Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale

PER BECH, LIS RAABAEK OLSEN, Psychiatric Research Unit, Frederiksborg General Hospital, Hillerod, Denmark

METTE KJOLLER, NIELS KRISTIAN RASMUSSEN, National Institute of Public Health, Copenhagen, Denmark

ABSTRACT *The health status questionnaire Short-Form 36 (SF-36) includes subscales measuring both physical health and mental health. Psychometrically, the mental health subscale contains a mixture of mental symptoms and psychological well-being items, among other things, to prevent a ceiling effect when used in general population studies. Three of the mental health well-being items are also included in the WHO-Five well-being scale. In a Danish general population study, the mental health subscale was compared psychometrically with the WHO-Five in order to evaluate the ceiling effect. Tests for unidimensionality were used in the psychometric analyses, and the sensitivity of the scales in differentiating between changes in self-reported health over the past year has been tested.*

The results of the study on 9,542 respondents showed that, although the WHO-Five and the mental health subscale were found to be unidimensional, the WHO-Five had a significantly lower ceiling effect than the mental health subscale. The analysis identified the three depression symptoms in the mental health subscale as responsible for the ceiling effect. The WHO-Five was also found to be significantly superior to the mental health subscale in terms of its sensitivity in differentiating between those persons whose health had deteriorated over the past year and those whose health had not.

In conclusion, the WHO-Five, which measures psychological well-being, reflects aspects other than just the absence of depressive symptoms.

Key words: mental health, well-being, SF-36, WHO-Five, ceiling effect

Introduction

Many questionnaires designed to measure negative affects, such as the Hospital Anxiety and Depression Scale (HAD – Zigmond et al., 1983), or to measure positive affects, such as the Psychological General Well-Being Schedule (PGWB – Dupuy, 1984), include a mixture of negatively worded (distress) items and positively worded (well-being) items. A well-being subscale (apart from the anxiety and depression subscales) has thus been derived from the HAD scale (Bech et al., 2002), and an anxiety and depression subscale (apart from the psychological well-being subscale) can be derived from the PGWB schedule (Dupuy, 1984). However, such a mixture of distress items and well-being items has become increasingly problematic because, on the one hand, well-being scales are now considered to be important dimensions of health-related quality-of-life instruments (Bech, 1995, 2001) and, on the other hand, depression scales have to cover authoritatively the universe of depression symptoms as found in the DSM-IV (APA, 1994) or the ICD-10 (WHO, 1993) algorithms of major depression (Bech et al., 2001).

The construct validity of the Medical Outcome Studies (MOS) Short-Form 36 (SF-36) Health Survey questionnaire (Ware and Sherbourne, 1992; Keller et al., 1998) is to measure physical functioning, mental functioning, and general well-being. Factor analysis from

the results of the International Quality of Life Assessment (IQOLA) project has indicated that the general well-being factor contained the positively worded items of the SF-36 Mental Health subscale and the vitality subscale (Keller et al., 1998). The psychometric argument for including both distress and well-being items in the Mental Health subscale of SF-36 (Ware and Sherbourne, 1992; Gandek et al., 1998) has been that it would decrease the floor and ceiling effects – to avoid having an excessively high percentage of respondents at the very low end of the scale (floor effect) or at the very high end of the scale (ceiling effect).

The Mental Health subscale is derived from the PGWB schedule and consists of five items of which three are distress (negative affect) items and two are well-being (positive affect) items.

From a pool of items including both distress and well-being items similar to the PGWB, the WHO-Five Well-Being Scale (WHO-Five) has been adapted into a unidimensional scale (Bech et al.,1996; Heun et al., 1999; Bech, 2001; Bech et al., 2003). It has recently been shown that the WHO-Five, with its five well-being items, fulfilled the Mokken analysis for scalability (Mokken, 1971) with a coefficient of homogeneity of 0.63 and a Cronbach coefficient alpha of 0.88 (Zierau et al., 2002). Table 1 shows firstly that three of the five well-being items in the WHO-Five are also included in the SF-36, namely two from the Mental Health subscale and one item from the vitality subscale. Table 1 also shows that the WHO-Five items represent positive affect compared to the corresponding depression symptoms (negative affect) of major depression. Table 1 shows thirdly that the WHO-Five covers the three core items of depression according to ICD-10 (WHO, 1993) – mood, interests, and energy –

whereas the Mental Health subscale only covers mood and energy. The concept of mental health obviously covers more than the three items, but they are important features of the area of affective states.

This study is part of a Danish general population survey conducted with SF-36 (Kjoller and Rasmussen, 2002) to which the two well-being items of the WHO-Five (Table 1) had been added. This version of the WHO-Five has been compared with the Mental Health subscale in terms of floor and ceiling effects. In the first Danish population study with SF-36 (Kjoller et al., 1995; Bjorner et al., 1998a) a mean score of 82 was obtained for the Mental Health subscale. In another Danish general population study (Olsen et al., 2000), with the PGWB, a mean score of 68 was obtained for the WHO-Five. From this background we hypothesized that the WHO-Five, although it contained pure well-being items, should have a lower ceiling effect than the Mental Health subscale. This psychometric problem reflects a conceptual problem of measuring mental health – to what extent is the absence of mental symptoms equal to a high degree of psychological well-being? A comparison between the WHO-Five and the SF-36 subscale of mental health was made by investigating their psychometric properties in a general population study.

Methods

Scales
The Danish version of the SF-36 has been reported earlier with regard to the validity of its translation procedure, its applicability in a general population study, and the norm data of its eight subscales (Kjoller et al., 1995, Bjorner et al., 1998a,b).

Table 1. Content validity of the SF-36 items and the WHO-Five compared to DSM-IV/ICD-10 symptoms of depression

| DSM-IV/ ICD-10 Major Depression (negative affect) | The Mental Health items | | The Vitality item | The WHO-Five items |
	Distress (negative affect)	Well-being (positive affect)	Well-being (positive affect)	Well-being (positive affect)
Depressed mood	Down in the dumps/ Downhearted and blue	Been happy		Been happy
Lack of interests				Interested in things
Lack of energy			Lots of energy	Lots of energy
Restless/ slowed down	Very nervous	Calm and peaceful		Calm and peaceful
Sleep disturbances				Fresh and rested
Number of separate items	*3*	*2*	*1*	*5*

As it was the objective of the present study to compare a scale containing a mixture of distress and well-being items (the Mental Health subscale) with a scale consisting of pure well-being items (WHO-Five), three other SF-36 subscales have been included, one of which ('general health') contains a mixture of such items and two containing pure distress items ('role emotional' and 'bodily pain').

One of the SF-36 items is not included in the various subscales because it is not a measure of the current state (past four weeks) but covers the change in global health over the last 12 months. This item asks: 'Compared to one year ago, how would you rate your health in general now?' This item was used to compare the sensitivity of the Mental Health subscale with that of the WHO-Five.

Data collection

Data on the general Danish population were collected in 2000 as part of a population health survey. A representative sample of 16,684 non-institutionalized Danish citizens over 15 years of age was drawn from the Central Civil Registration System, which registers addresses and other data on all people living in Denmark. The survey included a home visit with a 30-minute structured personal interview regarding social and demographic data, health behaviour, health status and diseases. After the interview, the respondents were given the SF-36 questionnaire that also included the two additional WHO-5 items and questions from other self-administered scales, among them the EuroQol (1990). The SF-36 was placed as the first in this sample of questionnaires. Of the 16,684 potential respondents, 4,351 (26.0%) persons did not take part in the personal interview, and 306 (1.8%) did not receive the questionnaires. Thus, in total, 12,027 persons received the questionnaires and 1,569 (13.0%) did not return them. Of the returned questionnaires, 916 were not fully completed; thus, in total, 9,542 (or 57.2% of the original sample and 79.3% of the persons who had received the questionnaires) had completed the questionnaires fully.

Psychometric analysis

The overall comparison of the WHO-Five with the SF-36 subscales of 'mental health', 'general health', 'role emotional' and 'bodily pain' has been based on the mean and standard deviation on a scale from 0 (lowest possible state) to 100 (best possible state). The

intercorrelations of these states have been analysed using the Spearman correlation coefficient.

The specific comparison of the WHO-Five with the Mental Health subscale comprised internal consistency, homogeneity or unidimensionality, and floor versus ceiling effects.

The Cronbach coefficient alpha (Cronbach, 1951) was used as a measure of the internal consistency of the scales. The number of items is lower than six, in both the WHO-Five and in the Mental Health subscale, so a coefficient alpha of 0.86 or higher has been considered adequate, although this is not a sufficient condition for homogeneity or unidimensionality (Cortina, 1993).

A non-parametric evaluation of homogeneity, the Loevinger coefficient, was used in accordance with the Mokken scale analysis (Mokken, 1971; Molenaar et al., 1994). The Mokken analysis of homogeneity or unidimensionality is a measure of the extent to which an extra item fits into the structure provided by the other items of the scale. The level of acceptance, according to the Mokken analysis, is a Loevinger coefficient of homogeneity of 0.40 or higher (Mokken, 1971).

As both the WHO-Five and the Mental Health subscale are transformed into 0–100 scales, the ceiling and floor effects have been statistically investigated in a stepwise comparison of the following sections of the 0–100 scale: 100; 99–76; 75–51; 50–26; 25–1; 0. The ceiling effect is defined as a higher percentage of observations in the score intervals of 100 and 99 to 76.

Non-parametric tests have been used (Siegel, 1956). In the stepwise comparison of ceiling and floor effects the chi-square test was applied. The Wilcoxon test has been used to compare two groups (male versus female) and the Kruskal–Wallis test was used to compare more than two groups.

Results

Of the 9,542 individuals who fully completed the SF-36 questionnaire, 4,681 were males and 4,924 females. The mean age was 44.9 years, with a standard deviation of 16.9 years.

Table 2 shows the means and standard deviations of the SF-36 subscales (mental health, role emotional, general health and bodily pain) as well as the WHO-Five. The scales are rank ordered, with the highest mean scores listed at the top of Table 2 and the lowest mean scores at the bottom. The number of distress items, rather than the mixture of distress and

well-being items, was associated with increasing mean scores. The mean scores were significantly higher for males than for females in all subscales listed in Table 2, apart form the general health subscale. This subscale includes an item measuring global physical well-being. On this global item, males scored significantly higher than females, as indicated in Table 2. The WHO-Five had significantly lower values (p < 0.01) than the four SF-36 subscales in Table 2.

Table 3 shows the comparison between WHO-Five and the Mental Health subscale in terms of ceiling and floor effects. The Mental Health subscale had a statistically significant higher percentage of respondents in the two highest categories of the 0–100 scale (in the interval from 76 and including 100) in which 9.9% plus 68.8%, or 78.9% of the observations had been placed. For the WHO-Five, 2.0% plus 47.6%, or 49.6% of the observations had been placed in the two highest categories.

The respondents were subdivided into two groups in order to investigate the sensitivity of the Mental Health subscale compared to the WHO-Five with regard to changes in global health over the past year. The percentage of responders answering that their health had been somewhat or much worse was 7.5% while 92.5% had better or about the same health (11.2% and 81.3% respectively). For the 7.5% with worse health, the mean score and (standard deviation) were 67.0 (20.3) and 46.9 (21.9) on the Mental Health subscale and WHO-Five respectively. For the other group with the same or better health, the results were 84.1 (13.4) and 70.5 (17.6) respectively. The difference in sensitivity between the mental health

mean scores (84.1 – 67.0, or 17.1) and WHO-Five (70.5 – 46.9 or 23.6) were statistically significant (p < 0.01).

The internal consistency of the WHO-Five and the Mental Health subscale in terms of the Cronbach coefficient alpha was 0.84 and 0.81, respectively, which is just acceptable when using the criteria for short scales (Cortina, 1993).

The Loevinger coefficient of homogeneity was 0.56 for both the WHO-Five and the Mental Health subscale, which is above the level of 0.40 for acceptance of unidimensionality. Table 4 shows the rank order of the eight items that define the WHO-Five and the Mental Health subscale. The items that are most inclusive (highest mean score values) are listed at the top of Table 4. The individual coefficients of homogeneity according to the Mokken analysis were all above 0.40, which indicates that the hierarchical order as listed in Table 4 is of statistical significance. Thus, most respondents had answered 'no' to the top-listed items ('down in the dumps', 'a very nervous person', and 'been downhearted and blue'). These items measure mental dysfunction or symptoms of depression. Most exclusive, with the fewest respondents, are the two WHO-Five items of being 'fresh and rested' and having felt 'a lot of energy'.

Table 5 shows the intercorrelation between the various SF-36 subscales including the WHO-Five. The highest Spearman coefficient was obtained for WHO-Five versus mental health (0.84) whereas the correlation of 'role emotional' versus WHO-Five or versus mental health showed rather low coefficients (below 0.40).

Table 2. Mean and standard deviation of the SF-36 subscale, WHO-Five and the SF-36 item of global physical well-being

| Scales | Type of items | | Males plus females | Males | Females | p[1] |
	Distress	Well-being	(N = 9,542)	(N = 4,681)	(N = 4,924)	
Role emotional	3	0	88.0 (26.7)	89.6 (25.0)	86.4 (28.1)	≤ 0.01
Mental health	3	2	82.4 (14.8)	84.4 (14.0)	81.3 (15.4)	≤ 0.01
Bodily pain	2	0	79.0 (23.5)	81.9 (22.1)	76.3 (24.4)	≤ 0.01
General health	2	3	76.9 (19.9)	77.4 (19.1)	76.5 (20.7)	ns
WHO-Five	0	5	68.7 (19.0)	70.6 (18.2)	66.9 (19.5)	≤ 0.01
Global physical well-being (SF-36 item 1)	0	1	64.5 (21.7)	65.2 (21.4)	63.8 (22.0)	≤ 0.05

1. Significance of gender difference

Table 3. Floor versus ceiling effects of the SF-36 Mental Health subscale and WHO-Five

Scale categories	Percentage of respondents		p
	WHO-Five (%)	Mental Health (%)	
Excellent: 100	2.0	9.9	≤ 0.01
Very good: 99–76	47.6	68.8	≤ 0.01
Good: 75–51	33.6	17.3	≤ 0.01
Fair: 50–26	13.2	3.5	≤ 0.01
Poor: 25–1	3.4	0.5	≤ 0.01
Minimal: 0	0.1	0.1	ns

Table 4. Rank order of the eight items of WHO-Five and Mental Health according to the Mokken analysis; the most inclusive items (highest score values) are listed at the top

Mental Health subscale	Have you been so down in the dumps that nothing could cheer you up?	(0.60)
	Have you been a very nervous person?	(0.51)
	Have you felt downhearted and blue?	(0.59)
	Have you been interested in things?	(0.48)
WHO-Five scale	Have you been a happy person?	(0.59)
	Have you been calm and peaceful?	(0.55)
	Have you been fresh and rested?	(0.52)
	Did you have a lot of energy?	(0.63)

Table 5. Intercorrelations (Spearman coefficients) of WHO-Five and the four SF subscales

	WHO-Five	Mental Health	Bodily Pain	General Health	Role Emotional
Mental health	0.84				
Bodily pain	0.40	0.36			
General health	0.50	0.43	0.48		
Role emotional	0.38	0.39	0.26	0.34	
Age	0.08	0.09	–0.08	–0.25	–0.14

Discussion

According to the World Health Organization (WHO, 1948), health is not only the absence of symptoms but is also social, mental and physical well-being. Within this framework of mental health, the WHO-Five reflects mental well-being and the SF-36 Mental Health subscale reflects mental symptoms. Thus, the study has shown that the ceiling effect, operating to a significantly higher degree in the Mental Health subscale than in the WHO-Five, was attributable to the three depression symptoms. In other words, most respondents denied being so unhappy that nothing could cheer them up, being a nervous person, or being downhearted or blue. In the general population, the mean score of the Mental Health subscale was found to be 82, or close to excellent, which is in agreement with Stewart et al. (1992). The mean score of the WHO-Five was found to be only 69, indicating that mental well-being is not equal to absence of symptoms.

Recently, a factor-analytical study on SF-36 performed within the IQOLA project (Keller et al., 1998) has indicated that the construct validity of the SF-36 items includes three factors: physical functioning, mental functioning, and general well-being. Although the general well-being factor covers the positively worded items of the Mental Health

subscale and the vitality subscale (included in WHO-Five) it also covers the positive item of physical well-being in the general health subscale. When this item was analysed separately the mean score was around 65, which is even lower than the mean score for the WHO-Five. Physical well-being can be measured reliably on the SF-36 by one single global item of health, but more than one item is necessary to measure psychological health. The unidimensionality of the WHO-Five found by the Mokken analysis in this study has also been shown by Zierau et al. (2002).

A scale measuring degrees of positive well-being seems to be easier to use as a screening instrument for mental dysfunction or disabilities than scales already using the language of distress or disability. In a comprehensive study, Henkel et al. (2002) have compared the WHO-Five with the General Health Questionnaire (Goldberg, 1972), which also includes a mixture of distress and well-being items, and with the PRIME MD (Spitzer et al., 1994), which only includes symptoms of disability when screening for symptoms of major depression in the primary care setting. They found that the time needed for completion was one to two minutes for the WHO-Five (and around five minutes for the other scales) without showing any inferiority concerning sensitivity, specificity, or positive or negative predictive value for major depression. Likewise, Folker and Jensen (2001) showed, when monitoring treatment outcome in patients with mental disorders, that the WHO-Five was much easier to complete and much more valid than the Sickness Impact Profile.

Furthermore, the WHO-Five was found to be more sensitive than the Mental Health subscale when the responders were subdivided into those with and those without a negative change in global health over the past year. As discussed elsewhere (Bech 1995), this emphasizes the importance of measuring the risk zone of decreased psychological well-being.

In conclusion, it is the number of distress items that is responsible for the ceiling effects in the SF-36 mental and physical health subscales. Thus, mental health is not only the absence of depression symptoms but also a reasonably high degree of psychological well-being.

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4 edn (DSM-IV). Washington DC: American Psychiatric Press, 1994, p. 32.

Bech P. Quality-of-life measurements for patients taking which drugs? The clinical PCASEE perspective. PharmacoEconomics 1995; 7: 141–51.

Bech P. Male depression: stress and aggression as pathways to major depression. In Dawson A, Tylee A (eds) Depression: Social and Economic Timebomb. London: BMJ Books, 2001, pp. 63–6.

Bech P, Gudex C, Staehr Johansen K. The WHO (Ten) Well-Being Index: Validation in Diabetes. Psychother Psychosom 1996; 65: 183–90.

Bech P, Hackett D, Pitrosky B. Efficacy of Venlafaxine on a derived self-rated scale on well-being in generalized anxiety disorder (GAD) and major depressive disorder (MDD). Poster Abstract. Int J Neuropsychopharmacol 2002a; 5 (suppl 1): S212.

Bech P, Moses R, Gomis R. The effect of prandial glucose regulation with repaglinide on treatment satisfaction, wellbeing and health status in patients with pharmacotherapy-naïve Type 2 diabetes: a placebo-controlled, multicentre study. Qual Life Res 2003; 12: 413–25.

Bech P, Rasmussen N-A, Olsen LR, Noerholm V, Abildgaard W. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001; 66: 159–64.

Bjorner JB, Damsgaard MT, Watt T, Groenvold. Tests of data quality, scaling assumptions and reliability of the Danish SF-36. J Clin Epidemiol 1998b; 51: 1001–11.

Bjorner JB, Thunedborg K, Kristensen T, Modvig J, Bech P. The Danish SF-36 health survey: translation and preliminary validity studies. J Clin Epidemiol 1998a; 51: 991–9.

Cortina JM. What is coefficient alpha? An examination of theory and applications. J Applied Psychol 1993; 78: 98–104.

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297–334.

Dupuy HJ. The Psychological General Well-Being (PGWB) Index. In Wenger NK, Mattson ME, Furberg CD, Elison J (eds) Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq Publishing, 1984, pp. 184–8.

EuroQoL Group. EuroQoL – a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: 199–208.

Folker H, Jensen BM. Undersøgelse af udvalgte metoder til selvvurdering af helbred, livskvalitet og tilfredshed med behandling. [Study of selected methods for the self-assessment of quality of life and treatment satisfaction – Danish]. Ugeskr Laeger 2001; 163: 3347–52.

Gandek B, Ware JE Jr, Aaronson NK, Alonso J, Apolone G, Bjorner J, Brazier J, Bullinger M, Fukuhara S, Kaasa S,

Leplege A, Sullivan M. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998; 51: 1149–58.

Goldberg D. The detection of psychiatric illness by questionnaires. Oxford: Oxford University Press, 1972.

Henkel V, Braun S, Kohnen R, Allgaier K, Poth E, Stefanek J, Maier W, Möller H-J. Identifying Depression in Primary Care: a Comparison of Different Methods, 2002 (submitted).

Heun R, Burkart M, Maier W, Bech P. Internal and external validity of the WHO Well-Being Scale in the elderly general population. Acta Psychiatr Scand 1999; 99: 171–8.

Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, Bjorner JB, Brazier J, Bullinger M, Kaasa S, Leplege A, Sullivan M, Gandek B. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998; 51: 1179–88.

Kjoller M, Rasmussen NK (eds) Sundhed og sygelighed i Danmark 2000 - og udviklingen siden 1994. [Health and morbidity in Denmark 2000 - and the development since 1994. Danish.] København: Statens Institut for Folkesundhed, 2002.

Kjoller M, Rasmussen NK, Keiding L, Petersen HC, Nielsen GA. Sundhed og sygelighed i Danmark 1994 - og udviklingen siden 1987. [Health and Morbidity in Denmark 1994 – and the Development since 1987 – Danish.] København: DIKE, 1995.

Mokken RJ. A Theory and Procedure of Scale Analysis. Paris: Mouton, 1971.

Molenaar IW, Debets P, Systma K, Hemker BT. User's Manual MSP: A Program for the Mokken Scale Analysis for Polytomous Items (Version 3.0). Groningen: Lec ProGamma, 1994.

Olsen LR, Noerholm V, Rasmussen N-A, Watt T, Groenvold M, Bjorner J. Major depression and psychological well-being in the general Danish population. Eur Psychiatry 2000; 15 (suppl 2): P02.248.

Siegel S. Non-parametric Statistics. New York: McGraw-Hill, 1956.

Spitzer RL, Williams JBW, Kroenke K, Linzer M, DeGruy FV 3rd, Hahn SR, Brody D, Johnson JG. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994; 272: 1749–56

Stewart AL, Ware JE, Sherbourne CD, Wells KB. Psychological distress/well-being and cognitive functioning measures. In Stewart AL, Ware JE (eds) Measuring Functioning and Well-being. Durham: Duke University Press, 1992, pp. 102–42.

Ware JE, Sherbourne C. The MOS 36-Item Short-Form Health Survey (SF-36). Conceptual framework and item selection. Med Care 1992; 30: 473–83.

World Health Organization. Construction in Basic Documents. Geneva: WHO, 1948.

World Health Organization. International Classification of Diseases. 10 edn (ICD-10). Geneva: World Health Organization, 1993.

Zierau F, Bille A, Rutz W, Bech P. The Gotland Male Depression Scale: a validity study in patients with alcohol abuse disorder. Nord J Psychiatry 2002; 56: 265–71.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983; 67: 361–70.

Correspondence: Per Bech, Psychiatric Research Unit, Frederiksborg General Hospital, DK-3400 Hillerod, Denmark.
Telephone (+45) 48 29 32 53.
Fax: (+45) 48 26 38 77.
Email pebe@fa.dk.