

RESEARCH ARTICLE

Test-retest reliability and minimal detectable change of the Beck Depression Inventory and the Taiwan Geriatric Depression Scale in patients with Parkinson's disease

Sheau-Ling Huang^{1,2}, Ching-Lin Hsieh^{1,2}, Ruey-Meei Wu³, Wen-Shian Lu^{4,5*}

1 School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan, **2** Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan, **3** Department of Neurology, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan, **4** School of Occupational Therapy, Chung Shan Medical University, Taichung, Taiwan, **5** Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung, Taiwan

* wenshian.lu@gmail.com



Abstract

Background

The Beck Depression Inventory II (BDI-II) and the Taiwan Geriatric Depression Scale (TGDS) are self-report scales used for assessing depression in patients with Parkinson's disease (PD) and geriatric people. The minimal detectable change (MDC) represents the least amount of change that indicates real difference (i.e., beyond random measurement error) for a single subject. Our aim was to investigate the test-retest reliability and MDC of the BDI-II and the TGDS in people with PD.

Methods

Seventy patients were recruited from special clinics for movement disorders at a medical center. The patients' mean age was 67.7 years, and 63.0% of the patients were male. All patients were assessed with the BDI-II and the TGDS twice, 2 weeks apart. We used the intraclass correlation coefficient (ICC) to determine the reliability between test and retest. We calculated the MDC based on standard error of measurement. The MDC% was calculated (i.e., by dividing the MDC by the possible maximal score of the measure).

Results

The test-retest reliabilities of the BDI-II/TGDS were high (ICC = 0.86/0.89). The MDCs (MDC%s) of the BDI-II and TGDS were 8.7 (13.8%) and 5.4 points (18.0%), respectively. Both measures had acceptable to nearly excellent random measurement errors.

Conclusions

The test-retest reliabilities of the BDI-II and the TGDS are high. The MDCs of both measures are acceptable to nearly excellent in people with PD. These findings imply that the BDI-II

OPEN ACCESS

Citation: Huang S-L, Hsieh C-L, Wu R-M, Lu W-S (2017) Test-retest reliability and minimal detectable change of the Beck Depression Inventory and the Taiwan Geriatric Depression Scale in patients with Parkinson's disease. PLoS ONE 12(9): e0184823. <https://doi.org/10.1371/journal.pone.0184823>

Editor: Yi Hu, Chinese Academy of Sciences, CHINA

Received: April 22, 2017

Accepted: August 31, 2017

Published: September 25, 2017

Copyright: © 2017 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

and the TGDS are suitable for use in a research context and in clinical settings to detect real change in a single subject.

Introduction

People with Parkinson's disease (PD) are clinically associated with movement-related symptoms (e.g., tremor, bradykinesia, rigidity, and postural instability). The recent literature shows that the impact of non-motor symptoms on daily function in the patients with PD is notable [1–3]. Depression is one of the most prevalent neuropsychiatric symptoms in patients with PD [4, 5]. Among non-motor symptoms, depression is one that has a great impact on the quality of life of patients with PD and their caregivers [3, 6, 7]. The assessment of depression can help clinicians detect early stages of depression in patients with PD and plan appropriate intervention. To screen and monitor the state of depression, clinicians need to routinely assess the depression in patients with PD. Moreover, to interpret the results of the assessments, clinicians have to decide whether the change scores in the depression scales are true, or beyond measurement error.

Test-retest reliability refers to the extent of agreement between measurements of a measure in successive sessions [8]. In addition, random measurement error exists in all assessments. If the measurement error of a measure is not determined, the interpretability of the test scores is limited [9, 10]. To calculate the random measurement error, the minimal detectable change (MDC) is used. The MDC, also called the smallest real difference (SRD) [11], is the smallest threshold of a change score that is beyond random error at a certain confidence level (usually 95%) [12]. The MDC can be used as a threshold to confirm whether a difference between consecutive assessments on a measure of a single patient indicates a real change or is due to random error [13]. Thus, the value of MDC also indicates the amount of random measurement error. Accordingly, the MDC of a measure is vital in the interpretation of data for both clinicians and researchers [9, 10].

The second edition of the Beck Depression Inventory (BDI-II) is widely used for screening, assessing, and monitoring the depression status and change trend in community-dwelling older adults [14] and in different populations [15–17]. The Taiwan Geriatric Depression Scale (TGDS) is frequently used to screen for depression in the elderly [18] and to assess depression in the elderly with depressive disorder [19]. To our knowledge, the test-retest reliability and the MDC of the TGDS have not been examined in people with PD. Although the test-retest reliability and the MDC of the BDI-II in people with PD have been investigated, the characteristics of the participants in the data analysis of test-retest reliability and MDC have not been clearly reported, limiting the generalization of these results [20]. A measure without a determined MDC does not allow clinicians and researchers to explain change scores in successive measurements reasonably. That is to say, lacking explicit MDCs, the BDI-II and the TGDS have limited applicability and interpretability in people with PD. Therefore, the aim of this study was to estimate and compare the extents of the test-retest reliability and the amounts of random measurement error of the two measures. Our results should help potential users to select appropriate measures for, and interpret the clinical data of, patients with PD.

Materials and methods

Participants

All of the participants in this study were recruited from special clinics for movement disorders in the Department of Neurology of a medical center located in northern Taiwan. The inclusion criteria were as follows: (1) Diagnosis of PD by a movement disorder specialist, based on the UK PDS Brain Bank Criteria for the diagnosis of PD; (2) Hoehn and Yahr stages between stages I and III; (3) ≥ 50 years; (4) Mini-mental state examination (MMSE) score > 20 points; and (5) agreement to participate in this study and signing of the consent forms, as approved by the medical ethics committee at the medical center. Participants were excluded from the study if they had any of the following conditions: (1) Atypical Parkinsonism, including diffuse Lewy body dementia, multiple system atrophy, progressive supranuclear palsy, and corticobasal degeneration; (2) Secondary Parkinsonism (e.g., vascular Parkinsonism); or (3) Dementia, based on the diagnostic criteria for research in the ICD-10 Classification of mental and behavioral disorders. We retrieved medical records of the patients to confirm whether they had been diagnosed with the above diagnoses by their physicians.

Procedures

The participants were screened and invited by the movement disorder specialists. All participants filled out two questionnaires during the “on” status at the same place in two sessions about 2 weeks apart. Half of the participants filled out the BDI-II before the TGDS; the order was reversed for the other half to control possible bias of the testing sequence. Prior to the second session assessment, the research assistant confirmed that each participant had no change of the medication for Parkinson’s disease and other neuropsychiatric disorders; no change of Parkinson features, such as tremor, rigidity, bradykinesia, and gait disturbance; and no experience of brain injury during the follow-up period of this study.

Measures

Beck Depression Inventory (BDI-II). The BDI-II has 21 items scored on a 4-point scale (from 0 to 3), and a total score ranging from 0 to 63. In accordance with the BDI-II manual, the severity of depression was distinguished by the total score of the BDI-II as minimal (0–13), mild (14–19), moderate (20–28), or severe (29–63) [21].

Taiwan Geriatric Depression Scale (TGDS). The Taiwan Geriatric Depression Scale (TGDS), based on the Geriatric Depression Scale, is a sensitive, valid, and reliable measure for use in elderly outpatients in Taiwan [22]. It consists of 30 items of 1 point each, with a higher score indicating more severe depression. It has adequate test-retest reliability, and the best cut-off value for depressive status for geriatric outpatients is 15 [22].

Data analysis

SPSS 16.0 (SPSS Inc., USA) was employed to analyze our data. We first investigated the test-retest reliability of the measures by calculating the intraclass correlation coefficient (ICC). An ICC value of more than 0.80 is considered to indicate high test-retest reliability of a measure [23].

The value of ICC was then employed to estimate the standard error of measurement (SEM), and thereafter, to estimate the MDC (95% level of confidence) [9, 12].

The MDC% was then calculated by dividing the MDC by the maximal score of the measure [24]. An MDC% $< 30\%$ is considered acceptable, and $< 10\%$ is considered excellent [24].

In addition, we made Bland-Altman plots with 95% limits of agreement (LOA) [25] to observe the variance between repeated assessments. Assuming the differences follow normal distribution, 95% of the differences would distribute between mean difference ± 1.96 $SD_{\text{difference}}$ (i.e., LOA). The plots are useful to present heteroscedasticity. Heteroscedasticity means that a trend of increasing or decreasing exists in the amounts of difference within two adjacent assessments as the mean scores of the two assessments increase [26]. We used Pearson's r to examine whether heteroscedasticity existed by calculating the level of association between the average scores and the absolute values of difference of adjacent assessments. If Pearson's $r > 0.3$, the data were considered heteroscedastic [27].

We used paired t tests to examine whether systematic bias existed between test-retest assessments.

Last, we investigated the correlation between the scores of the BDI-II and the TGDS using the Pearson correlation coefficient (r) at first and second section assessments, respectively. The Pearson $r \geq 0.8$ was considered to show very strong correlation [28].

Ethics

Ethics approval was obtained from Research Ethics Committee, National Taiwan University Hospital—Project Number: 200705055R.

Results

Seventy-five participants with PD agreed to participate in the study. Five participants were lost to follow-up due to loss of contact or refusal to re-test. The mean baseline scores of the TGDS between the remaining 70 participants and the 5 participants lost to follow-up were not statistically different ($p = 0.116$), but the difference of the BDI-II between the 70 participants and the 5 patients lost to follow-up was statistically significant ($p = 0.037$). Demographic and clinical characteristics of the 70 participants are shown in Table 1. Their mean age was 67.7 years, and 63.0% of the participants were male. The mean baseline scores of the BDI-II and the TGDS were 8.5 points and 8.2 points, respectively.

The test-retest reliability indices of the BDI-II and the TGDS are shown in Table 2. The ICCs for the BDI-II and the TGDS were 0.86 and 0.89, respectively. The mean differences for the BDI-II and the TGDS were 0.5 and -0.4, respectively. The MDCs (MDC%) of the BDI-II and the TGDS were 8.7 points (13.8%) and 5.4 points (18.0%), respectively.

Table 1. The demographic and clinical characteristics of the participants.

Characteristic	Participants completing two assessments (n = 70)	Participants lost to follow-up (n = 5)
Age, mean (SD)	67.7 (8.8)	66.4 (3.8)
Sex, male/female (n)	44/26	4/1
H&Y stage		
I	17	3
II	29	1
III (n)	24	1
PD evolution time (range)	6 months to 15 years	6 months to 10 years
BDI-II baseline scores (point), mean (SD)	8.5 (8.4)	2.6 (0.9)
TGDS baseline scores (point), mean (SD)	8.2 (6.0)	6.0 (9.0)
Interval between two assessments (day), mean (SD)	13.3 (2.7)	-

BDI-II: the Beck Depression Inventory (II)

TGDS: the Taiwan Geriatric Depression Scale

<https://doi.org/10.1371/journal.pone.0184823.t001>

Table 2. The test-retest reliability indices of the BDI-II and the TGDS (n = 70).

Test	First session Mean (SD)	Second session Mean (SD)	Difference Mean (SD)	ICC (95% CI)	MDC (MDC %)
BDI-II	8.5 (8.4)	9 (8.5)	0.5 (4.5)	0.86 (0.78–0.91)	8.7 (13.8%)
TGDS	8.2 (6.0)	7.8 (6.5)	-0.4 (2.9)	0.89 (0.83–0.93)	5.4 (18.0%)

BDI-II: the Beck Depression Inventory (II)

TGDS: the Taiwan Geriatric Depression Scale

ICC: intraclass correlation coefficient

$ICC_{(2,1)} = (BMS - EMS) / (BMS + EMS + 2(JMS - EMS) / n)$, BMS = between-participants mean square, EMS = the error mean square, JMS = observations mean square, and n = the number of participants

MDC: minimal detectable change

$MDC = 1.96 \times SEM \times \sqrt{2}$; $MDC\% = (MDC / \text{maximal score of measurements}) \times 100\%$

<https://doi.org/10.1371/journal.pone.0184823.t002>

In Fig 1, the differences of scores are plotted against the mean scores of the 2 measurements of both measures [25]. The LOAs (limits of agreements) ranged from 9.3 to -8.3 points for the BDI-II, and from 5.2 to -6.1 points for the TGDS. In addition, the Pearson's correlation coefficients (r) between the mean and the difference for the BDI-II and the TGDS were 0.4 and 0.3, respectively.

The results of paired t tests showed that the differences between the two sessions of the BDI-II and the TGDS were not significant ($p > 0.05$).

The Pearson's r values were 0.80 and 0.85 at the first and second sections, between the scores of the BDI-II and the TGDS, respectively. The results indicated that the correlation between the BDI-II and the TGDS was very strong.

Discussion

We compared the extent of the test-retest reliability of the two measures. The results showed that the ICC values of the BDI-II and the TGDS were high (ICC = 0.86/0.89). Our findings indicated that both the BDI-II and the TGDS are equally reliable when administered to patients with PD. Our finding of sound test-retest reliabilities of the BDI-II and the TGDS are consistent with those of previous studies when these two measures are applied to other populations [20, 21, 29].

Our results showed that the MDCs of the BDI-II and the TGDS were 8.7 points and 5.4 points, respectively. The results suggest that a score of ≥ 9 for the BDI-II and one of ≥ 6 points for the TGDS are needed to announce a real change (beyond random measurement error) between repeated measurements. Our results can be viewed as a threshold reference for the BDI-II and the TGDS to help clinicians and researchers reasonably and confidently determine real change between repeated measurements for patients with PD.

We further compared the proportion of random measurement error relative to possible maximal score (i.e., MDC%) of these two measures. The MDC% of the BDI-II (13.8%) and that of the TGDS (18.0%) were acceptable to nearly excellent [24], suggesting that both measures are suitable for detecting a true change between two successive tests for a single patient with PD in a clinical setting. Based on the number of items and the MDC%, the BDI-II seems to be more appropriate for use in detecting the status of depression of patients with PD in clinical settings.

In comparing our results to those of a previous study [20], we found that both studies showed that the test-retest reliability of the BDI-II for patients with PD was high, with ICCs above 0.86, further supporting the reliability of the BDI-II in patients with PD. On the other

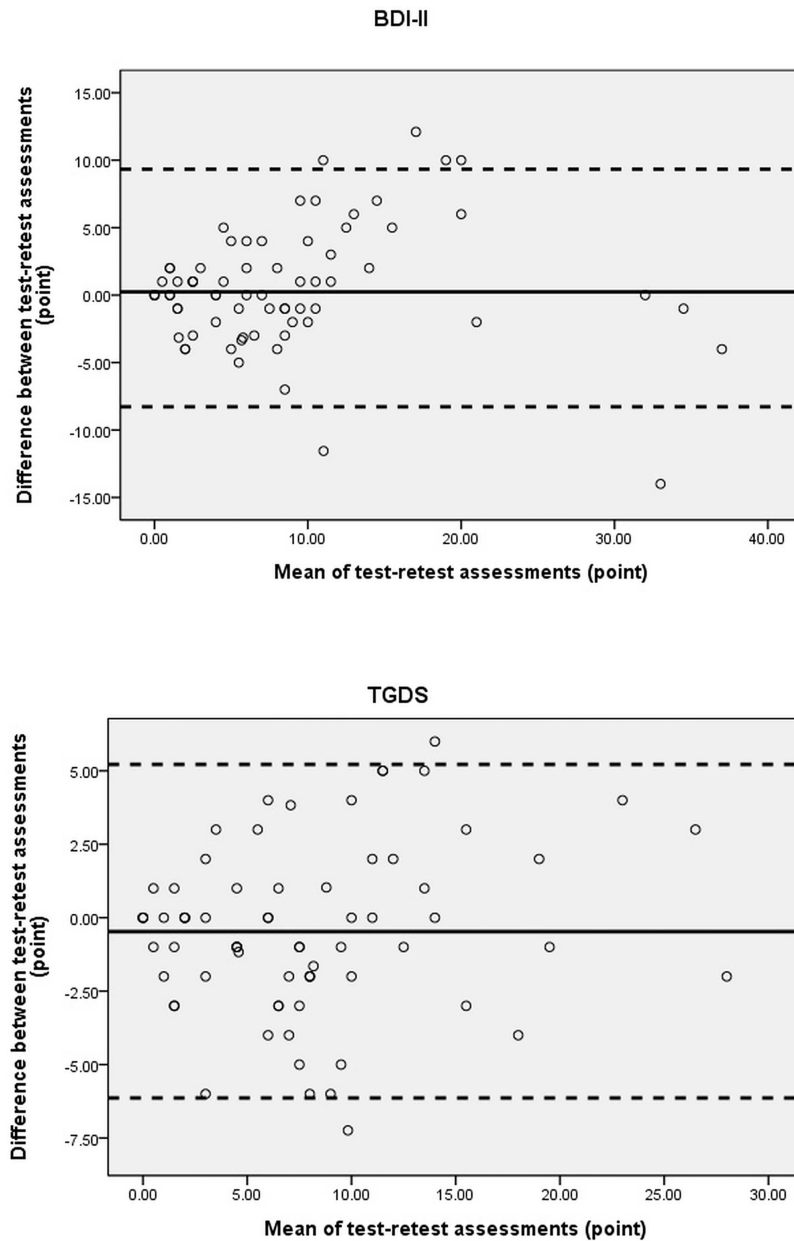


Fig 1. Bland & Altman plots of the difference scores against the mean scores of each pair of the BDI-II and the TGDS. The solid line represents the mean of the differences. The 2 dashed lines define limits of agreement (mean of the difference \pm 1.96 SD).

<https://doi.org/10.1371/journal.pone.0184823.g001>

hand, the MDC in our research (8.7 points) was somewhat greater than that of a previous study [20] (3.3 points). However, the characteristics of the sample (e.g., the number and severity of disease of participants) and detailed data (e.g., the mean scores and standard deviation of the test and the retest) included in the data analysis of test-retest reliability and MDC were not clearly reported in that study [20]. These shortcomings limit the generalization of those results, hindering any further comparison between our results and those of the previous study [20]. However, our finding of an MDC of 8.7 points could be viewed as a high standard to detect a real change between two sessions in patients with PD.

In terms of heteroscedasticity, the Bland-Altman plot (Fig 1) shows an obvious systematic trend for the BDI-II, and the Pearson's r values between the absolute value of difference and the mean of the BDI-II were > 0.3 , indicating the data were heteroscedastic (i.e., the mean and the difference of each pair of repeated measurements increased or decreased simultaneously). According to Flansbjer et al., if the data are heteroscedastic, a fixed value of MDC is not a proper threshold for determining a real change between repeated measurements because the amount of random measurement error depends on the initial test score of the measure [13]. For example, if a patient has an initial BDI-II score of 30 points, a change of more than 4.1 points (30×0.138) is needed to indicate a true change. The MDC of 8.7 points could be considered as the highest criterion for detecting a real change between successive tests of the BDI-II. As for TGDS, Pearson's r was 0.3, just close to the criterion of the presence of the heteroscedasticity. If prospective users concern such an issue, the users can use the same rules applied to the BDI-II mentioned above to estimate the true change between successive tests of the TGDS. These results should be helpful for clinicians to interpret the changes between repeated measurements for single patients more accurately.

According to the results of paired t tests between the test and retest sessions of the BDI-II and the TGDS ($p > 0.05$), no systematic bias was found in this research.

Regarding the issue of the patients lost to follow-up in this study, the number of patients lost to follow-up was five. Their mean score (SD) of the BDI-II was 2.6 (0.9), which was significantly different from that of the other 70 participants. We could not accurately estimate the degree to which the loss of the 5 patients to follow-up influenced our results. However, the proportion of patients lost to follow-up (5) to all participants (75) was about 6.7%, and the variation of these patients seemed small (SD = 0.9), implying a limited effect of the loss of the 5 patients to follow-up on the robustness of our results.

In the current study, we focused on determining the MDC to depict the extent of random error and a threshold of statistical significance for the BDI-II and the TGDS. However, in a clinical context, the minimal important difference (MID), which is the degree of a change meaningful to a patient, is equally crucial in decision making in treatment planning [30]. To improve the utility and applicability of the BDI-II and the TGDS for patients with PD, future research to investigate the MID of the two measures is warranted.

This study has one limitation. The characteristics of our sample (a convenience sample and participants with mild to moderate severity) might threaten the generalizability of our findings. Future research could enroll more patients with a more even distribution (e.g., from mild to severe disability) to cross-validate our findings.

Conclusion

Both the BDI-II and the TGDS are reliable measures of the severity of depression in people with PD. In addition, both measures have acceptable to nearly excellent random measurement error. These results imply that the BDI-II and the TGDS are suitable for use in research and clinical settings to detect real change in a single patient with PD.

Supporting information

S1 Dataset. Dataset of the study.
(XLSX)

Acknowledgments

We would like to thank the participants and raters for their work during data collection.

Author Contributions

Conceptualization: Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Wen-Shian Lu.

Data curation: Sheau-Ling Huang, Ching-Lin Hsieh.

Formal analysis: Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Wen-Shian Lu.

Funding acquisition: Sheau-Ling Huang, Ruey-Meei Wu.

Investigation: Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Wen-Shian Lu.

Methodology: Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Wen-Shian Lu.

Project administration: Sheau-Ling Huang, Ching-Lin Hsieh.

Resources: Sheau-Ling Huang, Ruey-Meei Wu.

Software: Ching-Lin Hsieh.

Supervision: Sheau-Ling Huang, Ching-Lin Hsieh.

Validation: Ching-Lin Hsieh.

Visualization: Sheau-Ling Huang, Ching-Lin Hsieh, Wen-Shian Lu.

Writing – original draft: Sheau-Ling Huang, Wen-Shian Lu.

Writing – review & editing: Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Wen-Shian Lu.

References

1. Nocera JR, Amano S, Vallabhajosula S, Hass CJ. Tai Chi Exercise to Improve Non-Motor Symptoms of Parkinson's Disease. *J Yoga Phys Ther.* 2013; 3:1–11. <https://doi.org/10.4172/2157-7595.1000137> PMID: [24278789](https://pubmed.ncbi.nlm.nih.gov/24278789/)
2. Duncan GW, Khoo TK, Yarnall AJ, O'Brien JT, Coleman SY, Brooks DJ, et al. Health-related quality of life in early Parkinson's disease: the impact of nonmotor symptoms. *Mov Disord.* 2014; 29(2):195–202. <https://doi.org/10.1002/mds.25664> PMID: [24123307](https://pubmed.ncbi.nlm.nih.gov/24123307/)
3. Liu WM, Lin RJ, Yu RL, Tai CH, Lin CH, Wu RM. The impact of nonmotor symptoms on quality of life in patients with Parkinson's disease in Taiwan. *Neuropsychiatr Dis Treat.* 2015; 11:2865–73. <https://doi.org/10.2147/NDT.S88968> PMID: [26635475](https://pubmed.ncbi.nlm.nih.gov/26635475/)
4. Leentjens AF. Depression in Parkinson's disease: conceptual issues and clinical challenges. *J Geriatr Psychiatry Neurol.* 2004; 17(3):120–126. <https://doi.org/10.1177/0891988704267456> PMID: [15312275](https://pubmed.ncbi.nlm.nih.gov/15312275/)
5. Aarsland D, Cummings JL. Depression in Parkinson's disease. *Acta Psychiatr Scand.* 2002; 106(3):161–162. <https://doi.org/10.1034/j.1600-0447.2002.2e009.x> PMID: [12197850](https://pubmed.ncbi.nlm.nih.gov/12197850/)
6. Karlsen KH, Tandberg E, Arslan D, Larsen JP. Health related quality of life in Parkinson's disease: a prospective longitudinal study. *J Neurol Neurosurg Psychiatry.* 2000; 69(5):584–589. <https://doi.org/10.1136/jnnp.69.5.584> PMID: [11032608](https://pubmed.ncbi.nlm.nih.gov/11032608/)
7. Aarsland D, Larsen JP, Karlsen K, Lim NG, Tandberg E. Mental symptoms in Parkinson's disease are important contributors to caregiver distress. *Int J Geriatr Psychiatry.* 1999; 14(10):866–874. [https://doi.org/10.1002/\(SICI\)1099-1166\(199910\)14:10<866::AID-GPS38>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1099-1166(199910)14:10<866::AID-GPS38>3.0.CO;2-Z) PMID: [10521886](https://pubmed.ncbi.nlm.nih.gov/10521886/)
8. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86(2):420–428. <https://doi.org/10.1037/0033-2909.86.2.420> PMID: [18839484](https://pubmed.ncbi.nlm.nih.gov/18839484/)
9. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006; 86(5):735–743. <https://doi.org/10.1093/ptj/86.5.735> PMID: [16649896](https://pubmed.ncbi.nlm.nih.gov/16649896/)
10. Jette AM, Tao W, Norweg A, Haley S. Interpreting rehabilitation outcome measurements. *J Rehabil Med.* 2007; 39(8):585–590. <https://doi.org/10.2340/16501977-0119> PMID: [17896048](https://pubmed.ncbi.nlm.nih.gov/17896048/)
11. Schuck P, Zwingmann C. The 'smallest real difference' as a measure of sensitivity to change: a critical analysis. *Int J Rehabil Res.* 2003; 26(2):85–91. PMID: [12799601](https://pubmed.ncbi.nlm.nih.gov/12799601/)
12. Schreuders TA, Roebroek ME, Goumans J, van Nieuwenhuijzen JF, Stijnen TH, Stam HJ. Measurement error in grip and pinch force measurements in patients with hand injuries. *Phys Ther.* 2003; 83(9):806–815. <https://doi.org/10.1093/ptj/83.9.806> PMID: [12940767](https://pubmed.ncbi.nlm.nih.gov/12940767/)

13. Flansbjerg UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med.* 2005; 37(2):75–82. <https://doi.org/10.1080/16501970410017215> PMID: 15788341
14. Segal DL, Coolidge FL, Cahill BS, O'Riley AA. Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. *Behav Modif.* 2008; 32(1):3–20. <https://doi.org/10.1177/0145445507303833> PMID: 18096969
15. Storch EA, Roberti JW, Roth DA. Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. *Depress Anxiety.* 2004; 19(3):187–189. <https://doi.org/10.1002/da.20002> PMID: 15129421
16. Arnau RC, Meagher MW, Norris MP, Bramson R. Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association.* 2001; 20(2):112–119. <http://dx.doi.org/10.1037/0278-6133.20.2.112>
17. Coles ME, Gibb BE, Heimberg RG. Psychometric evaluation of the Beck Depression Inventory in adults with social anxiety disorder. *Depression and anxiety.* 2001; 14(2):145–148. <https://doi.org/10.1002/da.1057> PMID: 11668668
18. Su TP, Huang SR, Chou P. Prevalence and risk factors of insomnia in community-dwelling Chinese elderly: a Taiwanese urban area survey. *The Australian and New Zealand journal of psychiatry.* 2004; 38(9):706–713. <https://doi.org/10.1080/j.1440-1614.2004.01444.x> PMID: 15324335
19. Chang YS, Liang SC, Chen MC, Lu MR. Quality of life in elderly with depressive disorder. *Taiwan Geriatr Gerontol.* 2006; 2(1):21–31.
20. Visser M, Leentjens AF, Marinus J, Stiggelbout AM, van Hilten JJ. Reliability and validity of the Beck depression inventory in patients with Parkinson's disease. *Mov Disord.* 2006; 21(5):668–672. <https://doi.org/10.1002/mds.20792> PMID: 16450355
21. Beck AT, Steer RA, Brown GK. BDI-II: Beck Depression Inventory Manual 2nd ed. San Antonio: Psychological Corporation; 1996.
22. Liao YC, Yeh TL, Yang YK, Lu FH, Chang CJ, Ko HC, et al. Reliability and Validity of the Taiwan Geriatric Depression Scale. *Taiwanese Journal of Psychiatry.* 2004; 18(1):30–41.
23. Prince B, Makrides L, Richman J. Research methodology and applied statistics. Part 2: the literature search. *Physiother Can.* 1980; 32(4):201–206. PMID: 10298052
24. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Deville WL, de Winter AF, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil.* 2002; 83(8):1145–1150. <https://doi.org/10.1053/apmr.2002.33728> PMID: 12161838
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1(8476):307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8) PMID: 2868172
26. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999; 8(2):135–160. <https://doi.org/10.1177/096228029900800204> PMID: 10501650
27. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998; 26(4):217–238. <https://doi.org/10.2165/00007256-199826040-00002> PMID: 9820922
28. Portney LG, Watkin MP. Foundations of clinical research: Applications to practice. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2009.
29. Levin BE, Llabre MM, Weiner WJ. Parkinson's disease and depression: psychometric properties of the Beck Depression Inventory. *J Neurol Neurosurg Psychiatry.* 1988; 51(11):1401–1404. PMID: 3236018
30. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Current opinion in rheumatology.* 2002; 14(2):109–114. PMID: 11845014