ORIGINAL ARTICLE

Beck Depression Inventory (BDI): A Reliability and Validity Test in the Malaysian Urological Population

K F Quek, M.Phil*, W Y Low, PhD*, A H Razack, FRCS, C S Loh, FRCS (Urol)***, *Health Research Development Unit, **Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, ***Sunway Medical Centre and Gleneagles Intan Medical Centre**

Summary

This study aimed to validate the Beck Depression Inventory (BDI) in the Malaysian urological population. Reliability and internal consistency were evaluated using the test-retest method and internal consistency was assessed using Cronbach's alpha. Responsiveness was expressed as the effect size. Internal consistency was high (Cronbach's alpha value = 0.56 to 0.87). Test-retest correlation coefficient and intraclass correlation coefficient were significant (ICC=0.56 to 0.87) and a high degree of sensitivity and specificity. The BDI is thus a reliable and a valid instrument to be used in Malaysia.

Key Words: Beck Depression Inventory, Cronbach's alpha, Sensitivity, Specificity, Test-retest reliability and Validity

Introduction

Benign prostatic hyperplasia (BPH) is a common disease, which is rarely life-threatening and impose a psychological impact on an individual such as depression[1]. It is no doubt that lower urinary tract symptoms (LUTS) can be a hindrance to patients and interfere with their daily activities. The physical symptoms such as the frequency, nocturia, urgency, urge incontinence and dribbling can have a profound psychological affect on the patient's quality of life which can cause tremendous emotional burden, namely anxiety, stress and a high level of depression[1,2].

There are several measures of depression such as the Beck Depression Inventory (BDI), Hamilton Depression Rating Scale[3]. All these have been widely used in other countries and therefore need to be validated for the local population. The Beck Depression Inventory (BDI) has thus become a commonly used instrument in multicentre, international clinical trials to assess psychiatric disorders. The BDI, developed by Beck & Steer, (1986) has been widely used in the studies of depression in many countries both in the community-based studies as well as the clinical studies[4].

The study was conducted at the University Hospital Kuala Lumpur and was aimed to assess the reliability and validity of the Beck Depression Inventory (BDI) among urological patients.

Materials and Methods

Study sample

The patients were selected based on the inclusion and exclusion criteria. For patients with LUTS, the inclusion criteria were patients who are stable,

This article was accepted: 11 June 2001

whereas the exclusion criteria were patients who were not treated with surgical and medical treatment for lower urinary tract symptoms prior to this study. Patients less than 40 years old who were illiterate were excluded in the study, as were patients with any chronic and acute diseases. For the control group, the inclusion criteria include patients who were free from all major chronic and acute diseases while the exclusion criteria, those with urological problems except renal stone with minimal severity.

Instrument

The BDI is designed to assess the severity of depression among the psychiatric patients as well as possible depression in normal population. This inventory measures cognitive, affective, somatic symptoms, neurovegetative and endogenous aspects of depression[4]. The self-report questionnaire is rated on a four-point scale ranging from 0 (no symptom) to 3 (severe symptom). The items were chosen to assess only the severity of depression, and were not selected to reflect any particular theory of depression. The 21 symptoms and attitudes assessed by the BDI include: (1) Mood; (2) Pessimism; (3) Sense of failure; (4) Self-dissatisfaction; (5) Guilt; (6) Punishment; (7) Self-dislike; (8) Self-accusation; (9) Suicidal ideas; (10) Crying; (11) Irritability; (12) Social withdrawal; (13) Indecisiveness; (14) Body image change; (15) Work difficulty; (16) Insomnia; (17) Fatigability; (18) Loss of appetite; (19) Weight loss; (20) Somatic preoccupation; and (21) Loss of libido.

The BDI score was obtained by summing the ratings given by the interviewer for each of the 21 items. The overall depression scores range from 0 to 63 and normally divided into four categories. Scores of 0 to 9 were considered within the normal range or asymptomatic, scores of 10 to 15 indicate mild depression, scores of 16 to 23 indicate moderate depression and scores of 24 to 63 indicate extremely severe depression. The questionnaire is easily administered and takes about 5 - 10 minutes to complete[5].

The psychometric properties of the BDI were assessed in three different samples. Validity and reliability were studied in a group of patients with LUTS (N=108) and control patients without LUTS (N=50) and responsiveness was assessed in a group of patients admitted for transurethral resection of the prostate (TURP)(N=79). Management of cases based on clinical criteria such as medical history, physical and rectal examinations was entirely done by a Urologist (A.H.R).

Procedure

The study protocol was approved by the Ethics Committee, University Hospital Kuala Lumpur. Written informed consent was obtained from all participants after being explained the nature of the study. The patients were then required to complete the BDI. All questionnaires were self-administered as well as assisted guidance if necessary by one of the authors (K.F.Q). All patients included in the validity study were scheduled for twelve weeks after the first administration of the BDI. In the sensitivity to change study, patients completed the questionnaires a week prior to surgical treatment and were retested at three months follow up.

Data Analysis

The Cronbach's alpha coefficient was used to assess the internal consistency of the BDI[6]. Test-retest reliability was assessed using the intraclass correlation coefficient (ICC), which is derived from the analysis of variance (ANOVA) model. Values of ICC varies from 1 (perfectly reliable) to 0 (totally unreliable)[7]. Responsiveness was analyzed by calculating the mean difference between BDI before and after TURP and dividing it by the mean standard deviation of the scores before TURP (effect size)[8].

Mean differences in BDI scores before and after TURP was also calculated for each individual item by means of a paired t test or by dividing it by the mean standard deviation of stable patients (Guyatt statistic)[9].

Sensitivity of the BDI was assessed by comparing between the means of pre-treatment and post-treatment item scores of patients who have undergone TURP whereas specificity was assessed by comparing the pre-treatment and post-treatment item scores in subjects rated as control.

Results

237 patients consisting of medical, surgical and control group participated in this study. The mean age for the medical group was 63.67 years old (SD=8.57), surgical group was 70.01 years old (SD=8.17) and control group was 50.04 years old (SD=12.29). By ethnicity, the Chinese formed the largest ethnic groups in all the three groups. Patients with LUTS in the surgical group were significantly more depressed as compared to control patients ($p<0.0001$). Of the 237, 108 LUTS patients has the total BDI mean scores of 8.07 (SD=5.58) compared to the 79 patients undergoing TURP (ξ=14.43, SD=7.87)($p<0.0001$). This showed that prior to treatment, the surgical group was significantly more depressed than the medical and control patients.

Internal consistency for the BDI was high for all the items of BDI indicating a high level of homogeneity among items in the scale. The internal consistency showed the resulting values of Cronbach's alpha for the scale when individual item were excluded from the analysis. Test-retest reliability was assessed in 108 patients after 12 weeks interval and the total scores of BDI had an ICC of 0.85 ($p<0.001$) (Table I). The total scores of BDI before (8.07) and after (7.67) showed most of the patients were not depressed according to the classification of the severity of BDI.

Responsiveness was assessed in patients undergoing TURP. Table II showed the pre and post scores, mean difference, effect size, and the Guyatt statistic for individual items, and for total scores. Before treatment, patients in the surgical group were depressed and subsequently improved after treatment. The mean pre intervention score on the total BDI (ξ= 14.43, sd=7.87) was significantly higher ($p<0.0001$) than the mean post intervention score (ξ=10.04, sd=5.41), giving an average improvement after TURP on the depression level of 4.39. Overall effect size and responsiveness were found to be high indicating improvement in TURP-induced urinary symptoms in these patients.

In terms of treatment responsiveness, sensitivity and specificity of the instrument was evaluated by comparing the change between baseline and end point scores following treatment. All items of the BDI demonstrated a high degree of sensitivity and specificity to the effects of treatment (Table III). Significant changes were observed across more than half of the BDI items in the LUTS group. The lowest magnitude of change was noted in item 21. In contrast, except for item 19 ($p<0.001$), none of the comparison in the treatment of the control subjects approached significance.

The result of the discriminant validity between the surgical and the control group is shown in Table IV. Significant differences between the two groups were noted in the total scores of BDI. The surgical group tend to be more depressed (mean BDI=14.43) than the control group (mean BDI=7.76) who were asymptomatic.

Discussion

The psychometric properties of the BDI validated here and other countries showed that they are virtually identical with respect to their measurement properties[4]. Although test-retest exhibited statistical significant at $p<0.05$ at some items of BDI, this is not surprising because the test-retest was done at 12 weeks interval, thus allowing the urinary symptoms and depression to improve or worsen. In contrast, the minimal changes of symptoms and depression would occur if test-retest were done at one week, two week or the latest one month after the initial assessment. The reasons for the test-retest to be carried out at twelve weeks interval because most patients in the TURP group would then have achieved the maximum benefit or total symptom improvement and this would make comparisons between the patients with LUTS and patients undergoing TURP much easier.

Table I
Validity and Reliability:
Mean Test-Retest Score, Intraclass Correlation Coefficient and Internal Consistency for Individual BDI items

BDI Items	Intraclass Correlation Coefficient* (ICC)	Internal Consistency**	Mean Test Score	SD	Mean Retest Score	SD	Mean Difference	SD	95% Confidence Interval Lower	Upper
1	0.67	0.67	0.32	0.47	0.30	0.50	0.028	0.48	-0.12	0.06
2	0.87	0.87	0.32	0.47	0.45	0.85	0.13	0.88	-0.04	0.30
3	0.82	0.82	0.14	0.44	0.18	0.47	0.037	0.36	-0.03	0.11
4	0.73	0.73	0.54	0.60	0.56	0.59	0.019	0.55	-0.08	0.12
5	0.69	0.69	0.35	0.50	0.34	0.53	0.0093	0.50	-0.10	0.09
6	0.75	0.75	0.28	0.53	0.21	0.63	0.065	0.52	-0.16	0.03
7	0.66	0.66	0.27	0.44	0.21	0.41	0.056	0.43	-0.14	0.03
8	0.73	0.73	0.46	0.62	0.39	0.56	0.074	0.54	-0.18	0.03
9	0.56	0.56	0.028	0.16	0.019	0.13	0.0093	0.17	-0.04	0.02
10	0.69	0.69	0.34	0.95	0.33	0.95	0.0093	0.92	-0.18	0.17
11	0.70	0.70	0.72	1.05	0.78	0.98	0.056	0.97	-0.13	0.24
12	0.68	0.68	0.21	0.41	0.29	0.45	0.074	0.43	-0.007	0.15
13	0.72	0.72	0.29	0.53	0.35	0.62	0.065	0.53	-0.04	0.17
14	0.79	0.79	0.16	0.44	0.18	0.43	0.019	0.36	-0.05	0.09
15	0.84	0.84	0.47	0.63	0.56	0.63	0.093	0.46	0.004	0.18
16	0.77	0.77	0.59	0.71	0.55	0.67	0.046	0.60	-0.16	0.07
17	0.84	0.84	0.43	0.60	0.51	0.66	0.074	0.47	-0.01	0.16
18	0.60	0.61	0.24	0.45	0.16	0.37	0.083	0.43	-0.17	-0.0003
19	0.83	0.84	0.20	0.51	0.13	0.43	0.074	0.35	-0.14	-0.006
20	0.76	0.77	0.53	0.70	0.39	0.61	0.14***	0.57	-0.25	-0.03
21	0.82	0.82	1.02	0.96	0.86	0.95	0.16***	0.74	-0.30	-0.02
Total scores	0.85	0.85	8.07	5.58	7.67	5.69	0.41***	4.08	-0.37	-0.37

p<0.001 for all ICCs(*)
Cronbach's alpha: note that Cronbach's alpha value given for each item represent the effect of removing that item from the calculation of the alpha value (eg if item 1 is omitted, the resulting value for the scale is 0.67, if item 2 is omitted, it is 0.87, and so forth)(**)
t test for paired comparisons significant (***)

Table II
Responsiveness: Mean Scores Before and After TURP, Effect Size and Guyatt Statistic

BDI Items	PreTURP Mean	SD	PostTURP Mean	SD	Mean Differences*	SD	Effect Size	Guyatt statistic
1	0.63	0.48	0.32	0.47	0.32	0.67	0.67	0.68
2	0.77	0.95	0.57	0.89	0.20*	1.36	0.21	0.43
3	0.27	0.55	0.076	0.31	0.19	0.64	0.35	0.43
4	1.01	0.61	0.66	0.57	0.35	0.89	0.57	0.58
5	0.68	0.54	0.38	0.49	0.30	0.79	0.56	0.60
6	0.49	0.57	0.23	0.50	0.27	0.73	0.47	0.51
7	0.59	0.49	0.21	0.41	0.38	0.67	0.78	0.86
8	0.67	0.63	0.37	0.56	0.30	0.79	0.48	0.48
9	0.025	0.16	0	0	0.025*	0.16	0.16	0.16
10	0.63	1.18	0.39	0.97	0.24*	1.59	0.20	0.25
11	0.81	1.03	0.51	0.84	0.30	1.36	0.29	0.29
12	0.82	0.69	0.63	0.60	0.19*	0.95	0.28	0.46
13	0.80	0.67	0.67	0.59	0.13*	0.95	0.19	0.25
14	0.28	0.48	0.21	0.41	0.063*	0.63	0.13	0.14
15	1.02	0.72	0.90	0.65	0.13*	0.94	0.18	0.21
16	1.15	0.75	0.77	0.68	0.38	0.94	0.56	0.54
17	0.61	0.54	0.47	0.57	0.14*	0.84	0.26	0.23
18	0.46	0.59	0.29	0.46	0.16*	0.77	0.27	0.36
19	0.51	0.77	0.23	0.42	0.28	0.90	0.36	0.55
20	0.77	0.75	0.33	0.55	0.44	0.84	0.59	0.63
21	1.51	1.01	1.54	0.94	0.038*	1.44	0.038	0.04
Total scores	14.43	7.87	10.04	5.41	4.39	5.90	0.56	0.79

Effect size=Mean difference/SD PreTURP
Guyatt statistics=Mean difference/SD of stable LUTS patients (medication group)
** t test for paired comparisons not significant*

BDI with its ability to discriminate between patients with LUTS and those without showed high levels of sensitivity and specificity likewise with the discriminant validity. The reasonably large effect size obtained indicated a high degree of sensitivity to change, likewise with other studies[8].

This study showed that the LUTS significantly contributed to the increased of the depression level especially those with severe LUTS. The improvement of the symptoms following treatment resulted in the improvement of the depression.

These findings provide substantial assurance that scores obtained using the BDI are reliable. The validity and reliability scores were consistent with scores of BDI validated in various countries[5].

Conclusion

This study showed that the intraclass correlation coefficient for total scores of the items of the BDI indicates high intraclass reliability and the high Cronbach's alpha coefficient suggest that BDI exhibits good internal consistency. BDI seems to be a valid,

Table III
BDI Items Characteristics of Patients Undergoing TURP and the Control Group: Sensitivity and Specificity

Sensitivity Items	N	Mean Changes	SEM	t statistics	p value
1	79	0.32	0.075	4.19	0.0001
2	79	0.20	0.15	1.32	0.19
3	79	0.19	0.072	2.63	0.01
4	79	0.35	0.10	3.53	0.001
5	79	0.30	0.089	3.42	0.001
6	79	0.27	0.082	3.24	0.002
7	79	0.38	0.075	5.07	0.0001
8	79	0.30	0.089	3.42	0.001
9	79	0.025	0.018	1.42	0.159
10	79	0.24	0.18	1.35	0.182
11	79	0.30	0.15	1.98	0.05
12	79	0.19	0.11	1.78	0.079
13	79	0.13	0.11	1.18	0.241
14	79	0.063	0.071	0.90	0.373
15	79	0.13	0.11	1.20	0.234
16	79	0.38	0.11	3.60	0.01
17	79	0.14	0.095	1.47	0.146
18	79	0.16	0.087	1.89	0.063
19	79	0.28	0.10	2.74	0.008
20	79	0.44	0.095	4.67	0.001
21	79	0.038	0.16	0.235	0.815
Total scores	79	4.39	0.66	6.61	0.0001
1	50	0.02	0.073	0.275	0.785
2	50	0.12	0.84	1.43	0.159
3	50	0	0.081	0	1
4	50	0.04	0.075	0.531	0.598
5	50	0.04	0.049	0.814	0.42
6	50	0.14	0.099	1.41	0.164
7	50	0.04	0.057	0.704	0.485
8	50	0.16	0.072	2.22	0.031
9	50	0.02	0.02	1	0.322
10	50	0.18	0.12	1.46	0.151
11	50	0	0.16	0	1
12	50	0.10	0.059	1.70	0.096
13	50	0.06	0.053	1.14	0.261
14	50	0	0.064	0	1
15	50	0.10	0.082	1.22	0.229
16	50	0.04	0.075	0.53	0.598
17	50	0	0.076	0	1
18	50	0.02	0.053	0.375	0.709
19	50	0.22	0.082	2.67	0.01
20	50	0.06	0.088	0.685	0.497
21	50	0.08	0.075	1.07	0.29
Total scores	50	0.68	0.73	0.925	0.36

Table IV
BDI Items Characteristics: Discriminant Validity

Items	Pre TURP Patients		Control						
	Mean	SEM	Mean	SEM	Mean Difference	SEM	95% Confidence Interval Lower	Higher	p value
1	0.63	0.055	0.46	0.082	0.17	0.099	0.022	0.37	0.08
2	0.77	0.11	0.42	0.099	0.35	0.15	0.064	0.64	0.05
3	0.27	0.062	0.22	0.077	0.046	0.099	0.15	0.24	0.64
4	1.01	0.069	0.54	0.087	0.47	0.11	0.25	0.69	0.0001
5	0.68	0.061	0.30	0.065	0.38	0.093	0.20	0.57	0.0001
6	0.49	0.065	0.32	0.097	0.17	0.11	0.048	0.39	0.12
7	0.59	0.056	0.28	0.064	0.32	0.085	0.15	0.48	0.0001
8	0.67	0.071	0.50	0.096	0.17	0.12	0.062	0.40	0.15
9	0.025	0.018	0	0	0.025	0.018	0.01	0.061	0.16
10	0.63	0.13	0.26	0.11	0.37	0.17	0.037	0.71	0.05
11	0.81	0.11	0.70	0.15	0.11	0.19	0.26	0.48	0.56
12	0.82	0.078	0.30	0.065	0.52	0.11	0.30	0.74	0.0001
13	0.80	0.075	0.30	0.66	0.50	0.11	0.28	0.71	0.0001
14	0.28	0.054	0.10	0.051	0.18	0.074	0.031	0.33	0.05
15	1.02	0.08	0.48	0.077	0.54	0.12	0.31	0.78	0.0001
16	1.15	0.085	0.52	0.087	0.63	0.13	0.38	0.88	0.0001
17	0.61	0.061	0.48	0.077	0.13	0.098	0.066	0.32	0.19
18	0.46	0.067	0.16	0.052	0.30	0.085	0.13	0.46	0.001
19	0.51	0.086	0.36	0.086	0.15	0.13	0.11	0.40	0.26
20	0.77	0.084	0.62	0.11	0.15	0.14	0.12	0.42	0.26
21	1.51	0.11	0.50	0.10	1.01	0.15	0.70	1.31	0.0001
Totol scores	14.43	0.88	7.76	0.74	6.67	1.15	4.39	8.95	0.0001

useful and a reliable instrument for assessing the severity of LUTS in the Malaysian urological population.

Acknowledgements

The authors would like to express their gratitude to the University Malaya's Research and Development Management Unit for providing the research grant for this study.

References

1. Quek KF, Low WY, Razack AH, Loh CS. Psychological effects of treatment for lower urinary tract symptoms. British Journal Urology International 2000; 86: 630-33.
2. Lukacs B, Leplege A, McCarthy C, Comet D. Construction and validation of a BPH specific health related quality of life scale, for medical outcome research studies. In Proceedings of the Second International Consultation on Benign Prostatic Hyperplasia (BPH) (eds ATK. Cockett , S. Khoury, Y. Aso *et al.*), Paris.: SCI. 1994; 139-43.
3. Hamilton M. A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry; 1960; 23: 56-61.
4. Steer RA, Beck AT and Garrison B. Applications of the Beck Depression Inventory; in Sartorius N, Ban TA (eds): Assessment of Depression. Berlin, Springer. 1986; 121-42.
5. Beck AT and Steer RA. Beck Depression Inventory: Manual. The Psychological Corporation, Harcourt Brace Jovanich Inc., San Diego. 1986.
6. Cronbach LJ. Coefficient alpha and the internal structure of test. Psychometrika 1951; 16: 297.
7. Deyo RA, Dichr P, Patrick DL. Reproducibility and responsiveness of health status measures. Control Clinical Trials 1991; 12 (supp 1): 142S-158S.
8. Cohen J. Statistical Power analysis for the Behavioral Sciences. New York: Academic Press. 1977; 75-105.
9. Guyatt G, Walter S, Norman G. Measuring changes overtime: assessing the usefulness of evaluative instruments. Journal Chronic Disease 1987; 40: 171-78.