

Regular Article

Reliability and validity of the Thai version of the WHO-Five Well-Being Index in primary care patients

Ratana Saipanish, MD,^{1*} Manote Lotrakul, MD¹ and Sutida Sumrithe, MD²Departments of ¹Psychiatry and ²Family Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand

Aims: Because of the high patient load in Thailand, we need a practical measurement to help primary physicians detect depression. This study aimed to examine the reliability and validity of the Thai version of the World Health Organization-Five Well-Being Index (WHO-5-T), which is short and easy to use as a screening tool for major depression in primary care patients.

Methods: The English version of the WHO-Five Well-Being Index was translated into Thai. Back-translations, cross-cultural adaptation and field testing of the pre-final version with final adjustments were performed accordingly. The WHO-5-T was administered randomly to 300 patients in our primary care clinic. Then the patients were further assessed using the Mini International Neuropsychiatric Interview and the Hamilton Rating Scale for Depression as the gold standard of diagnosis and symptom severity, respectively.

Results: Completed data were obtained from 274 respondents. Their mean age was 44.6 years

[standard deviation (SD) = 14.7] and 73.7% of them were female. The mean WHO-5-T score was 14.32 (SD = 5.26). The WHO-5-T had a satisfactory internal consistency (Cronbach's alpha = 0.87) and showed moderate convergent validity with the Hamilton Rating Scale for Depression ($r = -0.54$; $P < 0.001$). The optimal cut-off score of the WHO-5-T < 12 revealed a sensitivity of 0.89 and a specificity of 0.71 in detecting depression. The area under the curve in this study was 0.86 (SD = 0.03, 95% confidence interval 0.81 to 0.89).

Conclusions: The Thai version of the WHO-Five Well-Being Index was found to be a reliable and valid self-assessment to screen for major depression in primary care setting at a cut-off point of < 12 .

Key words: depression, primary care, screening, WHO-Five Well-Being Index.

DEPRESSION PRODUCES A significant proportion of disability and incurs significant public health and economic costs.¹ Even though depression is a treatable illness, it is frequently under recognized and inadequately treated, especially in primary care settings.^{2,3} The condition is even worse in Thailand, where primary physicians are faced

with a high patient load. More than 70% of Thai primary physicians have to take care of more than 50 patients a day.⁴ Hence, it is not only difficult to detect mental illness, including depression, but also to provide effective treatment to the particular patient.

In this difficult situation, we have been looking for case-finding instruments that might help primary physicians identify depressive disorder in their patients. Although a few depression screening questionnaires in the Thai language have been developed, such as the Health-Related Self-Report (HRSR) scale,⁵ the Thai Depression Inventory (TDI)⁶ and the Center for Epidemiological Studies Depression Scale

*Correspondence: Ratana Saipanish, MD, Department of Psychiatry, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand. Email: rarsp@mahidol.ac.th
Received 13 December 2007; revised 24 October 2008; accepted 1 December 2008.

(CES-D),⁷ all of them are too time-consuming for routine uses in our high patient load settings. A new case-detecting instrument that is brief and easy to administer to improve the recognition of depression in primary care is still needed.

One of the questionnaires we considered to suit our situation is the second version of the World Health Organization – Five Well-Being Index (WHO-5), a short and quick screening tool for detecting depression. It is a self-report measure that consists of five Well-Being Index items. It is derived from the original version of the Well-Being Index to measure health-related personal well-being, which consisted of 28 items. Following psychometric analysis, it was shortened to be the first version of WHO-5 and then reorganized to be the second version. Bonsignore *et al.*⁸ reported a good internal and external validity of the second version of the WHO-5 in detecting depression in the elderly population. The WHO-5 has been translated into many languages and used worldwide to detect depression in various health conditions such as elderly patients,⁹ diabetic patients,¹⁰ and patients in primary care.^{11,12}

This study aimed to report the translation and psychometric properties of the Thai version of the WHO-5 in screening for depression in a primary care setting.

METHODS

The translation and adaptation of the Thai version of the WHO-5

After obtaining permission from the copyright holder, the original English version of the WHO-5 was translated into Thai for a cross-cultural adaptation of the self-report measure, including forward-translation, synthesis of the translation, backward-translation, cross-cultural adaptation and pilot testing. The English version of the WHO-5 was translated into Thai by two independent bilingual translators, and then a synthesis of both versions was conducted to produce a common Thai version. The first draft was then translated back from Thai to English by another bilingual school teacher. The authors finally discussed and consolidated to reach the most suitable version and conducted pilot testing. Ten patients were invited to complete and comment on this pre-final version. Final modification and adjustment were thereafter made accordingly.

Subjects

The patients were recruited from the outpatient clinic of the Department of Family Medicine, Ramathibodi Hospital, Bangkok. This clinic functions as the primary care clinic of the hospital. It has 34 primary care physicians working with around 400 outpatients each day and its top ten diseases are similar to those of other primary care settings in Bangkok¹³ for example, hypertension, hyperlipidemia, diabetes mellitus, upper respiratory infection. The first of every five consecutive patients was invited to complete the WHO-5-T while they were waiting to see their doctors. Informed consent was obtained after the objectives of the study had been explained and the patients agreed to participate. Data were collected until a total of 300 cases had been reached.

After completing the questionnaire, patients were then assessed by a research assistant who was unaware of the patients' WHO-5-T scores. A trained research assistant in the study was a clinical psychologist who was trained in the use of the Mini International Neuropsychiatric Interview (MINI), Thai version¹⁴ and the Hamilton Rating Scale for Depression (HAM-D), Thai version,¹⁵ which were employed as gold-standard measures for the current diagnosis of major depressive disorder and symptom severity, respectively.

In total, 274 patients participated in the whole process and completed all data. The study was designed to collect a total of 300 cases, but 26 subjects were excluded due to incomplete data. The mean age of the subjects was 44.6 [standard deviation (SD) = 14.7]. There were 202 females (73.7%) and 72 males (26.3%). Most of them (60.1%) were married and one-third had graduated from secondary school.

This study was approved by the Ethics Committee of the Faculty of Medicine, Ramathibodi Hospital, Bangkok.

Instruments

The WHO-5-T is a short and quick self-report tool for detecting depression in the Thai language. It consists of the same five Well-Being Index items as the original English version. Each item of the WHO-5-T is scored on a scale of 5 (= all of the time) back to 0 (= at no time). A sum of the scores added up; it ranges from 0 to 25. A higher score means a healthier condition.

The MINI is a standardized clinical diagnostic interview schedule for the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition Axis-I disorders.¹⁶ It can be reliably administered by lay interviewers who pass the appropriate training. The depression modules of the schedule were used in the study. A study of the MINI Thai version showed that the Kappa, sensitivity and positive predictive value (PPV) on the diagnosis of a current major depressive episode were high (>0.75, >0.81 and >0.81, respectively). The specificity, the negative predictive value (NPV), and efficiency were also high (all >0.81).¹⁴

The HAM-D is a well accepted research tool for measuring the severity of depression and response to treatment.¹⁷ The Thai version of the HAM-D has good internal consistency (alpha coefficients = 0.74) and its concurrent validity as compared with the Global Assessment Scale is also satisfactory (Spearman's correlation coefficient = -0.82).¹⁵

Statistical analyses

Data were analyzed using the Statistical Package for the Social Sciences 10. The internal consistency of the WHO-5 was measured using Cronbach's alpha coefficient and factor analysis using principle component extraction and varimax rotation. To determine the best cut-off score the indices of sensitivity, specificity, PPV, and NPV were used along with the receiver operating characteristic (ROC) curve. Furthermore, positive-likelihood ratios (sensitivity/(1 - specificity)) and negative-likelihood ratios ((1 - sensitivity)/specificity) of the test were assessed. Pearson's correlation coefficient was used to assess the convergent validity between the HAM-D and the WHO-5. Significance level was set at $P < 0.05$ (two-tailed test).

RESULTS

The mean total WHO-5-T score of our sample was 14.32 (SD = 5.26) with a range of 2 to 25, the median score was 15.0, and there was a skewness of -0.34 (SE = 0.15).

Mean scores for each WHO-5-T item in this study are shown in Table 1. The Cronbach's alpha coefficient for the total scale was equal to 0.87. All items, if deleted, would consistently decrease the total scale alpha. A factor analysis showed only one factor, which explained 66.8% of the variance, by considering eigenvalues >1.0 (Table 1). These results indicated an adequate internal consistency reliability of the WHO-5-T.

The MINI was used as a gold standard in determining the concurrent validity of the test. According to the MINI, 19 patients (6.9%) met the diagnosis of major depression. The mean WHO-5-T score for these patients ($X = 7.58 \pm 4.21$) was significantly lower than patients without major depression ($X = 14.83 \pm 4.98$) ($t = 6.177$, d.f. = 272, $P < 0.0001$).

It is illustrated from the ROC curve that the WHO-5-T performed well in identifying patients with depressive disorders (Fig. 1). The area under the curve in this study was 0.86 (SD = 0.03, 95% CI 0.81 to 0.89). Table 2 demonstrates the sensitivity, specificity, PPV, NPV, and likelihood ratio for different WHO-5-T thresholds in diagnosing major depression. At a cut-off score of WHO-5-T <12, a sensitivity was 0.89 and specificity was 0.71. This threshold had a PPV of 0.19 and an NPV of 0.99. The positive likelihood ratio was 3.13 at this cut-off point.

Using Pearson's correlation coefficient to determine the convergent validity, the total score of the WHO-5-T and the HAM-D, which are scored in opposite directions, were negatively correlated

Table 1. Thai version of the WHO-5-T item level values, item-total correlations and factor matrix

WHO-5-T item	Mean	SD	Corrected item-total correlation	Alpha if item deleted	Factor loadings
1. I have felt cheerful and in good spirits	2.97	1.25	0.73	0.84	0.83
2. I have felt calm and relaxed	2.95	1.18	0.68	0.85	0.80
3. I have felt active and vigorous	2.72	1.31	0.76	0.83	0.86
4. I woke up feeling fresh and rested	2.81	1.35	0.66	0.86	0.78
5. My daily life has been filled with things that interested me	2.88	1.34	0.70	0.85	0.81

SD, standard deviation; WHO-5-T, World Health Organization-Five Well-Being Index (Thai version).

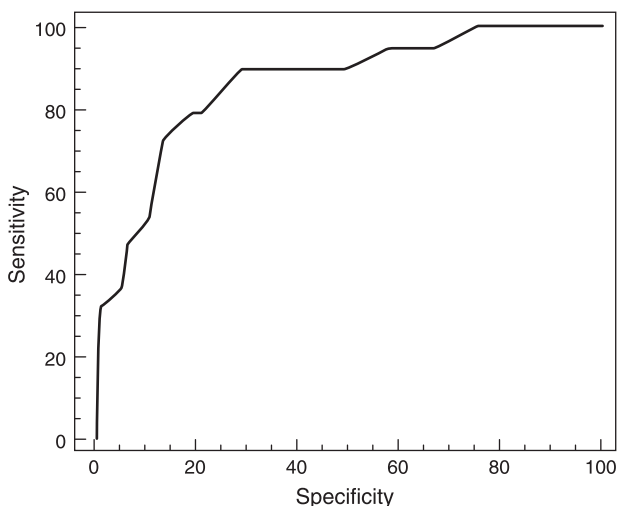


Figure 1. The receiver operating characteristic curve of the Thai version of the World Health Organization-Five Well-Being Index (WHO-5-T).

(-0.54 , $P < 0.001$). This indicated a negative association between the two instruments of moderate strength. We divided the HAM-D score into four groups according to the severity of depression.¹⁸ The ANOVA test revealed a significant difference between the mean WHO-5-T score for the HAM-D severity group ($F = 26.75$, $d.f. = 3$, $P < 0.0001$), especially the difference between the 'more than' 'less than' depression and the 'no' or 'mild depressive' groups (Table 3).

DISCUSSION

This study has shown a sufficient reliability and validity of the Thai version of the WHO-5 in detecting major depressive disorder in a primary care setting. The results of our study suggested that the WHO-5-T was well suited for use in the Thai primary-care population. It is shorter and easier than other measures of depression available in Thailand.

Table 2. The performance of various WHO-5-T cut-off scores in detecting major depression

Cut-off score	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Likelihood ratio – positive	Likelihood ratio – negative
7/8	0.74	0.86	0.29	0.98	5.37	0.3
8/9	0.79	0.81	0.24	0.98	4.19	0.26
9/10	0.79	0.78	0.21	0.98	3.66	0.27
10/11	0.84	0.75	0.20	0.99	3.36	0.21
11/12	0.89	0.71	0.19	0.99	3.13	0.15
12/13	0.89	0.65	0.16	0.99	2.56	0.16
13/14	0.89	0.60	0.14	0.99	2.22	0.18
14/15	0.89	0.51	0.12	0.99	1.81	0.21
15/16	0.94	0.42	0.11	0.99	1.64	0.12
16/17	0.9	0.33	0.10	0.99	1.41	0.16

WHO-5-T, World Health Organization-Five Well-Being Index (Thai version).

Table 3. Relationship between WHO-5-T mean scores and depression severity according to the HAM-D

HAM-D score	<i>n</i>	Mean WHO-5-T	SD	95% CI
No depression (0–7)	207	15.66 [†]	0.32	15.03–16.29
Mild (8–12)	35	11.91 [†]	0.88	10.12–13.71
Less than major depression (13–17)	19	8.37 [‡]	0.85	6.59–10.14
Major depression (18 or greater)	13	8.23 [‡]	1.43	5.11–11.35

[†]Different from other three groups. [‡]Different from 'no depression' and 'mild depression'.

The level of significance of ANOVA is <0.0001 , and the results of post hoc comparison are indicated as superscript symbols.

HAM-D, Hamilton Rating Scale for Depression; SD, standard deviation; CI, confidence interval;

WHO-5-T, World Health Organization-Five Well-Being Index (Thai version).

The internal consistency of the WHO-5-T (alpha coefficient = 0.87) was within the acceptable range.¹⁹ It was in the same range as in previous studies from Germany and Japan.^{9,10,20} From ROC analysis, it indicated that the WHO-5-T had a sufficient discriminatory validity as a screening tool for detecting major depression. At a standard cut-off point of <12, the WHO-5-T had the best sensitivity (89%) and specificity (71%). This cut-off score is one point lower than the original version, which is <13 (see <http://www.who-5.org>). The reason might be due to the limitation of our study population as mentioned below. We also determined the convergent validity of the WHO-5-T in relation to the HAM-D. The satisfactory correlation between these two scales confirmed the validity of the WHO-5-T. Even though the WHO-5-T cannot totally differentiate 'less than major depression' from 'major depression', because of a small number of cases in each group, both of these conditions need to be treated as a study in a primary care setting.²¹

When compared to the Japanese version of the WHO-Five Well-Being Index (WHO-5-J), WHO-5-T has a lower cut-off score. WHO-5-J can detect depression in diabetic patients at a cut-off point of <13 with very good sensitivity (100%) and specificity (78%).¹⁰ This difference might come from not only the living styles and environment between the two countries but also the subjects and settings of the studies. In another study on the WHO-5-J,⁹ a standard cut-off point of ≤ 12 has been used to detect depression with suicidal ideation in the elderly community. The assessment of perceived social support, which was significantly associated with the elderly suicidal ideation, needs to be combined to provide better sensitivity (87%) and specificity (75%). There is a study from Germany using the WHO-5 to detect depression in primary care.¹¹ Its higher cut-off point (≤ 13) was due to superior sensitivity (94%) and a superior false-negative rate (6%), which can be considered the most important aspects for screening purposes.

This study had some limitations. First, the study was conducted in a university hospital in the capital city of Thailand. Therefore, the respondents might not reflect the actual primary care patients seen in a rural and remote area, where lifestyle might be more peaceful. A peaceful lifestyle might influence the score on the second and the fifth items of the WHO-5, which ask about 'feeling calm and relaxed' and 'interesting things in daily life'. However, this is the first study to evaluate the reliability and validity

of the Thai version of the WHO-5, and it might require future studies from our countryside. Second, our study was a sample estimation. Hence, the use of this screening tool should be aware of the population estimation. We calculated the 95% confidence intervals of sensitivity and specificity, which were 75.2–100% and 65.4–76.6%, respectively. Even though the 95% confidence interval of sensitivity is quite large, it is in an acceptable range. Third, we did not examine discriminatory validity to detect other psychiatric disorders rather than major depression, as we know that other psychiatric disorders are also able to influence the respondents' subjective well-being. However, poor well-being at last can lead to depression, which we have concentrated on. Fourth, there was a possibility that our respondents under- or over-reported their inner feelings. This can happen in every study about subjective senses. The test-retest reliability of the WHO-5-T was not assessed. Generally, this type of reliability is used to measure the stability of a scale over time and is usually assessed over a short period of time. This was our limitation as most of our participants had a scheduled appointment for more than a month later.

CONCLUSION

This study has shown that the Thai version of WHO-5 is a valid and useful instrument in screening for depression. It has a satisfactory sensitivity and specificity with a cut-off point of <12 for detecting depression in the Thai primary-care population.

ACKNOWLEDGEMENTS

This study was supported by Grant 47018 of the Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

REFERENCES

- Ustun TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJ. Global burden of depressive disorders in the year 2000. *Br. J. Psychiatry* 2004; **184**: 386–392.
- Weiss MG, Raguram R, Channabasavanna SM. Cultural dimensions of psychiatric diagnosis. A comparison of DSM-III-R and illness explanatory models in south India. *Br. J. Psychiatry* 1995; **166**: 353–359.
- Patel V, Araya R, Bolton P. Treating depression in the developing world. *Trop. Med. Int. Health* 2004; **9**: 539–541.
- Lotrakul M, Saipanish R. Psychiatric services in primary care settings: A survey of general practitioners in Thailand. *BMC Fam. Pract.* 2006; **7**: 48.

- ⁵ Kasantikul D, Karnjanathanalers N, Limsuwan N *et al.* [Health-related self-report (HRSR) scale: The diagnostic screening test for depression in Thai population.] *J. Med. Assoc. Thai.* 1997; 80: 647–657 (In Thai.)
- ⁶ Lotrakul M, Sukanich P. [Development of the Thai depression inventory.] *J. Med. Assoc. Thai.* 1999; 82: 1200–1207 (In Thai.)
- ⁷ Kuptniratsaikul V, Chulakadabba S, Ratanavijitrasil S. [An instrument for assessment of depression among spinal cord injury patients: Comparison between the CES-D and TDI.] *J. Med. Assoc. Thai.* 2002; 85: 978–983 (In Thai.)
- ⁸ Bonsignore M, Barkow K, Jessen F, Heun R. Validity of the five item WHO Well-Being Index (WHO-5) in an elderly population. *Eur. Arch. Psychiatry Clin. Neurosci.* 2001; 251 (Suppl. 2): II/27–II/31.
- ⁹ Awata S, Bech P, Koizumi Y *et al.* Validity and utility of the Japanese version of the WHO-Five Well-Being Index in the context of detecting suicidal ideation in elderly community residents. *Int. Psychogeriatr.* 2007; 19: 77–88.
- ¹⁰ Awata S, Bech P, Yoshida S *et al.* Reliability and validity of the Japanese version of the World Health Organization-Five Well-Being Index in the context of detecting depression in diabetic patients. *Psychiatry Clin. Neurosci.* 2007; 61: 112–119.
- ¹¹ Henkel V, Mergl R, Kohnon R, Allgaier A-K, Möller H-J, Hegerl U. Use of brief depression screening tools in primary care: Consideration of heterogeneity in performance in different patient groups. *Gen. Hosp. Psychiatry* 2004; 26: 190–198.
- ¹² Henkel V, Mergl R, Kohnon R, Maier W, Möller H-J, Hegerl U. Identifying depression in primary care: A comparison of different methods in a prospective cohort study. *BMJ* 2003; 326: 200–201.
- ¹³ Medical Service Department. Annual report [Online]. 2006 (78 screens). [Cited 11 Aug 2008.] Available from URL: <http://www.dmsbma.go.th/downloadyearreports.html> (in Thai).
- ¹⁴ Kittirattanapaiboon P, Khamwongpin M. [The validity of the Mini International Neuropsychiatric Interview (M.I.N.I.)-Thai version 126.] *J. Ment. Health Thai.* 2005; 13: 126–136 (In Thai.)
- ¹⁵ Lotrakul M, Sukanit P, Sukying C. [The validity and reliability of the Hamilton Rating scale for depression, Thai version.] *J. Psychiatr. Assoc. Thai* 1996; 41: 235–246 (In Thai.)
- ¹⁶ Sheehan DV, Lecrubier Y, Sheehan KH *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 1998; 59 (Suppl. 20): 22–33, quiz 34–57.
- ¹⁷ Hamilton M. Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol.* 1967; 6: 278–296.
- ¹⁸ Bech P. Acute therapy of depression. *J. Clin. Psychiatry* 1993; 54 (Suppl.): 18–27; discussion 28.
- ¹⁹ Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 2nd edn. Oxford University Press, Oxford, 1995.
- ²⁰ Löwe B, Spitzer RL, Gräfe K *et al.* Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J. Affect. Disord.* 2004; 78: 131–140.
- ²¹ Paykel ES. Use of the Hamilton Depression Scale in general practice. In: Bech P, Copper A (eds). *The Hamilton Scale*. Springer, Berlin, 1990; 40–47.