbell, A., Steginga, S. K., Ferguson, M., Beeden, A., Walls, M., Cairns, W., & Dunn, J. (2009). Measuring distress in cancer patients: The distress thermometer in an Australian sample. *Progress in Palliative Care*, *17*(2), 61-68. https://doi.org/10.1179/096992609X392259

Chen, J. J., & Marsh, L. (2014). Anxiety in Parkinson's disease: identification and management. *Therapeutic Advances in Neurological Disorders*, *7*(1), 52–59. https://doi.org/10.1177/1756285613495723

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Donovan, K. A., Grassi, L., McGinty, H. L., & Jacobsen, P. B. (2014). Validation of the distress thermometer worldwide: state of the science. *Psycho-Oncology*, *23*(3), 241-250. https://doi.org/10.1002/pon.3430

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Holland, J. C., & Bultz, B. D. (2007). The NCCN guideline for distress management: a case for making distress the sixth vital sign. *Journal of the National Comprehensive Cancer Network*, *5*(1), 3-7. https://doi.org/10.6004/jnccn.2007.0003

Park, A., & Stacy, M. (2009). Non-motor symptoms in Parkinson's disease. *Journal of Neurology*, *256*(3), 293-298. https://doi.org/10.1007/s00415-009-5240-1

Patel, D., Sharpe, L., Thewes, B., Bell, M. L., & Clarke, S. (2011). Using the Distress Thermometer and Hospital Anxiety and Depression Scale to screen for psychosocial morbidity in patients diagnosed with colorectal cancer. *Journal of Affective Disorders*, *131*(1-3), 412-416. https://doi.org/ 10.1016/j.jad.2010.11.014

Ransom, S., Jacobsen, P. B., & Booth-Jones, M. (2006). Validation of the distress thermometer with bone marrow transplant patients. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, *15*(7), 604-612. https://doi.org/10.1080/13607863.2011.562175

Roth, A. J., Kornblith, A. B., Batel-Copel, L., Peabody, E., Scher, H. I., & Holland, J. C. (1998). Rapid screening for psychologic distress in men with prostate carcinoma: a pilot study. *Cancer: Interdisciplinary International Journal of the American Cancer*

Society, *82*(10), 1904-1908. https://doi.org/ https://doi.org/10.1002/(SICI)1097-0142(19980515)82:10<1904::AID-CNCR13>3.0.CO;2-X

San Giorgi, M. R., Aaltonen, L. M., Rihkanen, H., Tjon Pian Gi, R. E., van der Laan, B. F., Hoekstra-Weebers, J. E., & Dikkers, F. G. (2017). Validation of the distress thermometer and problem list in patients with recurrent respiratory papillomatosis. *Otolaryngology–Head and Neck Surgery*, *156*(1), 180-188. https://doi.org/10.1177/0194599816668307

Scottish Intercollegiate Guidelines Network. (2010). *Diagnosis and pharmacological management of Parkinson's disease: a national clinical guideline*. Scottish Intercollegiate Guidelines Network. Retrieved from https://www.sign.ac.uk/assets/sign113.pdf

Snowden, A., White, C. A., Christie, Z., Murray, E., McGowan, C., & Scott, R. (2011). The clinical utility of the distress thermometer: a review. *British Journal of Nursing*, *20*(4), 220-227. https://doi.org/10.12968/bjon.2011.20.4.220

van der Geest, I. M. M., van Dorp, W., Pluijm, S. M. F., & van den Heuvel-Eibrink, M. M. (2018). The distress thermometer provides a simple screening tool for selecting distressed childhood cancer survivors. *Acta Paediatrica*, *107*(5), 871-874. Retrieved from https://doi.org/10.1111/apa.14251

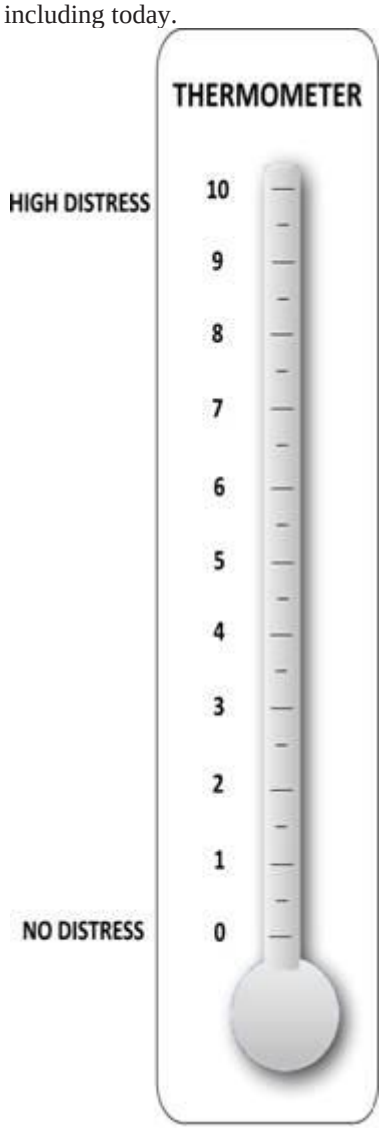
Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361-370.

Zwahlen, D., Hagenbuch, N., Carley, M. I., Recklitis, C. J., & Buchi, S. (2008). Screening cancer patients' families with the distress thermometer (DT): a validation study.

Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer, *17*(10), 959-966. https://doi.org/ 1002/pon.1320

Participant ID_____ Date completed: _____Pre/post:

1) PLEASE CIRCLE THE NUMBER (0-10) that best describes in general how much distress you have been experiencing over the past week, including today.



2) If any of the following has been a problem *for you over the past week*, including today, please tick the box next to it. Leave it blank if it does not apply to you.

Physical Problems:
- ☐ Your Appearance
- ☐ Bathing or Dressing
- ☐ Dribbling Saliva
- ☐ Swallow problems
- ☐ Eating/Appetite
- ☐ Change in weight
- ☐ Sore/Dry mouth
- ☐ Eating or Appetite
- ☐ Nausea/Vomiting
- ☐ Urinary problems
- ☐ Bowel problems
- ☐ Sleep problems
- ☐ Nightmares
- ☐ Acting out in sleep
- ☐ Need to move legs at night
- ☐ Day time sleepiness
- ☐ Fatigue or Tiredness
- ☐ Swollen legs
- ☐ Pain
- ☐ Sweats
- ☐ Sexual concerns
- ☐ Taking medication

Motor issues:
- ☐ Tremor
- ☐ Fine motor control
- ☐ Walking
- ☐ Stiffness
- ☐ Weakness
- ☐ Freezing
- ☐ Bed/Chair mobility
- ☐ Falls

Cognitive Issues:
- ☐ Memory
- ☐ Speed of thinking
- ☐ Concentration& attention
- ☐ Judging distance/Space
- ☐ Odd bizarre behaviour
- ☐ Impulsive

Practical Problems:
- ☐ Caring responsibilities
- ☐ Finances, work
- ☐ Housing
- ☐ Transport/Driving

Family Problems:
- ☐ Relationship with children
- ☐ Relationship with partner
- ☐ Relationship with relatives / friends
- ☐ Burden (on family, friends etc)

Emotional Problems:
- ☐ Sadness or depression
- ☐ Loneliness or isolation
- ☐ Hopelessness
- ☐ Worry, fear or anxiety
- ☐ Loss of control or freedom
- ☐ Anger or frustration
- ☐ Seeing/Hearing things not there

Spiritual Problems:
- ☐ Spiritual concerns
- ☐ Religious concerns
- ☐ Anything else

Any other problems not mentioned above?

_____

3) If you ticked any boxes above, please choose *three main* problems *you would like to discuss* at your appointment today:

1._____**2.**_____**3.**_____

Appendix B

RESEARCH EQUIPMENT, CONSUMABLES AND EXPENSES

Trainee ……

Year of Course ……2nd…year……. **Intake Year**……2017……..

Please refer to latest stationary costs list (available from student support team)

Item	**Details and Amount Required**	**Cost or Specify if to Request to Borrow from Department**
Stationary	A4 envelopes 100 (box of 250 quoted on 'Stationary costs 2015' handout)	Subtotal: £9.01
Postage	0.62 X 205	Subtotal: £127.10
Photocopying and Laser Printing	Information sheet (2 pages) x 150 + DT x 150 (time point 1) + Consent form (2 copies, one for primary researcher and one for participant) x 110 +	Subtotal: £50.00

	DT x 110 (time point 2) = 780 HADS x 110 + Additional Information Questionnaire x 110 = 220 Total =1,000	
Equipment and Software		Subtotal:
Measures		Subtotal:
Miscellaneous		Subtotal:
Total		£186.11

For any request over £200 please provide further justification for all items that contribute to a high total cost estimate. Please also provide justification if costing for an honorarium:

Trainee Signature…… … Date……

Supervisor's Signature …….. Date ……

Appendix C

WEST OF SCOTLAND/ UNIVERSITY OF GLASGOW

DOCTORATE IN CLINICAL PSYCHOLOGY

HEALTH AND SAFETY FOR RESEARCHERS

1. Title of Project	**Validation of the Distress Thermometer with People with Parkinson's Disease**
2. Trainee	
3. University Supervisor	#####
4. Other Supervisor(s)	
5. Local Lead Clinician	
6. Participants: (age, group or sub-group, pre- or post-treatment, etc)	Idiopathic Parkinson's Disease individuals
7. Procedures to be applied (e.g., questionnaire, interview, etc)	Three measures – Distress Thermometer (time point one – before clinic appointment and time point two – after clinic appointment); Hospital Anxiety and Depression Scale (HADS); Brief demographic questionnaire (asking about marital and employment status and whether distress levels have changed since DT time point one and if so, why.

8. Setting (where will procedures be carried out?) i) Details of all settings	Movement Disorder Clinic – room booked in MDC. Home visits – if necessary i.e. participant cannot stay after their MDC appointment.
ii) Are home visits involved	Y

WEST OF SCOTLAND / UNIVERSITY OF GLASGOW

DOCTORATE IN CLINICAL PSYCHOLOGY

HEALTH AND SAFETY FOR RESEARCHERS

9. Potential Risk Factors Considered (for researcher and participant safety): i) Participants ii) Procedures iii) Settings	Potential for violence and aggression from participants. Potential for participants to become fatigued during the course of the study.

10. . 10. Actions to minimise risk (refer to 9) i) Participants ii) Procedures iii) Settings	Participants energy levels will be observed and monitored by the primary researcher For home visits: Both the Lone Working Policy for NHS Ayrshire and Arran and the Lone Study Procedure for students of University of Glasgow will be followed if home visits are necessary. Assess potential risk before going on home visit e.g. if there are any dogs on the premises. Participants who have been seen recently by a member of the clinical team involved with the patient and a risk assessment has been carried out. If the participant has had no recent involvement with a clinical team then a home visit will not be conducted. Discuss potential for risk with a member of the clinical team who has seen the patient recently. If there is doubt regarding level of risk, I will discuss with their University supervisor and/or a senior member of the clinical team that have responsibility for management of the patient. The overall appraisal of risk will take into account what is known about the participant, a risk assessment of their living environment by the clinical team and consideration of the geographical setting of the visit. This will include assessment of any risk associated with travelling to and from the participant's home. Home visits will be undertaken only within normal work hours i.e. 9am -5pm.

Trainee signature: ..... Date: .....

University supervisor signature: ..... Date: .....