The Hopkins Symptom Checklist (SCL-90-R): A Patient-Reported Outcome Measure in Parkinson's Disease

The Hopkins Symptom Checklist (SCL-90-R):

A Patient-Reported Outcome Measure in Parkinson's disease

Danilo Carrozzino, PhD[1], Kaj Sparle Christensen, MD, PhD[2], Chiara Patierno, PhD[1], Chiara Siri, PsyD[3], Anna Zecchinelli, MD[4], Gianni Pezzoli, MD[4,5], and Fiammetta Cosci, MD, PhD[6,7]

[1]Department of Psychology "Renzo Canestrari", Alma Mater Studiorum, University of Bologna, Bologna, Italy

[2]Research Unit for General Practice and Section for General Medical Practice, Department of Public Health, Aarhus University, Aarhus, Denmark

[3]Moriggia Pelascini Hospital, Rehabilitation Unit, Como, Italy

[4]Parkinson Institute, Pini-CTO, Milan, Italy

[5]Fondazione Grigioni per il Morbo di Parkinson, Milan, Italy

[6]Department of Health Sciences, University of Florence, Florence, Italy

[7]Department of Psychiatry & Neuropsychology, Maastricht University, Maastricht, The Netherlands

Corresponding author:

Danilo Carrozzino, PhD

Department of Psychology "Renzo Canestrari",

Alma Mater Studiorum, University of Bologna

Viale Berti Pichat 5, 40127 Bologna (Italy).

E-mail address: danilo.carrozzino@unibo.it

Abstract

Objective: This is the first study applying Clinimetric Patient-Reported Outcome Measures (CLIPROM) criteria to evaluate the construct validity, sensitivity, and clinical utility of the SCL-90-R in patients with Parkinson's disease (PD). ~~patients.~~ **Methods:** A Rasch analysis was conducted ~~and~~ using a sample of 488 PD outpatients. ~~used~~. **Results:** Testing for dimensionality revealed that less than 5% of *t*-tests were significant, indicating that the SCL-90-R subscales entailed the property of construct validity. As to the total score, a Person Separation Reliability Index of 0.96 was found. **Conclusions:** The SCL-90-R total score is a sensitive screening measure that can be used not only to differentiate healthy stress reactions from symptoms of psychological distress but also to detect PD patients with an increased risk for psychiatric complications. As to the subscales, the brief versions that did not include misfitting items should be used to assess the severity of specific symptoms of psychological distress affecting PD patients.

Keywords: Clinimetrics, CLIPROM criteria, Parkinson's disease, Patient-Reported Outcome Measures, Psychological distress, SCL-90-R.

Introduction

Patient-Reported Outcome Measures (PROMs), any report coming directly from patients about how they function or feel in relation to a health condition or its therapy, [1,2] are ~~self-rated scales or~~ self-reported questionnaires that were ~~specifically~~ developed to improve the detection of the subjective ~~burden and~~ impact of symptoms on quality of life, emotional stability, and psychological well-being of patients. [3-5] Since their introduction, there have been an increasing number of studies using PROMs to evaluate symptoms of psychological distress in Parkinson's disease (PD) patients.[6-14] ~~As the use of these assessment instruments proliferated,~~ Greater attention has also been devoted to the evaluation of their measurement properties.[9-11,15-18] Several studies were conducted but most of them focused on classical psychometric criteria rather than on clinimetric principles to evaluate the ~~reliability and~~ validity of PROMs in PD patients.[9,15,17] Clinimetrics is the term originally coined by Alvan R. Feinstein [19-21] to introduce an innovative assessment method that has been later refined as the science of clinical measurements. [22] Clinimetric Patient-Reported Outcome Measures, the CLIPROM criteria, have been recently introduced to evaluate the ~~measurement~~ clinical properties ~~and clinical usefulness~~ of PROMs.[23] Such criteria[23] do not rely on the psychometric assumption of homogeneity of components and represent a step forward to the ~~assessment~~ evaluation of sensitivity, construct validity, and clinical utility of ~~PROMs as they challenge the traditional views of how~~ PROMs these assessment instruments. ~~should be developed, validated, and used.[23]~~ According to CLIPROM criteria, [23] Item Response Theory (IRT) models (i.e., Rasch and Mokken analyses) are used to evaluate the construct validity of PROMs. The clinimetric analysis of construct validity implies the assessment of whether each item provides unique/distinctive clinical information, the evaluation of the extent to which each symptom included in a rating scale belongs to an underlying clinical dimension, and testing if PROMs are statistically sufficient and clinically valid indices of the severity of the clinical condition under examination.[23-29]

Two studies focused on clinimetric principles to evaluate the validity of the revised version of the Hopkins Symptom Checklist (SCL-90-R), one of the most widely used PROMs for the assessment of self-reported symptoms of psychological distress.[10,11] These studies showed that the SCL-90-R was found to be a valid measure of a wide range of psychiatric symptoms (e.g., depression, phobic anxiety, somatization) affecting patients with PD.[10,11] It should be noted, however, that CLIPROM criteria were only partially addressed in the process of assessment of the SCL-90-R: the authors of these studies [10,11] used the Mokken analysis, which is a weaker test than the Rasch one, [24] implying that ~~measurement~~ clinical properties such as the local dependency ~~(i.e., evaluating whether the response to one item was dependent on the response to another item after controlling for the underlying trait)~~ and the differential functioning (~~i.e., testing whether a certain form of item bias can occur when different groups within the sample respond differently to an item despite equal levels of the underlying trait)~~ or transferability of items (e.g., assessing whether the rating scale continues to measure the same clinical dimension across different groups of patients) were not examined. There is, therefore, a need for a clinimetric reanalysis o~~f the sensitivity and construct validity~~ of the SCL-90-R using CLIPROM criteria.[23] ~~in patients with PD.~~

Aims

~~This is the first study using CLIPROM criteria[23] to analyze the clinimetric properties and clinical utility of the SCL-90-R in the process of assessment of symptoms of psychological distress in PD patients.~~ This is the first study in which a Rasch analysis was conducted according to CLIPROM criteria[23] to evaluate ~~the construct validity testing~~ whether the SCL-90-R and its subscales consisted of locally independent and transferable items providing distinctive clinical information. The Rasch analysis was also performed to assess the dimensionality or construct validity of the SCL-90-R evaluating whether its total score and clinical subscales were valid measures of symptoms of psychological distress. The clinimetric sensitivity [23,30] of the SCL-90-R ~~a clinimetric property that refers to the ability of the rating scale to differentiate patients from healthy~~

subjects, to discriminate between different groups of patients suffering from the same illness, to detect changes in clinical (i.e., drug or psychotherapy) trials, to discriminate between an active treatment and placebo, and to differentiate between wanted and unwanted effects of treatments, was also assessed, particularly to evaluate whether the SCL-90-R its ability to discriminate between PD patients displaying different levels of psychological distress.

Material and Methods

Sample and procedure

The sample that is part of the study conducted by Siri et al.[6] consisted of 488 outpatients (mean age of 65.86 ± 10.26 years) with a medical diagnosis of PD formulated according to the UK Brain Bank clinical diagnostic criteria.[31] Data were collected by mailing a self-reported questionnaire, the SCL-90-R, to all outpatients who consulted the Parkinson Institute, ASST "Gaetano Pini CTO", Milan, Italy. An accompanying letter explaining the aims of the study and including a sheet for demographic and clinical data was enclosed with the SCL-90-R. Patients had also the opportunity to call the Institute for further explanations about the study and talk with investigators and clinicians involved in the research project. Patients signed a written informed consent for study participation, filled out the survey, and mailed it back to the Institution. Demented patients with PD were excluded, mainly for their inability to fill out self-reported questionnaires. Therefore, only non-demented PD patients according to the DSM-IV criteria were included in the present study. (mean disease duration of 12.09 ± 5.98 years)

The optimal number of PD patients to recruit was determined using methodological recommendations, which suggest a sample size of at least 100 respondents for performing validation studies on PROMs. [32] The optimal number of patients to be included in the present study was also determined based on a previous research [33] that recommended a sample size ranging from 250 to 500 participants for conducting Rasch analyses. As to the study design, this is a cross-

sectional research ~~study~~ that was conducted according to the Declaration of Helsinki [34] and approved by the Ethical Committee of the Parkinson Institute, ASST "Gaetano Pini CTO", Milan, Italy.

Measure

The Hopkins Symptom Checklist is a self-reported questionnaire evaluating symptoms of psychological distress.[35] The first version of this rating scale was originally developed by Parloff et al.[36] to be used as an outcome measure in psychotherapy trials with depressed or anxious patients. Over the years, a number of revised versions were developed and came into use. [37-41] The revised one, the SCL-90-R, was introduced by Derogatis[42] and consisted of 90 items rated on a 5-point Likert scale ranging from 0 (i.e., "not at all") to 4 (i.e., "extremely"). The patient is asked to report how much a given item (i.e., symptom) distressed or bothered him/her during the last week. [42] The Italian version of the SCL-90-R[43,44] was used in the present study.

Statistical analyses

The Rasch analysis was conducted using the Rasch Unidimensional Measurement Models (RUMM2030) software.[45] The following properties were tested:

1. *The overall fit to the Rasch model*, which was evaluated using the chi-square item-trait interaction statistics.[46,47] The overall fit provided a summary measure of how the scale under assessment conforms to the Rasch model expectations.[25] A non-significant chi-square probability value indicated a good level of overall fit.[46,47]

2. *Individual item and person fit*: standardized fit residual values for items and subjects were examined for any indication of misfit (i.e., values outside ± 2.5).[48]

3. *Construct validity*: to determine whether the SCL-90-R was a valid measure of the underlying construct of psychological distress, Principal Component Analysis (PCA) of residuals was performed to identify the two most different subsets of items (i.e., the most positively and negatively factor-loading items on the first component). Paired *t*-tests were then performed to

compare scores on the two subsets of items. If more than 5% of *t*-tests were significant, the SCL-90-R was not considered unidimensional.[25,27,48]

4. *Local dependency* as the evidence that the response to any item was dependent on the response to another item after controlling for the underlying construct under examination.[48] A residual correlation value of > 0.3 was considered indicative of local dependency between items.[49]

5. *Differential Item Functioning* (DIF) was examined for each item of the SCL-90-R with respect to age (dichotomized in young and old PD patients based on 50% percentile), education (dichotomized in short and long duration based on 50% percentile), and sex. We term it *uniform DIF* when one subgroup (e.g., males) consistently scores differently on an item across all levels of the trait under examination and *non-uniform DIF* when the DIF varies across different levels of the trait.[33,48]

6. *Person Separation Reliability Index* (PSI) was tested to evaluate the clinimetric sensitivity of the SCL-90-R and its subscales.[23,29]

Results

Sample characteristics

The sample consisted of 488 PD outpatients, 58.4% (*n* = 285) were males (mean age of 65.19 ± 10.52 years), and 41.6% (*n* = 203) were females (mean age of 66.81 ± 9.82 years). The clinical characteristics of patients, including the mean age of the total sample, the age at onset, disease duration, the Hoehn and Yahr (H&Y) staging scale, the Unified Parkinson's Disease Rating Scale motor score (UPDRS III) and the daily dose of levodopa, are reported in Table 1.

Fit to the Rasch model

The initial analysis of the SCL-90-R revealed a significant item-trait interaction statistic (χ^2 = 1434.49, degrees of freedom [df] = 630, $p < 0.001$), indicating misfit to the Rasch model (Table 2, Analysis 1). Standardized fit residuals for items (SD = 2.27) were just within acceptable limits.

Standardized fit residuals for persons (SD = 0.74) were found to be within acceptable limits. Rescoring all items, ordered response categories were achieved but without significantly improving the overall fit to the Rasch model (Table 2, Analysis 2). Even after excluding misfitting items (Table 2, Analysis 14), fit to the Rasch model was not achieved for the SCL-90- R total scale (χ^2 = 709.77, df = 546, $p < 0.001$). Model fit statistics for the SCL-90-R subscales are reported in Table 3. The analysis of the SCL-90-R subscales of phobic anxiety, interpersonal sensitivity, hostility, sleep, and psychoticism revealed a significant item-trait interaction statistic, indicating misfit to the Rasch model (Table 3). The initial analysis of the SCL-90-R anxiety subscale showed a significant item-trait interaction statistic (χ^2 = 174.42, df = 70, $p < 0.001$), indicating misfit to the Rasch model. The fit to the Rasch model was achieved (χ^2 = 63.18, df = 56, p = 0.24) after the exclusion of misfitting symptoms (items 17, and 33). The initial analysis of the SCL-90-R depression subscale showed a significant item-trait interaction statistic (χ^2 = 166.14, df = 105, $p < 0.001$), indicating misfit to the Rasch model. The fit to the Rasch model was achieved (χ^2 = 100.33, df = 91, p = 0.23) after the exclusion of misfitting symptoms (items 19, and 5). Rasch analysis of the SCL-90-R somatization subscale revealed a non-significant item-trait interaction statistic (χ^2 = 98.47, df = 84, p = 0.13), indicating adequate fit to the model, with no misfitting items. The initial analysis of the SCL-90-R subscale on cognitive performance deficits showed a significant item-trait interaction statistic (χ^2 = 105.40, df = 63, $p < 0.001$), indicating misfit to the Rasch model. The fit to the Rasch model was achieved (χ^2 = 65.35, df = 56, p = 0.18) after the exclusion of misfitting symptoms (item 65). Rasch analysis of the SCL-90-R subscale on paranoid ideation revealed a non-significant item-trait interaction statistic (χ^2 = 46.52, df = 36, p = 0.11), indicating adequate fit to the model, with no misfitting items.

Dimensionality

Testing for dimensionality revealed significant t-tests outside the critical value of 5%, suggesting that the total score of the SCL-90-R was multidimensional (Table 2, Analysis 1-14).

Concerning the SCL-90-R subscale on phobic anxiety, less than 5% of t-tests were significant, indicating that this subscale was unidimensional. As to the anxiety subscale, after the exclusion of items 17 ("trembling") and 33 ("feeling fearful"), less than 5% of t-tests were significant, suggesting that the 8-item version of the SCL-90-R anxiety subscale was a unidimensional measure. Regarding the depression subscale, after the exclusion of items 19 ("poor appetite") and 5 ("loss of sexual interest or pleasure"), less than 5% of t-tests were significant, suggesting that the 14-item version of the SCL-90-R depression subscale was unidimensional. As to the SCL-90-R somatization subscale, paired t-tests comparisons revealed that more than 5% of t-tests were significant (Table 3), indicating that this subscale was a multidimensional measure. Concerning the subscale on cognitive performance deficits, less than 5% of t-tests were significant, suggesting that the 9- and 8-item versions of this subscale were unidimensional measures. As to the subscale on interpersonal sensitivity, even after excluding the misfitting item (Table 3), more than 5% of t-tests were significant, indicating that this subscale was a multidimensional measure. Concerning the SCL-90-R subscales of hostility, paranoid ideation, and psychoticism, less than 5% of t-tests were significant, suggesting that these measures were unidimensional.

Local dependency

Indication of local dependency between 18 item-pairs of the SCL-90-R was found with residual correlations > 0.30. The following item-pairs were found to be locally dependent: items 7 ("the idea that someone else can control your thoughts") and 35 ("other people being aware of your private thoughts"), items 13 ("feeling afraid in open spaces or on the streets") and 25 ("feeling afraid to go out of your house alone"), items 14 ("feeling low in energy or slowed down") and 56 ("feeling weak in parts of your body"), items 15 ("thoughts of ending your life") and 59 ("thoughts of death or dying"), items 23 ("suddenly scared for no reason") and 33 ("feeling fearful"), items 23 ("suddenly scared for no reason") and 72 ("spells of terror or panic"), items 25 ("feeling afraid to go out of your house alone") and 47 ("feeling afraid to travel on buses, subways, or trains"), items

36 (“feeling others do not understand you or are unsympathetic”) and 37 (“feeling that people are unfriendly or dislike you”), items 43 (“feeling that you are watched or talked about by others”) and 61 (“feeling uneasy when people are watching or talking about you”), items 44 (“trouble falling asleep”) and 66 (“sleep that is restless or disturbed”), items 48 (“trouble getting your breath”) and 53 (“a lump in your throat”), items 51 (“your mind going blank”) and 55 (“trouble concentrating”), items 61 (“feeling uneasy when people are watching or talking about you”) and 69 (“feeling very self-conscious with others”), items 67 (“having urges to break or smash things”) and 81 (“shouting or throwing things”), items 68 (“having ideas or beliefs that others do not share”) and 74 (“getting into frequent arguments”), items 69 (“feeling very self-conscious with others”) and 70 (“feeling uneasy in crowds, such as shopping or at a movie”), items 69 (“feeling very self-conscious with others”) and 73 (“feeling uncomfortable about eating or drinking in public”), items 85 (“the idea that you should be punished for your sins”) and 89 (“feelings of guilt”).

Differential Item Functioning

Items 23 (“suddenly scared for no reason”) and 75 (“feeling nervous when you are left alone”) showed a uniform DIF for education.

Uniform DIF for sex was observed for items 20 (“crying easily”), 21 (“feeling shy or uneasy with the opposite sex”), 27 (“pains in lower back”), 39 (“heart pounding or racing”), 49 (“hot or cold spells”), 84 (“having thoughts about sex that bother you a lot”), 88 (“never feeling close to another person”), and 90 (“the idea that something is wrong with your mind”).

There was no indication of uniform or non-uniform DIF for age.

Person Separation Reliability Index

PSI was 0.96 (Table 2, Analysis 1), indicating that the original SCL-90-R total scale could reliably discriminate between respondents displaying different levels of the underlying trait under examination. PSI indices of the SCL-90-R subscales of phobic anxiety, hostility, sleep, paranoid

ideation, and psychoticism were found to range from 0.41 to 0.61 (Table 3), suggesting that they could not be reliably used to discriminate between different groups of subjects with different levels of the underlying construct. PSI indices of the SCL-90-R subscales on anxiety, depression, somatization, cognitive performance deficits, and interpersonal sensitivity were found to range from 0.74 to 0.82 (Table 3), indicating that these measures could reliably distinguish between different groups but not between different subjects.

Discussion

This is the first study applying CLIPROM criteria [23] to evaluate the clinimetric sensitivity, construct validity, and clinical usefulness of the SCL-90-R in PD patients. The findings of the present study indicate that the SCL-90-R is a multidimensional measure of psychological distress. This is in line with previous studies showing that the total score of this rating scale was found to cover both bodily and emotional components of psychological distress.[10,38,44] It should be also noted that excellent PSI indices were found, indicating that the SCL-90-R sensitively discriminated between PD patients with different levels of psychological distress. This implies that the total score of the SCL-90-R can be used as a screening measure ~~not only~~ to differentiate healthy stress reactions from symptoms of psychological distress ~~but also~~ and to identify PD patients with an increased risk to develop psychiatric complications. Important sources of information may therefore derive from the use of this rating scale in patients with PD. The findings of the present study also indicate that the SCL-90-R consisted of clinical subscales that were found to entail the clinimetric property of construct validity. Such measures were the SCL-90-R subscales of phobic anxiety, anxiety, depression, cognitive performance deficits, hostility, paranoid ideation, and psychoticism. However, caution should be paid when administering these subscales to PD patients as locally dependent, and non-transferable items were detected. As to the SCL-90-R subscale on phobic anxiety, findings demonstrate that the item 75 (“feeling nervous when you are left alone”) showed a uniform DIF with PD patients with a higher education scoring significantly higher than those with a

lower one despite equal levels of the underlying trait of psychological distress. Another major problem of the SCL-90-R phobic anxiety subscale has to do with local dependency or redundancy of items.[23] This measure was found to include locally dependent items (i.e., "feeling afraid in open spaces or on the streets", "feeling afraid to go out of your house alone", "feeling afraid to travel on buses, subways, or trains", "feeling uneasy in crowds, such as shopping or at a movie") that evaluate parallel forms of the same symptom, implying that only a narrow part of the clinical condition under examination can be detected by the use of the SCL-90-R subscale of phobic anxiety. The same problem was observed with the SCL-90-R subscales of anxiety, depression, cognitive performance deficits, hostility, paranoid ideation, and psychoticism, which were found to include a large number of inter-correlated items providing clinically redundant information. The evidence that the SCL-90-R consisted of locally dependent items is not surprising since this measure has been originally developed according to psychometric criteria.[50] In the psychometric model, a statistically significant correlation between items is often required to increase the internal reliability of the rating scale. [51,52] However, as has been widely demonstrated, the redundant nature of items may increase the internal consistency of the rating scale but is likely to decrease its sensitivity.[23,26,28,51-53] This is the main reason why previous studies [10,11,54] suggested using the SCL-90-R subscales in conjunction with highly sensitive clinician-rated scales such as the revised version of the Diagnostic Criteria for Psychosomatic Research (DCPR-R), [55] which was found to be a valid measure of symptoms of psychological distress (i.e., demoralization) in patients with PD. [56-58] Findings of the present study also indicate that the SCL-90-R subscales of anxiety, depression, cognitive performance deficits, interpersonal sensitivity, and psychoticism included misfitting items. As to the SCL-90-R anxiety subscale, in particular, caution should be paid with the use of the misfitting item 33 (i.e., "trembling"), which may reflect motor symptoms of PD rather than pure symptoms of anxiety. This implies that the 8-item version of the SCL-90-R anxiety, which did not include misfitting items, rather than the traditional one consisting of 10 items should be used to detect the subjective burden of pure symptoms of anxiety affecting patients with PD. For the same

reason the 14-item version of the SCL-90-R depression subscale, the 8-item version of the SCL-90-R cognitive performance deficits subscale, and the 6-item versions of the SCL-90-R hostility, paranoid ideation, and psychoticism subscales should be preferred to assess the severity of ~~peculiar~~ specific symptoms of psychological distress in PD patients since these measures were found to not include misfitting items.

The spectrum of psychiatric disorders in PD and the clinical utility of the SCL-90-R

Psychiatric disorders, particularly depression and anxiety, are part of non-motor symptoms of PD and are highly prevalent in this clinical population.[59-63] Psychiatric disturbances represent a significant source of suffering[64] that is so subjective that, as Parloff et al. [36] noted, no one can judge it but the patient. The SCL-90-R and its subscales are easy-to-use and no-time consuming clinimetric indices that can be used not only to detect vulnerability to depression and anxiety in patients with PD but also to support clinicians in the process of assessment of this subjective state of distress. It is, however, important to note that the use of PROMs such as the SCL-90-R should be constantly guided by the clinical reasoning of experienced clinicians. This implies that the SCL-90-R and its subscales can be used jointly with observer-rated scales and clinical judgment to provide a comprehensive assessment of psychiatric disorders affecting patients with PD.

Limitations

The present study has some limitations. First, the cross-sectional design, precluding the evaluation of the incremental and predictive validity of the SCL-90-R. Future studies, using a longitudinal design, are needed. The second limitation is the lack of a control group of healthy subjects, which is useful to further evaluate the clinimetric sensitivity of the SCL-90-R, particularly to test the ability of this rating scale to discriminate PD patients from healthy controls. Third, only a self-reported measure was used, precluding the assessment of the concurrent and clinical validity of

the SCL-90-R. Future studies, making use not only of other self-reported questionnaires but also of clinician-rated scales, are highly encouraged.

Conclusions

Time has come to apply CLIPROM criteria [23] for assessing the clinimetric sensitivity, construct validity, and clinical utility of PROMs to be used in patients with PD. The findings of this study indicate that the SCL-90-R is a highly sensitive clinimetric index that should be considered as an item bank including valid PROMs that can be used to evaluate a wide spectrum of symptoms of psychological distress. It should be noted, however, that the SCL-90-R total score and its subscales were found to entail different clinimetric properties. The total score appears to be particularly useful to discriminate between PD patients displaying different levels of psychological distress. It can be, therefore, used as a first-line screening measure not only to differentiate healthy stress reactions from symptoms of psychological distress but also to detect PD patients with an increased risk for psychiatric complications. As to the SCL-90-R subscales, these measures, particularly the brief versions that did not include misfitting items, should be used to assess the severity or subjective burden of specific symptoms of psychological distress affecting the quality of life and psychological well-being of PD patients.

Acknowledgements: Chiara Siri would like to thank the "Fondazione Grigioni per il Morbo di Parkinson", Milan, Italy.

Declaration of conflicting interests: The authors declare that there is no conflict of interest.

Funding: This research received no specific grant from funding agency in the public, commercial, or not-for-profit sectors.

ORCID IDs:

Danilo Carrozzino https://orcid.org/0000-0002-8164-8671

Kaj Sparle Christensen https://orcid.org/0000-0002-0409-2702

Chiara Patierno https://orcid.org/0000-0002-7043-8126

Anna Zecchinelli https://orcid.org/0000-0002-7092-7034

Fiammetta Cosci https://orcid.org/0000-0002-5022-0488

References

1. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007;45(5 Suppl 1):S3-S11. doi: 10.1097/01.mlr.0000258615.42478.55.

2. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* 2010;63(11):1179-1194. doi: 10.1016/j.jclinepi.2010.04.011.

3. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res.* 2009;18(1):115-123. doi: 10.1007/s11136-008-9430-6.

4. Ciani O, Federici CB. Value lies in the eye of the patients: the why, what, and how of patient-reported outcomes measures. *Clin Ther.* 2020;42(1):25-33. doi: 10.1016/j.clinthera.2019.11.016.

5. Carlton J, Peasgood T, Khan S, Barber R, Bostock J, Keetharuth AD. An emerging framework for fully incorporating public involvement (PI) into patient-reported outcome measures (PROMs). *J Patient Rep Outcomes.* 2020;4:4. doi: 10.1186/s41687-019-0172-8.

6. Siri C, Cilia R, De Gaspari D, et al. Psychiatric symptoms in Parkinson's disease assessed with the SCL-90R self-reported questionnaire. *Neurol Sci.* 2010;31(1):35-40. doi: 10.1007/s10072-009-0165-0.

7. Onofrj M, Thomas A, Tiraboschi P, et al. Updates on Somatoform Disorders (SFMD) in Parkinson's Disease and Dementia with Lewy Bodies and discussion of phenomenology. *J Neurol Sci.* 2011;310(1-2):166-171. doi: 10.1016/j.jns.2011.07.010.

8. Bugalho P, Da Silva JA, Cargaleiro I, Serra M, Neto B. Psychiatric symptoms screening in the early stages of Parkinson's disease. *J Neurol.* 2012;259(1):124-131. doi: 10.1007/s00415-011-6140-8.

9. Abraham DS, Gruber-Baldini AL, Harrington D, Shulman LM. The factor structure of the Brief Symptom Inventory-18 (BSI-18) in Parkinson disease patients. *J Psychosom Res.* 2017;96:21-26. doi: 10.1016/j.jpsychores.2017.03.002.

10. Carrozzino D, Morberg BM, Siri C, Pezzoli G, Bech P. Evaluating psychiatric symptoms in Parkinson's Disease by a clinimetric analysis of the Hopkins Symptom Checklist (SCL-90-R). *Prog Neuropsychopharmacol Biol Psychiatry.* 2018;81:131-137. doi: 10.1016/j.pnpbp.2017.10.024.

11. Carrozzino D, Siri C, Bech P. The prevalence of psychological distress in Parkinson's disease patients: The brief symptom inventory (BSI-18) versus the Hopkins symptom checklist (SCL-90-R). *Prog Neuropsychopharmacol Biol Psychiatry.* 2019;88:96-101. doi: 10.1016/j.pnpbp.2018.07.012.

12. Kwok JYY, Choi EPH, Chau PH, Wong JYH, Fong DYT, Auyeung M. Effects of spiritual resilience on psychological distress and health-related quality of life in Chinese people with Parkinson's disease. *Qual Life Res.* 2020;29(11):3065-3073. doi: 10.1007/s11136-020-02562-x.

13. van der Heide A, Meinders MJ, Bloem BR, Helmich RC. The impact of the COVID-19 pandemic on psychological distress, physical activity, and symptom severity in Parkinson's disease. *J Parkinsons Dis.* 2020;10(4):1355-1364. doi: 10.3233/JPD-202251.

14. Polo-Morales A, Alcocer-Salas Á, Rodríguez-Violante M, Pinto-Solís D, Solís-Vivanco R, Cervantes-Arriaga A. Association Between Somatization and Nonmotor Symptoms Severity in People with Parkinson Disease. *J Geriatr Psychiatry Neurol.* 2021;34(1):60-65. doi: 10.1177/0891988720901787.

15. Marinus J, Leentjens AF, Visser M, Stiggelbout AM, van Hilten JJ. Evaluation of the hospital anxiety and depression scale in patients with Parkinson's disease. *Clin Neuropharmacol.* 2002;25(6):318-324. doi: 10.1097/00002826-200211000-00008.

16. Martinez-Martin P, Rodriguez-Blazquez C, Frades-Payo B. Specific patient-reported outcome measures for Parkinson's disease: analysis and applications. *Expert Rev Pharmacoecon Outcomes Res.* 2008;8(4):401-418. doi: 10.1586/14737167.8.4.401.

17. Leentjens AF, Dujardin K, Marsh L, Richard IH, Starkstein SE, Martinez-Martin P. Anxiety rating scales in Parkinson's disease: a validation study of the Hamilton anxiety rating scale, the Beck anxiety inventory, and the hospital anxiety and depression scale. *Mov Disord.* 2011;26(3):407-415. doi: 10.1002/mds.23184.

18. De Roos P, Bloem BR, Kelley TA, et al. A consensus set of outcomes for Parkinson's disease from the International Consortium for Health Outcomes Measurement. *J Parkinsons Dis.* 2017;7(3):533-543. doi: 10.3233/JPD-161055.

19. Feinstein AR. The Jones criteria and the challenge of clinimetrics. *Circulation.* 1982;66(1):1-5. doi: 10.1161/01.cir.66.1.1.

20. Feinstein AR. An additional science for clinical medicine: IV. The development of clinimetrics. *Ann Intern Med.* 1983;99(6):843-848. doi: 10.7326/0003-4819-99-6-843.

21. Feinstein AR. *Clinimetrics.* New Haven: Yale University Press; 1987.

22. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *Int J Clin Pract.* 2012;66(1):11-15. doi: 10.1111/j.1742-1241.2011.02825.x.

23. Carrozzino D, Patierno C, Guidi J, et al. Clinimetric Criteria for Patient-Reported Outcome Measures. *Psychother Psychosom.* 2021;90(4):222-232. doi: 10.1159/000516599.

24. Bech P. *Clinical psychometrics*. Oxford: Wiley-Blackwell; 2012. doi: 10.1002/9781118511800.

25. Nielsen MG, Ørnbøl E, Vestergaard M, Bech P, Christensen KS. The construct validity of the Major Depression Inventory: A Rasch analysis of a self-rating scale in primary care. *J Psychosom Res.* 2017;97:70-81. doi: 10.1016/j.jpsychores.2017.04.001.

26. Fava GA, Carrozzino D, Lindberg L, Tomba E. The clinimetric approach to psychological assessment: a tribute to Per Bech, MD (1942–2018). *Psychother Psychosom.* 2018;87(6):321-326. doi: 10.1159/00049374.

27. Christensen KS, Oernboel E, Nielsen MG, Bech P. Diagnosing depression in primary care: a Rasch analysis of the major depression inventory. *Scand J Prim Health Care.* 2019;37(1):105-112. doi: 10.1080/02813432.2019.1568703.

28. Carrozzino D, Patierno C, Fava GA, Guidi J. The Hamilton Rating Scales for Depression: A critical review of clinimetric properties of different versions. *Psychother Psychosom.* 2020;89(3):133-150. doi: 10.1159/000506879.

29. Carrozzino D, Christensen KS, Cosci F. Construct and criterion validity of Patient-Reported Outcomes (PROs) for depression: A clinimetric comparison. *J Affect Disord.* 2021;283:30-35. doi: 10.1016/j.jad.2021.01.043.

30. Kellner R. Part 2. Improvement criteria in drug trials with neurotic patients. *Psychol Med.* 1972;2(1):73-80. doi: 10.1017/s003329170004564.

31. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry.* 1992;55(3):181-184. doi: 10.1136/jnnp.55.3.181.

32. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes.* 2014;12:2. doi: 10.1186/s12955-014-0176-2.

33. Hagell P, Westergren A. Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (RUMM) Program in Health Outcome Measurement. *J Appl Meas.* 2016;17(4):416-431.

34. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194. doi: 10.1001/jama.2013.281053.

35. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behav Sci.* 1974;19(1):1-15. doi: 10.1002/bs.3830190102.

36. Parloff MB, Kelman HC, Frank JD. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *Am J Psychiatry.* 1954;111(5):343-352. doi: 10.1176/ajp.111.5.343.

37. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nord J Psychiatry.* 2003;57(2):113-118. doi: 10.1080/08039480310000932.

38. Olsen LR, Mortensen EL, Bech P. The SCL-90 and SCL-90R versions validated by item response models in a Danish community sample. *Acta Psychiatr Scand.* 2004;110(3):225-229. doi: 10.1111/j.1600-0447.2004.00399.x.

39. Müller JM, Postert C, Beyer T, Furniss T, Achtergarde S. Comparison of eleven short versions of the Symptom Checklist 90-Revised (SCL-90-R) for use in the assessment of general psychopathology. *J Psychopathol Behav Assess.* 2010;32(2):246-254. doi: 10.1007/s10862-009-9141-5.

40. Prinz U, Nutzinger DO, Schulz H, Petermann F, Braukhaus C, Andreas S. Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders. *BMC Psychiatry.* 2013;13:104. doi: 10.1186/1471-244X-13-104.

41. Bech P, Bille J, Møller SB, Hellström LC, Østergaard SD. Psychometric validation of the Hopkins Symptom Checklist (SCL-90) subscales for depression, anxiety, and interpersonal sensitivity. *J Affect Disord.* 2014;160:98-103. doi: 10.1016/j.jad.2013.12.005.

42. Derogatis LR. Symptom checklist-90-revised (SCL-90-R). In: Rush AJ, First MB, Blacker D, eds. *Handbook of Psychiatric Measures*. Washington: American Psychiatric Publishing; 2008:73-76.

43. Prunas A, Sarno I, Preti E, Madeddu F, Perugini M. Psychometric properties of the Italian version of the SCL-90-R: a study on a large community sample. *Eur Psychiatry.* 2012;27(8):591-597. doi: 10.1016/j.eurpsy.2010.12.006.

44. Carrozzino D, Vassend O, Bjørndal F, Pignolo C, Olsen LR, Bech P. A clinimetric analysis of the Hopkins Symptom Checklist (SCL-90-R) in general population studies (Denmark, Norway, and Italy). *Nord J Psychiatry.* 2016;70(5):374-379. doi: 10.3109/08039488.2016.1155235.

45. Andrich D, Lyne A, Sheridan B, Luo G. *RUMM 2030*. Perth: RUMM Laboratory; 2010.

46. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol.* 2007;46(1):1-18. doi: 10.1348/014466506x96931.

47. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007;57(8):1358-1362. doi: 10.1002/art.23108.

48. Christensen KS, Oernboel E, Zatzick D, Russo J. Screening for depression: Rasch analysis of the structural validity of the PHQ-9 in acutely injured trauma survivors. *J Psychosom Res.* 2017;97:18-22. doi: 10.1016/j.jpsychores.2017.03.117.

49. Marais I, Andrich D. Effects of Varying Magnitude and Patterns of Response Dependence. *J Appl Meas.* 2008;9(2):105-124.

50. Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL): factors derived from the HSCL-90. *J Affect Disord.* 1979;1(1):9-24. doi: 10.1016/0165-0327(79)90021-1.

51. Bech P. Modern psychometrics in clinimetrics. *Psychother Psychosom.* 2004;73(3):134-138. doi: 10.1159/000076448.

52. Fleck MP, Carrozzino D, Fava GA. The challenge of measurement in psychiatry: the lifetime accomplishments of Per Bech (1942-2018). *Braz J Psychiatry.* 2019;41(5):369-372. doi: 10.1590/1516-4446-2019-0509.

53. Bech P. Clinimetric dilemmas in outcome scales for mental disorders. *Psychother Psychosom.* 2016;85(6):323-326. doi: 10.1159/000448810.

54. Carrozzino D, Bech P, Patierno C, et al. Somatization in Parkinson's disease: a systematic review. *Prog Neuropsychopharmacol Biol Psychiatry.* 2017;78:18-26. doi: 10.1016/j.pnpbp.2017.05.011.

55. Fava GA, Cosci F, Sonino N. Current psychosomatic practice. *Psychother Psychosom.* 2017;86(1):13-30. doi: 10.1159/000448856.

56. Koo BB, Chow CA, Shah DR, et al. Demoralization in Parkinson disease. *Neurology.* 2018;90(18):e1613-e1617. doi: 10.1212/WNL.0000000000005425.

57. Elfil M, Ahmed N, Alapati A, et al. Suicidal risk and demoralization in Parkinson disease. *J Neurol.* 2020;267(4):966-974. doi: 10.1007/s00415-019-09632-2.

58. Zhu B, Kohn R, Patel A, Koo BB, Louis ED, De Figueiredo JM. Demoralization and Quality of Life of Patients with Parkinson Disease. *Psychother Psychosom.* 2021;1-7. doi: 10.1159/000514270.

59. Chaudhuri KR, Healy DG, Schapira AHV. Non-motor symptoms of Parkinson's disease: diagnosis and management. *Lancet Neurol.* 2006;5(3):235-245. doi: 10.1016/S1474-4422(06)70373-8.

60. Fernandes M, Pierantozzi M, Stefani A, et al. Frequency of non-motor symptoms in Parkinson's patients with motor fluctuations. *Front Neurol.* 2021;12:678373. doi: 10.3389/fneur.2021.678373.

61. Jeong W, Kim H, Jo JH, Jang S-I, Park E-C. Association between depression and risk of Parkinson's disease in South Korean adults. *J Affect Disord.* 2021;292:75-80. doi: 10.1016/j.jad.2021.05.038.

62. Pontone GM, Mills KA. Optimal treatment of depression and anxiety in Parkinson's disease. *Am J Geriatr Psychiatry.* 2021;29(6):530-540. doi: 10.1016/j.jagp.2021.02.037.

63. Schönenberg A, Zipprich HM, Teschner U, Grosskreutz J, Witte OW, Prell T. Impact of subthreshold depression on health-related quality of life in patients with Parkinson's disease based on cognitive status. *Health Qual Life Outcomes.* 2021;19(1):107. doi: 10.1186/s12955-021-01753-5.

64. Chuquilín-Arista F, Álvarez-Avellón T, Mendéndez-González M. Impact of depression and anxiety on dimensions of health-related quality of life in subjects with Parkinson's disease enrolled in an association of patients. *Brain Sci.* 2021;11(6):771. doi: 10.3390/brainsci11060771.

Table 1. Sample characteristics (n=488)

	Mean	SD
Age	65.86	10.26
Age at onset	53.45	10.88
Disease duration	12.09	5.98
H&Y	2.33	0.72
UPDRS III	20.63	11.35
L-Dopa (mg/day)	521.06	278.51

H&Y: Hoehn and Yahr staging scale; L-Dopa: levodopa; SD: standard deviation;

UPDRS: Unified Parkinson's Disease Rating Scale motor score.

Table 2. Model fit statistics for SCL-90-R items (n=488)

Action	Analysis	Model fit (overall)	Item fit residual, mean (SD)	Person fit residual, mean (SD)	PSI	Dimensionality, significant t-tests (%)
Original sample	1	$\chi 2(630)=1434.49$, $p<0.001$	-0.15 (2.27)	-1.20 (0.74)	0.96	28.69
Rescoring all items	2	$\chi 2(630)=1056.88$, $p<0.001$	-0.38 (1.72)	-2.04 (1.11)	0.95	26.02
Delete item 19	3	$\chi 2(623)=1024.11$, $p<0.001$	-0.39 (1.69)	-2.03 (1.12)	0.95	27.46
Delete item 27	4	$\chi 2(616)=1009.19$, $p<0.001$	-0.37 (1.69)	-2.06 (1.13)	0.95	25.61
Delete item 60	5	$\chi 2(609)=971.82$, $p<0.001$	-0.37 (1.67)	-2.07 (1.14)	0.95	25.61
Delete item 17	6	$\chi 2(602)=936.44$, $p<0.001$	-0.35 (1.67)	-2.09 (1.16)	0.95	25.61
Delete item 5	7	$\chi 2(595)=889,88$, $p<0.001$	-0.36 (1.63)	-2.10 (1.16)	0.95	25.41
Delete item 84	8	$\chi 2(588)=850.21$, $p<0.001$	-0.37 (1.58)	-2.11 (1.17)	0.95	25.00
Delete item 9	9	$\chi 2(581)=801.31$, $p<0.001$	-0.36 (1.56)	-2.13 (1.18)	0.95	25.20

Delete item 40	10	χ2(574)=800.65, p<0.001	-0.38 (1.53)	-2.12 (1.19)	0.95	25.20
Delete item 44	11	χ2(567)=775.64, p<0.001	-0.38 (1.48)	-2.14 (1.20)	0.95	23.98
Delete item 8	12	χ2(560)=753.05, p<0.001	-0.40 (1.46)	-2.13 (1.21)	0.95	25.61
Delete item 12	13	χ2(553)=725.89, p<0.001	-0.42 (1.41)	-2.12 (1.21)	0.95	23.77
Delete item 1	14	χ2(546)=709.77, p<0.001	-0.43 (1.38)	-2.12 (1.22)	0.95	23.36

χ2: chi-square; p: probability; SD: standard deviation; PSI: person separation index (with extremes)

Table 3. Model fit statistics for the SCL-90-R subscales (n=488)

SCL-90-R subscales	K	Model fit (overall)	Item fit residual, mean (SD)	Person fit residual, mean (SD)	PSI	Dimensionality, significant t-tests (%)
Phobia	7	χ2(42)=67.22 p=0.008	-0.24 (0.99)	-2.35 (1.41)	0.54	2.26
Anxiety	10	χ2(70)=174.42, p<0.001	-0.84 (1.93)	-1.91 (1.55)	0.76	6.17
- Item 17	9	χ2(63)=114.23, p<0.001	-0.69 (1.56)	-2.14 (1.72)	0.76	5.97
- Item 33	8	χ2(56)=63.18, p=0.24	-0.53 (1.26)	-2.06 (1.70)	0.73	4.32
Depression	16	χ2(105)=166.14 p<0.001	-0.54 (2.06)	-1.84 (1.41)	0.81	6.37
- Item 19	*15*	χ2(98)=126.65, p=0.027	-0.62 (1.92)	-1.79 (1.47)	0.81	7.19
- Item 5	14	χ2(91)=100.33, p=0.236	-0.75 (1.47)	-1.86 (1.56)	0.81	4.11
Somatization	12	χ2(84)=98.47, p=0.134	-0.23 (1.22)	-1.69 (1.44)	0.78	6.97
Cognitive performance deficits	9	χ2(63)=105.40, p<0.001	-1.37 (1.71)	-1.56 (1.54)	0.74	3.90
- Item 65	8	χ2(56)=65.35, p=0.184	-1.57 (1.53)	-1.39 (1.60)	0.72	2.67

Interpersonal Sensitivity	18	χ2(126)=209.08, p<0.001	-0.45 (1.77)	-2.83 (1.49)	0.82	6.37
- Item 8	17	χ2(119)=161.09, p=0.006	-0.45 (1.59)	-2.82 (1.53)	0.82	5.54
Hostility	6	χ2(30)=78.34, p<0.001	-1.27 (1.41)	-3.76 (1.64)	0.45	1.64
Sleep	3	χ2(14)=29.88, p= 0.007	-1.16 (0.52)	-0.71 (1.64)	0.41	NA
Paranoid ideation	6	χ2(36)=46.52, p=0.113	-0.09 (1.17)	-2.48 (1.45)	0.48	2.91
Psychoticism	10	χ2(70)=173.02, p<0.001	-0.92 (1.83)	-3.07 (1.32)	0.61	2.87
- Item 90	9	χ2(54)=160.28, p<0.001	-1.01 (1.66)	-3.14 (1.26)	0.53	2.47
- Item 87	8	χ2(48)=110.47, p<0.001	-0.31 (1.81)	-3.43 (1.24)	0.41	1.24
- Item 84	7	χ2(42)=89.79, p<0.001	-0.22 (1.25)	-3.68 (1.29)	0.35	1.03
- Item 7	6	χ2(30)=57.53, p=0.002	0.09 (0.79)	-3.16 (1.25)	0.24	0.41

K: number of items; χ2: chi-square; p: probability; SD: standard deviation; PSI: person separation index (with extremes)