Eur Arch Psychiatry Clin Neurosci (2001) 251: Suppl. 2, II/27–II/31

Marzia Bonsignore · Katrin Barkow · Frank Jessen · Reinhard Heun

Validity of the five-item WHO Well-Being Index (WHO-5) in an elderly population

Abstract *Background* Depression has a high prevalence in the elderly population; however it often remains undetected. The WHO 5-item Well-Being Index (WHO-5) is a short screening instrument for the detection of depression in the general population, which has not yet been evaluated. The goals of the present study were:1) to assess the internal and external validity of WHO-5 and 2) to compare the two recent versions of WHO-5. *Study population and methods* 367 subjects above 50 years of age were examined with the WHO-5. ICD-10 diagnoses were made using a structured interview (CIDI). The internal validity of the well-being index was evaluated by calculating Loevinger's and Mokken's homogeneity coefficients. External validity for detection of depression was evaluated by ROC analysis. *Results* The scale was sufficiently homogeneous (Loevinger's coefficient: version 1 = 0.38, version 2 = 0.47; Mokken coefficient > 0.3 in nearly all items). ROC analysis showed that both versions adequately detected depression. Version 1 additionally detected anxiety disorders, version 2 being more specific for detection of depression. *Conclusion* The WHO-5 showed a good internal and external validity. The second version is a stronger scale and was more specific for the detection of depression. The WHO-5 is an useful instrument for identifying elderly subjects with depression.

Key words Depression · WHO Well-Being Index · elderly general population· anxiety · screening

When meeting Per Bech, one has to admire his great enthusiasm for clinical and epidemiological research aimed at helping people.

Reinhard Heun (✉) · Marzia Bonsignore
Katrin Barkow · Frank Jessen
Department of Psychiatry
University of Bonn
Sigmund-Freud-Str. 25, 53115 Bonn, Germany
E-Mail: heun@uni-bonn.de
Tel.: +49 228 287 6526

Introduction

The prevalence of depression in the elderly general population is estimated to be around 15% (5, 13, 19). Considering that many affected individuals remain undetected (2, 11, 18), an instrument for screening subjects at risk might be helpful. A number of structured interviews and tests have been developed (1, 20, 24), which are all quite extensive and time consuming. As the acceptance of such tests in both the patient and examiner rises with their brevity, a broader and quicker screening instrument is required. A candidate is the WHO Well-Being Index, initially developed to measure well-being in a WHO study on different therapies for patients with diabetes. The design of the questionnaire was based on Zung's self-rating scales for depression, anxiety and psychological distress (25, 26) including additional items. The first version of Well-Being Index consisted of 28 items, but following psychometrical analysis of the first study data it was reduced to 22 (6). In the diabetes study the scale proved to be a reliable tool for the measurement of health-related personal well-being (6) which has also been confirmed for testing populations with other chronic disorders (22). Additionally, in a sample from the general elderly population this scale has been shown to discriminate between subjects with and without psychiatric disorders (9). Following additional psychometrical analysis, Bech et al. first proposed a further shortened version of 10 items (2) in 1995 (WHO-10), then of 5 items (3) the same year (WHO-5), and finally a reviewed version of the WHO-5 in 1998 (23). The second version (1998) has been proposed for the use of screening for depression in the general population (23). Since the two versions of the 5-item Well-Being Index have both not yet been evaluated, it was the primary aim of the present study to investigate their internal and external validity and to compare the two versions. We chose a sample with an increased risk for depression, i.e., elderly relatives of patients with psychiatric disorders.

Materials and methods

Sample description and recruitment procedure

The sample has previously been recruited in the context of a large family study in the elderly general population; the methods of this family study have been described in detail elsewhere (10). Briefly, the present sample consisted of 367 first degree relatives aged above 50, who were re-examined 5 years after the first examination and were asked to additionally complete the WHO-5. There were 111 patients who completed version 1 (1995) and 256 who completed version 2 (1998). There was no difference in gender ($\chi^2 = 0.860$, df = 1, p = 0.354) or age (t = 0.142, df = 365, p = 0.887) between the two subgroups completing the two different versions. In total, 28 subjects (7.6%) had a current psychiatric ICD-10 disorder at the time of interview. Two subjects had more than one current diagnosis. Depression was observed in 14 subjects (3.8%; bipolar mood disorder: n = 1, depressive episode: n = 7, recurrent depressive disorder: n = 6), anxiety disorder in 11 subjects (3.0%; phobic disorder: n = 8, panic disorder: n = 3). Five persons suffered from another current psychiatric disorder (1.4%; dementia: n = 2, alcohol dependence syndrome: n = 2, somatoform disorder: n = 1). Of the subjects with no current diagnosis, 48 (14.2%) had experienced depression and 16 (4.7%) anxiety disorders in their past. The demographic characteristics of the 367 participants who completed the WHO-5 are given in Table 1.

Tab. 1 Description of study sample and distribution of diagnoses

	No current diagnosis	Current depression	Current anxiety	All
N	339	14	11	367
Gender [N (% female)]	211 (62.2%)	8 (57.1%)	8 (72.7%)	228 (62.1%)
Age [years: mean ± SD]	70.1 ± 8.1	70.8 ± 9.4	68.4 ± 6.3	70.2 ± 8.0
Duration of formal education [years: mean ± SD]	9.6 ± 2.1	8.9 ± 2.1	9.3 ± 1.4	9.5 ± 2.1
Social situation: Living with a partner [N (%)]	218 (64.3%)	7 (50.0%)	7 (63.6%)	234 (63.8%)
Having at least 1 children [N (%)]	290 (85.5%)	11 (78.6%)	9 (81.8%)	314 (85.6%)

Tab. 2 WHO-5 version 1 (1995)

	All of the time	More than half of the time	Less than half of the time	At no time
I feel downhearted and sad	0	1	2	3
I feel calm and can sit still easily	3	2	1	0
I feel energetic, active or vigorous	3	2	1	0
I wake up feeling fresh and rested	3	2	1	0
My daily life is full of things that were interesting to me	3	2	1	0

User's instruction: Please circle a number on each of the following statements to indicate how often you feel each of them has applied to you in the last few weeks.

Assessment of diagnosis and well-being

Psychiatric diagnoses were made using the Composite International Diagnostic Interview (CIDI) (21), a structured diagnostic interview for the detection of various psychiatric disorders according to the ICD-10 definition, that has been designed to be performed by lay interviewers. The two versions of the WHO-5 are shown in Tables 2 and 3. We used a German translation: version 1 had been translated by the senior author of this paper (R.H.), version 2 represented the official translation by the WHO. The main difference appears in the first item, which has been changed from a negative formulation into a positive one in the second version, with the aim to obtain a more homogenous scale. Further differences are the slightly different formulation of the other items and a different scoring system, which allows more graduation

Tab. 3 WHO-5 version 2 (1998)

Over the last two weeks	All of the time	Most of the time	More than half of the time	Less than half of the time	Some of the time	At no time
I have felt cheerful and in good spirits	5	4	3	2	1	0
I have felt calm and relaxed	5	4	3	2	1	0
I have felt active and vigorous	5	4	3	2	1	0
I woke up feeling fresh and rested	5	4	3	2	1	0
My daily life has been filled with things that interested me	5	4	3	2	1	0

User's instruction: Please circle a number on each of the following statements to indicate how often you feel each of them has applied to you in the last few weeks.

Tab. 4 Loevinger's coefficient of homogeneity and Mokken scores for individual items of the WHO Well-Being Index WHO-5

WHO-5 Version 1 (1998)		WHO-5 Version 2 (1999)	
Loevinger's coefficient of homogeneity: H = 0.38		Loevinger's coefficient of homogeneity: H = 0.47	
Item	Hi (Mokken score)	Item	Hi (Mokken score)
I feel downhearted and sad	0.46	I have felt cheerful and in good spirits	0.50
I feel calm and can sit still easily	0.23	I have felt calm and relaxed	0.50
I feel energetic, active or vigorous	0.41	I have felt active and vigorous	0.50
I wake up feeling fresh and rested	0.46	I woke up feeling fresh and rested	0.42
My daily life is full of things that were interesting to me	0.32	My daily life has been filled with things that interested me	0.45

in the second version. A sum score is calculated by adding up the figures of the five answers; it ranges from 0 to 15 for version 1, and from 0 to 25 for version 2. A high sum score indicates a status of optimal well-being.

The interviewers were medical students in their sixth year of study. Their training consisted of a 4-week stay in a gerontopsychiatric ward and intensive training with interview modules. Interviewers were continuously supervised during the study.

Statistical analysis

Assessment of the internal validity

The internal consistency of the two different WHO-5 was assessed by calculation of the Loevinger's coefficient and Mokken's coefficient of homogeneity. The Mokken coefficient is calculated for each individual item and indicates to which extent the respective item lies on the same dimension as the other items (16). We have used the procedure for the analysis of polychotomous items (7). A coefficient of 0.3 to 0.39 is regarded as acceptable, while a coefficient of 0.4 or more indicates an item that is adequately included in a scale. The Loevinger coefficient (H) is a measure of scalability for the whole scale as it indicates to which extent the items represent just one dimension (14). For the classification of scales on the basis of coefficients, Mokken suggests the following system: $H \geq 0.5$ indicates a strong scale; $0.4 \leq H < 0.5$ a medium scale, $0.3 \leq H < 0.4$ a weak scale (17).

Assessment of external validity

The ability of WHO-5 to detect elderly people with depression was estimated by performing receiver operating characteristic analyses (ROC) (12). This method was developed in the context of signal-detection theory and has been adopted for use in biological and behavioral research (8). The area under the ROC curve is an indicator of test performance, a value of 0.5 indicates that the examined instrument does not discriminate between the states of interest. The optimal cut-off score is defined as the point on the ROC curve, which is the furthest from the diagonal. The areas under the ROC curves were compared using Z-tests (15).

Results

Internal validity

The consistency of both versions was adequate, as indicated by Mokken coefficients (see Table 4). One item of version 1 did not fit adequately in the scale ($Hi < 0.3$). The coefficients of version 2 were all above 0.4 and were higher than the ones of version 1. According to Mokken (17), the second version would be judged as a medium scale, while the first version is a weak scale. Contrary to our expectations, the changing of the first item from a negative formulation to a positive one did not affect the Mokken score of this item.

External validity

In agreement with our expectations, subjects with current depression had lower total WHO-5 scores than subjects with no current depression (version 1: 7.5 vs. 12.3, $t = 3.6$, $df = 109$, $p = 0.001$, version 2: 9.1 vs 17.7, $t = 5.36$, $df = 254$, $p < 0.001$). Subjects with anxiety disorders who completed the first version had also lower total scores than their corresponding reference group (7.3 vs. 12.3, $t = 3.8$, $df = 109$, $p = 0.000$). However, version 2 did not discriminate between subjects with and without anxiety disorders (anxiety disorders: 14.3, no anxiety disorders. 17.4, $t = 1.57$, $df = 254$, $p = 0.118$). ROC analysis (see Fig. 1) revealed that both versions have a good external validity for the detection of depression: the area under the ROC curve (AUC) was statistically different from 0.5. Comparing the AUC of the two version with each other showed that they detected depression equally well ($z = 0.82$, $df = 1$, $p = 0.412$). In contrast to version 2, version 1 detected also subjects with anxiety disorders (version 1 $p = 0.007$, version 2 $p = 0.200$). In agreement with our expectations, both versions of WHO-5 did not detect lifetime diagnoses (i.e., previous, but no current diagnoses)

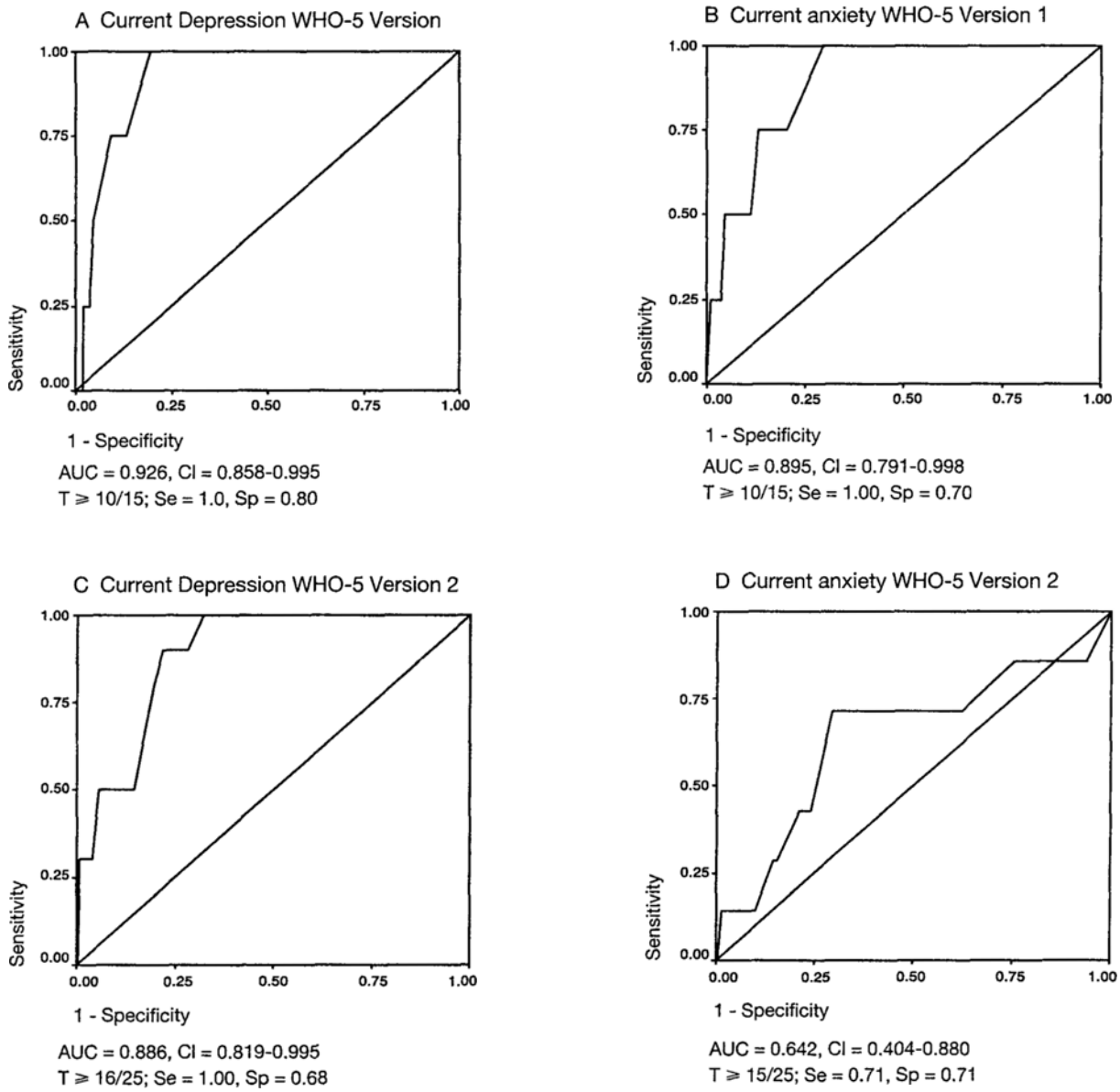

Fig. 1 Receiver Operating Characteristic (ROC) curves for the detection of depression (version 1: **A**, version 2: **C**) and anxiety disorder (version 1: **B**, version 2: **D**), including the area under the curve (AUC), the 95% confidence interval (CI) and the optimal threshold score (T) with corresponding sensitivity (Se) and specificity (Sp) values.

of depression or anxiety disorders (AUC = 0.5, $p > 0.2$, ROC results not presented).

Conclusion

The WHO-5 has a good internal consistency and homogeneity, equivalent to the longer precursor versions of Well-Being Index (3, 9). Due to its higher Loevinger and Mokken coefficients the second version seemed superior to the first version. The external validity ranked highly, as indicated by ROC analyses. WHO-5 scores were related to the absence or presence of depression. In addition to depression, version 1 significantly detected subjects with anxiety disorders, therefore version 2 is to be regarded as more specific for the detection of depression. These results suggest that the second version may be preferred in the future as a screening instrument for depression. The results are restricted to an elderly population at risk for psychiatric disorders, the transferability to other samples needs to be assessed in future.

Acknowledgment This paper has been prepared to celebrate the 60th birthday of P. Bech.

References

1. Alexopoulos GS, Abrams RC, Young RC et al. (1988) Cornell Scale for Depression and Dementia. Biol Psychiat 23: 271–284
2. Angst J (1999) Major depression in 1998: are we providing optimal therapy? J Clin Psychiatry 60 (Suppl 6): 5–9
3. Bech P (1996) The Bech, Hamilton and Zung Scales for Mood Disorders: Screening and Listening. 2nd Edition, Springer, Berlin, pp 395–424
4. Bech P, Staehr-Johansen K, Gudex C (1996) The WHO (ten) Well-Being Index: validation in diabetes. Psychother Psychosom 665: 183–190
5. Bond J (1987) Psychiatric illness in later life: a study of prevalence in a Scottish population. Int J Geriat Psychiat 2: 39–57
6. Bradley C (1994) Handbook of Psychology and Diabetes: A Guide to Psychological Measurement in Diabetes Research and Practice. Harwood, London
7. Debets P, Brouwer E (1989) User's Manual MSP: A Program for Mokken Scale Analysis for Polytomous Items (version 1.5). Iec ProGamma, Groningen
8. Erdreich LS, Lee ET (1981) Use of relative operating characteristic analysis in epidemiology. A method for dealing with subjective judgement. Am J Epidemiol 114: 649–662
9. Heun R, Burkardt M, Maier W, Bech P (1999) Internal and external validity of the WHO Well-Being Scale in the elderly general population. Acta Psychiatr Scand 99: 171–178
10. Heun R, Papassotiropoulos A, Jessen F, Maier W, Breitner JCS (2001) A family study of Alzheimer disease and early- and late-onset depression in elderly patients. Arch Gen Psychiatry 58: 190–196
11. Katon W (1987) The epidemiology of depression in medical care. Int J Psychiatr Med 17: 93–112
12. Kraemer HC (1988) Assessment of 2x2 associations: generalisation of signal detection methodology. Am Stat 42: 37–49
13. Livingston G, Hawkins A, Graham N et al. (1990) The Gospel Oak study: prevalence rates of dementia, depression and activity limitation among elderly residents in Inner London. Psychol Med 20: 137–146
14. Loevinger J (1948) The technique of homogenous test compared with some aspects of scale analysis and factor analysis. Psychol Bull 45: 507–529
15. McClish DK (1987) Comparing the areas under more than two independent ROC curves. Med Decis Making 7: 149–155
16. Mokken RJ (1971) A Theory and Procedure of Scale Analysis with Applications in Political Research. Mouton, Paris
17. Mokken RJ (1982) A nonparametric approach to the analysis of dichotomous item responses. Appl Psychol Measurement 6: 417–430
18. Ormel J, Brink WVD, Koeter MWJ et al. (1990) Recognition, management and outcome of psychological disorders in primary care: a naturalistic follow-up study. Psychol Med 20: 909–923
19. Pahkala K, Kesti E, Köngäs-Saviaro PJ et al. (1995) Prevalence of depression in an aged population in Finland. Soc Psychiat Epidemol 30: 99–106
20. Sunderland T, Alterman IS, Yount D et al. (1988) A new scale for the assessment of depressed mood in demented patients. Am J Psychiat 145: 955–959
21. World Health Organisation (1990) Composite International Diagnostic Interview. World Health Organisation, Division of Mental Health, Geneva
22. World Health Organization: Regional Office for Europe (1995) Quality Assurance Indicators in Mental Health Care. Consensus meeting, Stockholm
23. World Health Organization: Regional Office for Europe (1998) Well-Being measures in primary health care: The DepCare Project. Consensus meeting, Stockholm
24. Yesavage J, Brink T, Rose T et al. (1983) Development and validation of a geriatric depression screening scale. A preliminary report. J Psychiat Res 17: 37–49
25. Zung WWK (1974) The measurement of affects: depression and anxiety. In: Pichot P (ed) Psychological Measurement in Psychopathology. Karger, Basel, pp 170–188
26. Zung WWK (1983) A self-rating pain and distress scale. Psychosomatics 24: 887–894