Psychometric properties of the Japanese version of the Clinical Outcomes in Routine Evaluation–Outcome Measure

Masayo Uji[a],*, Ayuko Sakamoto[b], Keiichiro Adachi[c], Toshinori Kitamura[d]

[a]Department of Bioethics, Kumamoto University Graduate School of Medical Sciences, 1-1-1, Honjo, Kumamoto 860-8556, Japan
[b]Asoyamanami Hospital, 115-1, Miyaji, Ichinomiya, Aso, Kumamoto, 869-2612, Japan
[c]Department of Clinical Psychology, Kyushu Lutheran College, 3-12-16, Kurokami, Kumamoto, 860-8520, Japan
[d]Kitamura Mental Health Institute, 8-5-13-101, Akasaka, Tokyo, 107-0052, Japan

Abstract

This study examined the reliability and validity of the Japanese version of the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM), a standardized, brief, but comprehensive outcome measurement. The target population consisted of 1684 Japanese company employees, hospital staff, and university students. A confirmatory factor analysis proved that our data fit the factor structure of the original CORE-OM. We also examined its internal consistency, test-retest reliability, convergent validity, and sensitivity in discriminating between clinical and nonclinical samples. After demonstrating these results, we discuss how the Japanese version of the CORE-OM can be used both in clinical and research settings.
© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) was developed as a standardized brief outcome measure by Evans et al [1] for use in both routine clinical practice and psychotherapy research. The underlying rationale for its development was detailed by Barkham et al [2]. They pointed out that despite the existence of a vast number of psychologic and psychiatric measures, most of these have not been generally accepted and have no psychometric data. In addition, a discrepancy has existed between clinicians' and researchers' interests, leading to user-unfriendly measures that are too abstract or unreadable or that contain too many items. Furthermore, because they have been generally too long or too specific and unnecessarily complicated [1], clients have found them to be tiresome. These factors resulted in clinicians avoiding the use of the measurements in their practice and being consequently less involved in data collection [2]. Without widely accepted standardized measures, researchers have had difficulty in comparing the effectiveness of treatments used in different studies. For clinicians, there have been no appropriate indicators to assist in the decision to commence treatment or assess any changes

after treatment. Under the call for evidence-based practice, Barkham et al [2] emphasized the need for a standardized measure that covers core symptoms and life functioning. They also emphasized the importance of collaboration between clinicians and researchers in the development of a core outcome measure. With all of this in mind, the CORE-OM was developed by a group consisting of a variety of mental health professionals [1]. In this study, we translated the CORE-OM into Japanese and sought to verify its reliability and validity.

The CORE-OM is composed of 4 subcategories: subjective well-being (SWB), symptomatic problems, functioning, and risk. The symptomatic problem subcategory consists of 4 components: anxiety, depression, physical symptoms, and traumatic symptoms. In the same way, the functioning subcategory comprises 3 components: social functioning, general functioning, and close relationships. The risk subcategory includes both risk to self and risk to others. Examples include "I have felt OK about by myself" (SWB item), "I have felt tense, anxious, and nervous" (anxiety item), "I have felt totally lacking in energy and enthusiasm" (depression item), "I have been troubled by aches, pains, or other physical problems" (physical symptom item), "I have been disturbed by unwanted thoughts and feelings" (trauma item), "I have felt criticized by other people" (social functioning item), "I have been able to do most things I needed to" (general functioning item), "I have felt warmth

* Corresponding author. Tel.: +81 0 96 373 5523; fax: +81 0 96 373 5523.
E-mail address: ujimasayo@hotmail.co.jp (M. Uji).

Fig. 1. Confirmatory factor analysis of SWB subscale.
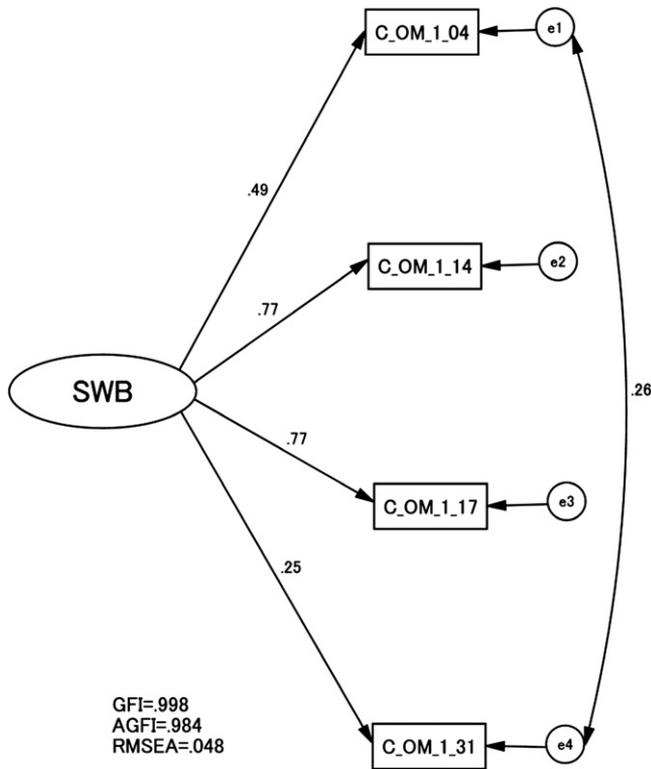
The Japanese version of the CORE-OM is expected to provide a standardized brief measure of the entire range of concepts that comprise a person's general health. In this study, we examined the reliability and validity of the Japanese version of the CORE-OM. Confirmatory factor analysis (CFA) was used to verify the validity of each CORE-OM subscale. Furthermore, we investigated the influence of sociodemographic variables on CORE-OM complete and subscale scores.

2. Methods

2.1. Procedures

This study consists of self-reported questionnaires administered on 2 occasions. The interval between the first and second occasions was 4 weeks. Participation was voluntary, and anonymity was assured.

2.2. Respondents

The respondents were medical staff from 7 private hospitals in Kumamoto, Fukuoka, students from 3 universities in Kumamoto, and employees from 8 companies in Kumamoto and Kagoshima. Questionnaires were distributed to 1684

and affection for someone" (close relationship item), "I have thought it would be better if I were dead" (risk to self item), and "I have threatened or intimidated another person" (risk to others item). For assessment, CORE-OM requires respondents to answer 34 items using 5-point scales. The wide range of concepts that are targeted for evaluation helps clinicians and researchers evaluate the degree of an individual's health and their acceptance of and satisfaction with their life.

The reliability and validity of the CORE-OM are confirmed by Evans et al [3]. They demonstrated favorable Cronbach α coefficients, in all subcategories, the lowest being a Cronbach α coefficient of .75 for the SWB subcategory. Despite the subscale being composed of only 4 items, it maintained a relatively high internal consistency [3] and also showed that each subscale possessed good test-retest stability. The lowest correlation coefficient between 2 occasions was .64 for the risk subcategory, a result the authors attributed to the situational and reactive nature of its items.

Concerning convergent validity, they showed appropriate correlations between each subcategory score and conceptually similar measures. They further showed its accuracy in discriminating between clinical and nonclinical populations. Although they found a sex difference on the SWB and risk subcategories, it was smaller than that between clinical and nonclinical groups. All the correlations between subscale scores were positively correlated, although the risk subscale score showed relatively lower correlations with other subscale scores.
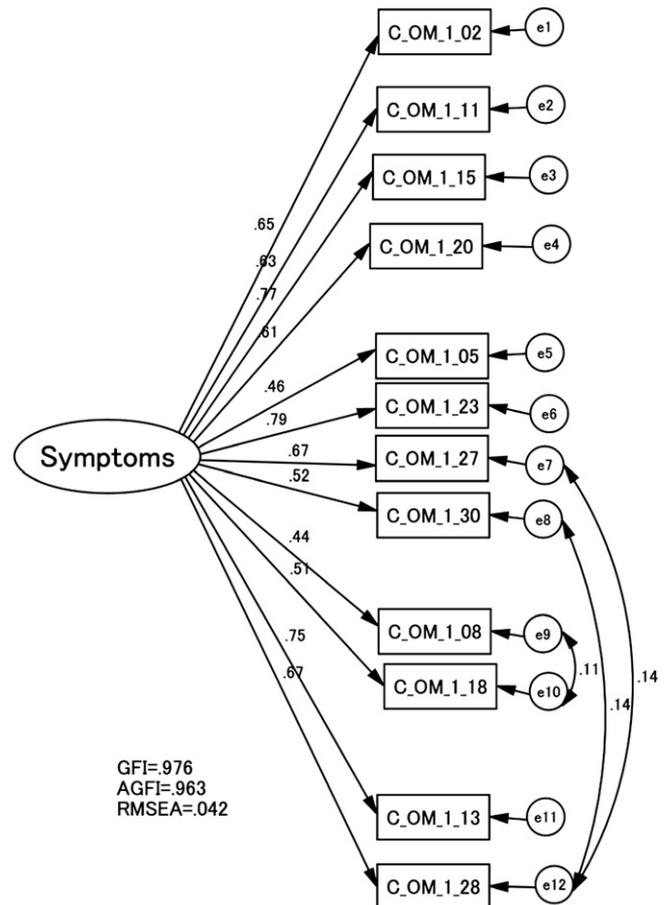

Fig. 2. Confirmatory factor analysis of symptomatic problems subscale.
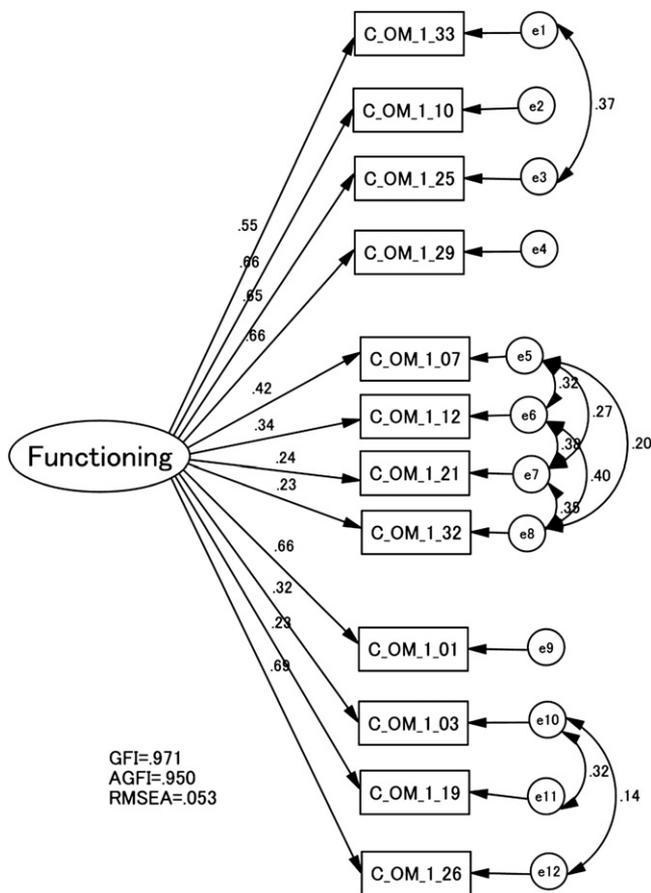
Fig. 3. Confirmatory factor analysis of functioning subscale.

2.4.1. Scales of Psychological Well-Being

The short version of the Scales of Psychological Well-Being (SPWB) [4] was used to verify the convergent validity of the SWB and the functioning subscales of the CORE-OM. We chose 3 items from each of the 6 SPWB subcategories: autonomy, environment mastery, personal growth, positive relation with others, purpose in life, and self-acceptance. The short version of the SPWB consists of 18 items using a 5-point scale and therefore has a total score ranging from 18 to 90. Contrary to the SWB subscale of the CORE-OM, a higher SPWB score indicates a higher level of psychologic well-being.

2.4.2. Hospital and Anxiety Depression Scale

The Hospital and Anxiety Depression Scale (HADS) was developed by Zigmond and Snaith [5] and is a self-assessment inventory for depression and anxiety. Physical symptoms that are usually concomitant with other depressive symptoms are excluded from the HADS to avoid the "noise" from somatic disorders [6]. This scale is composed of 7 items categorized in the anxiety subscale (HADS-A) and 7 items categorized in the depression subscale (HADS-D). Each item score ranges from 1 to 4. Therefore, the combined score of the HADS-A and HADS-D ranges from 7 to 28. The validity and reliability of the Japanese version of the HADS were confirmed by Kugaya et al [7]. The CORE-OM anxiety and depression components were expected to evaluate content

people. A total of 1357 people participated at least on either of the 2 occasions. Of these, the number of male and female respondents was 433 and 881, respectively, and the others did not answer to the question concerning their sex. The mean (SD) age was 35.0 (13.9). The respondents of the 1020 who completed every item of CORE-OM on the first occasion were used for CFA of the Japanese version of CORE-OM.

2.3. Measurements

2.3.1. The CORE-OM

The CORE-OM was included in the questionnaires on both occasions. As explained in the "Introduction," it consists of 34 items using 5-point scales. Therefore, its total score ranges from 34 to 170 with a higher score indicating more serious problems. The original English version of the CORE-OM was translated by one of the authors of this article (MU). To verify the accuracy of the Japanese translation, it was translated back into English by an individual unfamiliar with the original document.

2.4. Measurements of convergent validity of the CORE-OM

To confirm the convergent validity of the Japanese version of the CORE-OM, we used 5 measurements as described below.
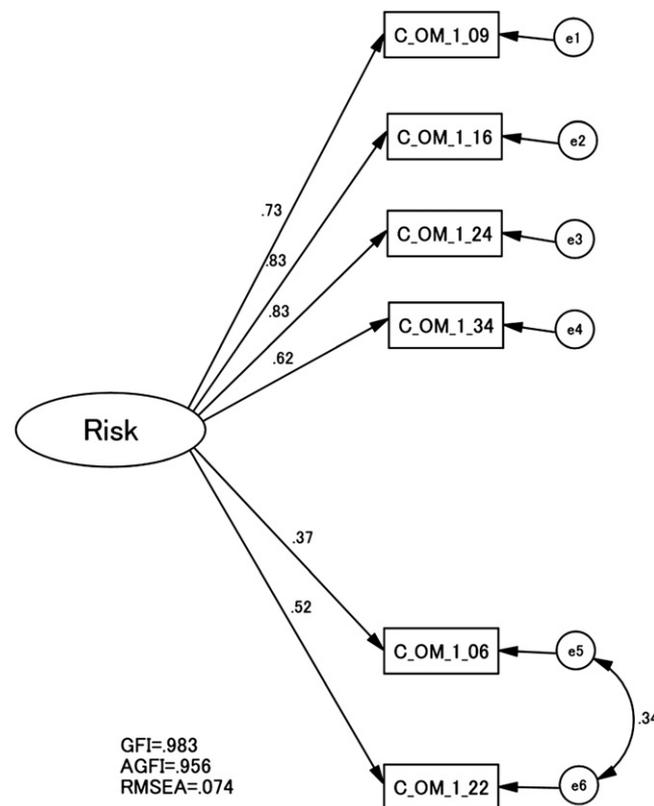

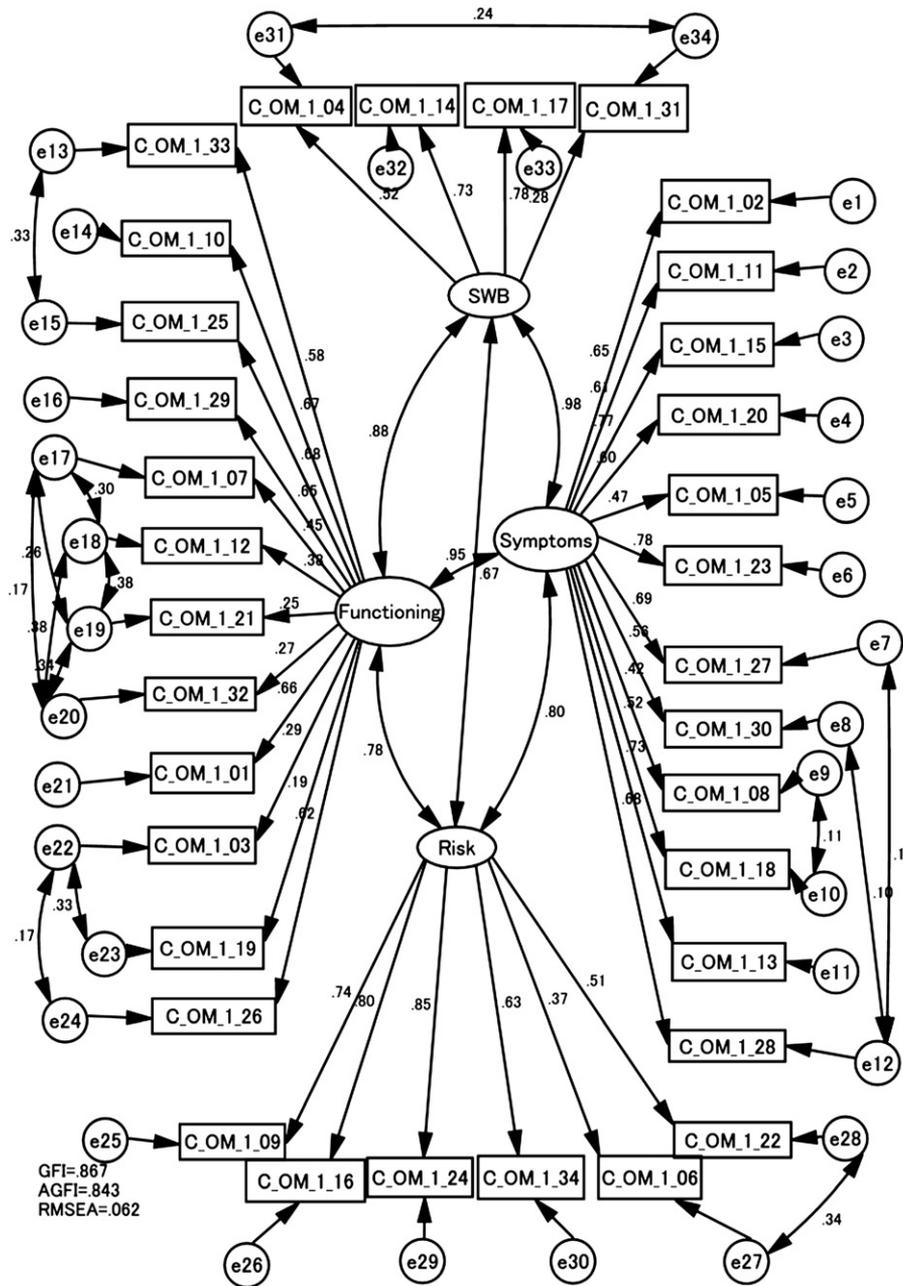Fig. 4. Confirmatory factor analysis of risk subscale.

Fig. 5. Confirmatory factor analysis of CORE-OM.

conceptually similar to the depression and anxiety subscales of the HADS.

2.4.3. Self-Rating Depression Scale

The Self-Rating Depression Scale (SDS) [8] is a self-report measure of depressive symptoms that consists of items on a 4-point scale from "never" (scored "1") to "almost always" (scored "4"). Using a Japanese university student population, Kitamura et al [9] have reported a 3-factor structure for the scale. They identified the 3 factors as affective, cognitive, and somatic. Three SDS somatic category items were selected because they were regarded as being conceptually similar to the CORE-OM physical subscale items.

2.4.4. Impact of Event Scale–Revised

The Impact of Event Scale–Revised (IES-R) [10] was developed as a revision of the original Impact of Event Scale

Table 1
Internal consistency of each subscale

Subcategory	n	Coefficient α (95% CI)
SWB	1301	.68 (.65-.71)
Symptomatic problems	1277	.89 (.88-.89)
Functioning	1280	.81 (.79-.83)
Risk to self and others	1299	.83 (.81-.84)
All items	1248	.94 (.93-.94)

Table 2
Item-subscale score correlations

No.	Item	Item-subcategory score correlation	Subcategory
4	...OK about myself	.69	SWB
14	...felt like crying	.80	SWB
17	...overwhelmed by my problems	.78	SWB
31	...optimistic about my future	.57	SWB
2	...tense, nervous, and anxious	.69	Problems
11	Tension and anxiety prevented me from...	.68	Problems
15	...panic or terror...	.76	Problems
20	...impossible to put to one side.	.64	Problems
5	...lacking in energy and enthusiasm.	.53	Problems
23	...despairing and hopeless.	.78	Problems
27	...unhappy.	.70	Problems
30	I am to blame for...	.60	Problems
8	...aches, pains, and other physical problems.	.55	Problems
18	...difficulty getting to sleep...	.59	Problems
13	...unwanted thoughts and feelings.	.76	Problems
28	Unwanted images and memories...	.72	Problems
33	...humiliated or shamed...	.60	Functioning
10	Talking to people...too much for me.	.63	Functioning
25	...criticized by other people.	.67	Functioning
29	...irritable when with other people.	.64	Functioning
7	...able to cope when things go wrong.	.57	Functioning
12	...happy with the things I have done.	.54	Functioning
21	...able to do most things...	.48	Functioning
32	...achieved the things...	.45	Functioning
1	...terribly alone and isolated.	.66	Functioning
3	...someone to turn to for support...	.49	Functioning
19	...warmth and affection for someone.	.40	Functioning
26	...no friends.	.67	Functioning
9	...hurting myself.	.78	Risk
16	...plans to end my life.	.81	Risk
24	...be better if I were dead.	.82	Risk
34	...hurt myself or taken dangerous risks.	.74	Risk
6	...physically violent to others.	.57	Risk
22	...threatened or intimidated another person.	.66	Risk

Item sentences are abbreviated.

(IES) [11]. The IES consisted of 7 intrusion and 8 avoidance items. The IES-R comprises 22 items after the addition of 7 items including 6 that tapped hyperarousal and 1 relating to flashback-like reexperience. Each item is rated from 1 to 5. Therefore, the total score of the IES-R ranges from 22 to 110. A higher score indicates a greater intensity of traumatic symptoms. Asukai et al [12,13] translated the IES-R into Japanese and verified its reliability and validity. Respondents were instructed to recall the most stressful event they had ever experienced and score it according to each item of the IES-R. This scale was chosen as a referential measure to test the convergent validity of the trauma components of the CORE-OM.

2.4.5. Buss-Perry Aggression Questionnaire

As a revised version of the Hostility Scale [15], the Buss-Perry Aggression Questionnaire (BAQ) [14] was developed for the purpose of evaluating different types of aggression. By conducting exploratory factor analysis, they extracted 4 subscales: physical aggression, verbal aggression, anger, and hostility. Ando et al [16] developed the Japanese version of the BAQ and demonstrated its reliability and validity. Factor analysis confirmed its 4-factor structure, corresponding to the 4 subscales extracted by [14].

In this study, we chose 6 physical aggression items and 5 verbal aggression items. The sum of these items was expected to correlate with the "risk to others" score of the CORE-OM. This is a 5-point scale; therefore, the total scores of the physical and verbal aggression subscales range from 6 to 30 and 5 to 25, respectively.

2.9. Statistics

To examine whether our data fit the factor structure of the original CORE-OM, we first conducted an individual CFA for each subscale: SWB, symptomatic problems, functioning, and risk. The subscales were then combined to obtain a comprehensive measure of the CORE-OM structure's validity. The 1020 respondents who completed every item of the CORE-OM of the first occasion were the subjects of the CFA. SPSS version 18.0 (SPSS) and Amos version 18.0 (SPSS, an IBM Company, Tokyo, Japan) were used for the CFA.

Each subscale and the total score of the CORE-OM were examined for internal consistency and test-retest reliability, with Cronbach α coefficient calculated as the index of internal consistency. Item-subcategory score correlations were examined using Pearson correlation. Furthermore, correlations between subcategory scores, as well as between each subcategory and the total scores, were examined.

We used t test in comparing clinical and nonclinical samples in terms of their subtotal and total CORE-OM scores. Sex differences were examined in the same way. Pearson correlation was adopted as the index for investigating the influence of age on the subtotal and total CORE-OM scores.

To assess convergent validity, we calculated the correlation between each CORE-OM subscale and other referential scales that could be regarded as assessing similar concepts.

Table 3
Correlations between subcategory scores and the CORE-OM total score

	SWB	Symptomatic problems	Functioning	Risk	CORE-OM total score
SWB					
Symptomatic problems	.76*				
Functioning	.70*	.76*			
Risk	.49*	.68*	.60*		
CORE-OM total score	.83*	.95*	.90*	.77*	

* $P < .001$.

Table 4
Correlations between subscale scores and complete scores on 2 occasions (test-retest reliability)

Subcategory	SWB	Symptomatic problems			Functioning			Risk	CORE-OM
	.67*	.82*			.72*			.66*	.85*
Subcomponent	Anxious	Depressed	Somatic	Trauma	Close	General	Social	Self	Others
	.68*	.69*	.60*	.70*	.65*	.61*	.65*	.68*	.50*

* $P < .001$.

3. Results

3.1. Confirmatory factor analysis of each subscale

Figs. 1 to 4 demonstrate the results of the CFA. The fitness of our data to the original model is shown using the following indices: Goodness of Fit Index (GFI) (range, .971-.998), Adjusted Goodness of Fit Index (AGFI) (range, .950-.984), and Root Mean Square Error of Approximation (RMSEA) (range, .042-.074), all of which indicate favorable or acceptable fitness [17].

3.2. Confirmatory factor analysis of the CORE-OM

The CFA was conducted by including the 4 previously mentioned subcategories as latent variables (Fig. 5). The fitness to the model was as follows: GFI, .867; AGFI, .843; and RMSEA, .062. With the exception of GFI, these indices showed acceptable fitness of the data to the model [17]. The path coefficient from functioning to item 19 was relatively low (standardized coefficient was .19) in comparison with that of the other 33 items.

Internal consistency results are shown in Table 1. All subscales except the SWB subscale displayed favorable reliability with Cronbach α coefficients of >.81 and <.94. The SWB scale showed a relatively lower Cronbach α coefficient of .68.

Item-subcategory score correlations ranged from .40 to .82 (Table 2). When compared with the items in other categories, functioning items showed lower correlations with the subcategory scores.

Correlations between subcategory scores and the total score were examined (Table 3). The combination of any 2 scores showed significant correlation. The risk subscale showed a relatively lower correlation with other subscales, similar to results obtained by Evans et al [3].

The correlations between subtotal and total scores on the 2 survey occasions were calculated to examine test-retest reliability (Table 4). The correlation coefficients ranged from .50 to .85. In accordance with Evans et al [3], among the 4 subscales, the risk subscale (the sum of risk to self and others) showed the lowest test-retest reliability with a correlation coefficient of .66. This result seems to have been strongly influenced by the low test-retest reliability of the risk to others category, with its correlation coefficient of .50. The CORE-OM total score showed a favorable test-retest reliability with a correlation coefficient of .85.

3.3. Difference between sexes

Female respondents scored lower on risk to others and close relationship subscales, which indicates fewer severe problems for females in these domains (Table 5). On the

Table 5
Difference in mean scores (SD) between sexes and between clinical and nonclinical populations

		Sex				Clinical and Nonclinical			
		Male (n = 433)	Female (n = 881)	t	P	Clinical (n = 27)	Nonclinical (n = 1276)	t	P
SWB	Total	10.7 (3.1)	11.4 (3.2)	−3.68***	.000	14.5 (4.3)	11.1 (3.2)	5.44***	.000
Symptomatic problems	Anxiety	9.4 (3.4)	9.6 (3.3)	−0.71	.479	14.0 (3.7)	9.4 (3.3)	6.99***	.000
	Depression	9.2 (3.1)	9.2 (3.0)	0.06	.950	12.8 (3.9)	9.1 (3.0)	6.17***	.000
	Somatic	4.3 (4.3)	4.5 (4.5)	−1.84	.066	7.3 (2.2)	4.4 (1.9)	7.95***	.000
	Trauma	4.4 (1.9)	4.5 (1.9)	−1.24	.217	6.9 (2.5)	4.4 (1.9)	6.74***	.000
	Total	27.2 (8.9)	27.8 (8.7)	−1.01	.315	41.1 (11.1)	27.3 (8.5)	8.15***	.000
Functioning	Close	8.8 (3.0)	8.3 (2.6)	3.24**	.001	10.7 (3.6)	8.4 (2.7)	4.12***	.000
	General	12.1 (2.8)	12.3 (2.6)	−1.44	.149	14.4 (3.7)	12.2 (2.6)	4.25***	.000
	Social	8.2 (3.1)	8.4 (3.0)	−1.35	.177	11.5 (3.8)	8.3 (3.0)	5.39***	.000
	Total	29.1 (6.8)	29.1 (6.4)	0.14	.893	36.6 (9.1)	28.9 (6.4)	5.95***	.000
Risk	Risk to self	6.5 (3.0)	6.4 (3.0)	0.44	.657	10.8 (4.7)	6.3 (2.9)	7.41***	.000
	Risk to others	3.2 (1.4)	3.0 (1.3)	2.41*	.016	3.1 (1.5)	3.1 (1.3)	0.18	.857
	Total	9.6 (3.7)	9.4 (3.8)	1.08	.280	13.9 (5.4)	9.4 (3.7)	6.10***	.000
CORE-OM total		76.4 (20.1)	77.6 (19.7)	−0.99	.325	105.8 (27.9)	76.6 (19.3)	7.44***	.000

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Table 6
Correlation between age and total and subtotal CORE-OM scores

Subcategory	SWB	Symptomatic problems					Functioning				Risk			CORE-OM total
Subcategory components	SWB	Anxious	Depressed	Physical	Trauma	Total	Close	General	Social	Total	Self	Others	Total	
Pearson correlation	−.21	−.18	−.15	.09	−.16	−.14	−.01	−.15	−.02	−.08	−.03	.04	−.01	−.12
P value	.000	.000	.000	.002	.000	.000	.633	.000	.528	.008	.343	.124	.822	.000

other hand, male respondents scored lower on the SWB subscale, which indicates that males experience higher levels of SWB. Other subscale scores and the total score were free from sex differences.

3.4. Differences between clinical and nonclinical samples

Of the above 1357 respondents, 27 were outpatients at mental hospitals or clinics at the time of the first questionnaire, whereas 1276 were not. The others did not answer to the question whether they are outpatients. The clinical and nonclinical populations were compared in terms of subtotal scores and total scores. With the exception of risk to others, all of the subtotal and total scores clearly differentiated these 2 groups.

3.5. The influence of age on total and subtotal CORE-OM scores

We calculated Pearson correlations between age and the 4 subscales: SWB, symptomatic problems, functioning, and risks as well as between age and total CORE-OM score (Table 6). On most of the subscales, older people scored significantly lower, except on the physical symptom subcategory score.

3.6. Convergent validity

The correlations between subtotal and total scores of the CORE-OM and other referential measurement scores are shown in Table 7. Our research suggests that the strength of the Japanese version of the CORE-OM lies in its ability to assess psychologic symptoms rather than physical symptoms. The trauma subcomponent and IES-R score had a correlation coefficient of .62, whereas the physical symptom subcomponent and SDS somatic had a correlation coefficient of .28.

The SWB and functioning subscales as well as the total CORE-OM score highly correlated with "positive relation with others" and "self acceptance" from the SPWB, but not with the SPWB's other 4 subscales (autonomy, environment mastery, personal growth, and purpose in life).

Risk to self correlated highly with HAD anxiety, HAD depression, and IES-R scores. Risk to others, however, did not correlate significantly with BAQ verbal aggression and had a relatively low but significant correlation coefficient of .30 with BAQ physical aggression.

4. Discussion

The CFA proved that Japanese population data fit the 4-structure model acceptably (Fig. 5), which is consistent with the theory behind the original CORE-OM introduced by Evans et al [1]. A possible reason why GFI fell below the acceptable fit value is that GFI decreases with increasing model complexity [17]. Compared with the CFAs for each subscale (Figs. 1-4), the CORE-OM CFA described in Fig. 5 is complicated, which seems to have resulted in the low GFI.

A reason for the low coefficient value from functioning to item 19 might be that, of the 4 close relationship items, only item 19 focuses on the emotional bond that exists in parent-child or spousal or romantic relationships. The other items assess more comprehensive concepts, as explained by the keywords *social support*, *friend*, and *isolation*.

Regarding internal consistency, the complete CORE-OM scale and each of its subscales, with the exception of the SWB subscale, demonstrated favorable Cronbach α coefficients. The relatively lower internal consistency of the SWB subscale can be attributed to the particularly small number of SWB items (4). As pointed out by Barkham et al [2], it is easy to obtain a high Cronbach α coefficient by increasing the number of items. Evans et al [3] also showed that the SWB subscale had the lowest Cronbach α coefficient (.75) of all the subscales in the original version of CORE-OM.

The CFAs indicated that each subscale could be regarded as an independent single-factor structure measurement (Figs. 1-4). Therefore, the CORE-OM can be regarded as an integrated scale composed of 4 independent subscales. Each subscale score correlated highly with the other subscale scores as well as with the total score of the CORE-OM (Table 3), and high covariances were observed between any combination of 2 of the 4 latent variables in the CFA (Fig. 5). This means that, first, all the subcategories included in the CORE-OM are independent subscales, and second, each subscale is significantly influenced by the others. This implies the appropriateness of using either the total score or selected subscale scores of the CORE-OM as outcome measurement indices.

4.1. Sex differences

In comparison with males, females scored lower on close relationships with others and risk to others (Table 5). The reason why females appeared to be more satisfied with regard to close relationships with others may be related to Japanese culture. In Japan, males have traditionally been expected to be independent, less talkative, noncomplaining, and able to solve problems by themselves. On the other hand, females are allowed to be dependent on others; therefore, they are more communicative. These social expectations could influence the degree of satisfaction with close relationships.

Table 7
Correlation between each component of CORE-OM subscales and other scale scores (convergent validity)

		SWB	Symptomatic problems					Functioning				Risk			Total CORE-OM
		SWB	Anxious	Depressed	Physical	Trauma	Total	General	Social	Close	Total	Self	Others	Total	
SPWB	Autonomy	-.21***	-.15***	-.17***	-.04	-.11***	-.15***	-.27***	-.06*	-.11***	-.18***	-.08**	-.01	-.07**	-.18***
	Mastery	-.31***	-.26***	-.30***	-.15***	-.22***	-.29***	-.34***	-.25***	-.17***	-.32***	-.18***	-.08**	-.17***	-.31***
	Growth	-.14***	-.13***	-.24***	-.15***	-.14***	-.20***	-.26***	-.20***	-.26***	-.30***	-.19***	-.17***	-.21***	-.25***
	Relation	-.35***	-.37***	-.44***	-.26***	-.40***	-.44***	-.30***	-.49***	-.49***	-.55***	-.40***	-.25***	-.40***	-.51***
	Purpose	-.24***	-.26***	-.38***	-.26***	-.30***	-.36***	-.24***	-.35***	-.35***	-.41***	-.36***	-.21***	-.36***	-.40***
	Acceptance	-.51***	-.45***	-.56***	-.32***	-.46***	-.54***	-.50***	-.44***	-.42***	-.58***	-.40***	-.19***	-.38***	-.59***
	Total	-.45***	-.42***	-.54***	-.30***	-.42***	-.51***	-.48***	-.46***	-.47***	-.60***	-.42***	-.23***	-.41***	-.58***
HAD	Anxiety	.66***	.71***	.62***	.51***	.63***	.74***	.42***	.59***	.44***	.63***	.55***	.23***	.41***	.74***
	Depression	.54***	.52***	.58***	.48***	.51***	.62***	.41***	.55***	.49***	.63***	.49***	.24***	.47***	.66***
SDS somatic		.30***	.25***	.27***	.28***	.23***	.31***	.25***	.25***	.24***	.31***	.25***	.05	.21***	.34***
IES-R		.55***	.60***	.55***	.47***	.62***	.66***	.32***	.52***	.38***	.53***	.52***	.20***	.48***	.65***
BAQ	Verbal	-.12***	-.15***	-.14***	-.07*	-.12***	-.15***	-.19***	-.09*	-.15***	-.18***	-.07*	.03	-.04	-.16***
	Physical	.10***	.15***	.18***	.08**	.16***	.16***	.13***	.19***	.06*	.17***	.20***	.30***	.26***	.19***

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Concerning the SWB subcategory, male respondents scored lower, which indicates that they were more satisfied with their psychologic well-being. Although the reasons for this result are unclear, it could be partly attributed to economic issues and, surprisingly, the progress made by international women's movements. Contrary to tradition, Japanese women are now expected to work and help support their family financially. On the surface, the idea of working woman appears in line with the equal opportunity goals of the women's movement. However, Japan does not have well-developed social support systems for working women. In addition, in most Japanese families, husbands still stick to the traditional "husband role," that is, they do not contribute to child rearing and housework duties. These factors lead to women feeling overwhelmed, inducing lower levels of satisfaction with their psychologic well-being. One of the limitations of this study is that we did not survey each respondent regarding on their marital status, number of children, or their level of cooperation with their husbands in terms of housework.

The fact that males often score higher in risk to others can be thought to derive from both biologic and social factors, that is, males are expected to be aggressive and dominant. For females, in Japan, it is regarded as virtue to be more passive and less decisive.

4.2. Discrimination between clinical and nonclinical subjects

On the complete scale and all subscales, except risk to others, nonclinical subjects scored lower than clinical ones, indicating that the Japanese version of CORE-OM is potentially useful in discriminating between these 2 populations. Considering the relatively small size of our clinical sample (n = 27), it would be inappropriate to form a definitive conclusion concerning the scale's discrimination capability.

4.3. Influence of age

The subscale scores indicated that older individuals tended to score on the lower end of the scale on general functioning as well as on all the problematic symptoms excluding physical symptoms. In other words, the older people demonstrated the higher general functioning levels and fewer mental symptoms. The questions of the 4 general functioning items were related to whether an individual feels that they can accomplish their purpose or in hindsight perceives that they were actually able to achieve what they wanted to, as well as their level of satisfaction with their previous actions. It is plausible that abundant life experience and modest self-expectations or ideals could contribute to a high self-acceptance and self-efficacy among older people.

The same factors could contribute to the lower level of mental symptoms among older people. They do not have high expectations of themselves or their lives because they have come to terms with loss and, in doing so, have given up a large part of their narcissism. Without unrealistically high aspirations, they do not become seriously depressed or

anxious and overcome traumatic experience by compromising with reality.

The physical symptom subscale was the only exception in terms of correlation with age: the older the individual, the higher the subscale score (Table 6). This is not surprising because quite naturally, aging people have a variety of lifestyle-related diseases.

4.4. Convergent validity

The complete scale and each subscale score correlated well with conceptually similar measurement scores (Table 7). The exception was no significant correlation between the risk to others and the BAQ verbal aggression subscales. This may be because the BAQ is suitable for assessing the behavioral traits that form part of the personality, whereas the CORE-OM focuses on current immediate aggressive impulses. It should be noted that all CORE-OM subscales had relatively low correlation coefficients with the BAQ-verbal and BAQ-physical subscales. In this study, we were not able to prove the convergent validity of the risk to others subcategory, although its factor structure was confirmed.

Considering the above results, we outline the use of the Japanese version of the CORE-OM. Its strengths include readability and user-friendliness, both noted as important factors by Barkham et al [2]. The Japanese version of the CORE-OM is as readable as the original and so can be adapted to a wide range of populations. The fact that it includes only 34 items means that it poses only a small burden on the respondent and is therefore suitable for use in both research and clinical settings.

As emphasized by Barkham et al [2], this measurement is not intended to eliminate other existing measurements. Rather, it should be used as a basic measure that both clinicians and researchers can use in combination with additional measurements to assess the specific symptoms or concepts they want to focus on.

In conclusion, this study demonstrated the validity and reliability of the Japanese version of the CORE-OM, and its use will help in bridging the gap between researchers and clinicians.

References

[1] Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: Clinical Outcomes in Routine Evaluation. J Ment Health Nurs 2000;9:247-55.

[2] Barkham M, Evans C, Margison F, MaGrath G, Mellor-Clark J, Milne D, et al. The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. J Ment Health Nurs 1998;7:35-47.

[3] Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardized brief outcome measure: psychometric properties and utility of the CORE-OM (Clinical Outcomes in Routine Evaluation-Outcome Measure). Br J Psychiatry 2002;180:51-60.

[4] Ryff CD. Happiness is everything, or it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 1989;57: 1069-81.

[5] Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67:361-70.

[6] Bjelland I, Dahl AA, Hang TT, Neckelmann D. The validity of the Hospital and Depression Scale. an updated review. J Psychosom Res 2002;52:69-77.

[7] Kugaya A, Akechi T, Okuyama T, Okamura H, Uchitomi Y. Screening for psychological distress in Japanese cancer patients. Jpn J Clin Oncol 1998;28:333-8.

[8] Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63-70.

[9] Kitamura T, Hirano H, Chen Z, Hirata M. Factor structure of the Zung Self-rating Depression Scale in first year university students in Japan. Psychiatry Res 2004;128:281-7.

[10] Weiss DS, Marmar CR. The Impact of Event Scale—revised. In: Wilson J, & Keane TM, editors. Assessing psychological trauma and PTSD. New York: Guilford Press; 1997. pp. 399-453.

[11] Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measurement of subjective stress. Psychosom Med 1979;41:209-18.

[12] Asukai N, Nishizono MF, Miyake Y. Kaitei dekigoto inpakuto shakudo (IES-R) nihongoban oyobi PTSD rinshosindanmensetsushakudo (CAPS) nihongoban no kaihatsu (in Japanese). [The development of the Japanese version of Impact of Event Scale (IES-R) and Clinician-Administered PTSD Scale for DSM-IV.]. Heisei 10 nenndo koseisyo seisinshinkeisikkankenkyuhi "gaishogo sutoresushogai-kanrenshogaino byotaito tiryo gaidorainni kansuru kenkyuhan" kenkyu hokokukai shoroku. Proceeding of the Ministry of Health, Labour, and Welfare 1998 Annual Meeting of Research on the Psychophysiology and Treatment Guidline of PTSD; 1999.

[13] Asukai N. Huanshogai gaishogo sutoresushogai (PTSD) (in Japanese) [anxiety disorder and posttraumatic stress disorder (PTSD)] Rinshoseisinigaku Zokango. Clin Psychiatry Suppl 1999:171-7.

[14] Buss AH, Perry M. The aggression questionnaire. Pers Processes Individ Differ 1992;63:452-9.

[15] Buss AH, Durkee A. An inventory for assessing different kinds of hostility. J Couns Psychol 1957;21:343-9.

[16] Ando A, Soga S, Yamasaki K, Shimai S, Shimada H, Utsuki N, et al. Nihongoban Buss-Perry kougekisei shitsumon yoshi (BAQ) no sakusei to datousei shinraisei no kento (Development of the Japanese version of the Buss-Perry Aggression Questionnaire (BAQ)). Shinrigaku kenkyu: The Jpn J Psychol 1999;70:384-92 [in Japanese].

[17] Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. MPR Online 2003;8:23-74.