

Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population

Alexandra Martin, Ph.D.^{a,*}, Winfried Rief, Ph.D.^a,
Antje Klaiberg, Dipl. Psych.^b, Elmar Braehler, Ph.D.^b

^aPhilipps-University, Faculty of Psychology, Section of Clinical Psychology and Psychotherapy, Gutenbergstrasse 18, D-35032 Marburg, Germany

^bUniversity of Leipzig, Medical Faculty, Stephanstrasse II, D-04103 Leipzig, Germany

Received 22 April 2005; accepted 13 July 2005

Abstract

Objective: The aim of this study was to assess the validity of the Patient Health Questionnaire depression module (PHQ-9). It has been subject to studies in medical settings, but its validity as a screening for depression in the general population is unknown.

Method: A representative population sample (2066 subjects, 14–93 years) filled in the PHQ-9 for diagnosis [major depressive disorder, other depressive disorder, depression screen-positive (DS+) and depression screen-negative (DS–)] and other measures for distress (GHQ-12), depression (Brief-BDI) and subjective health perception (EuroQOL; SF-36).

Results: A prevalence rate of 9.2% of a current PHQ depressive disorder (major depression 3.8%, subthreshold other depressive disorder 5.4%) was identified. The two depression groups had higher Brief-BDI and GHQ-12 scores, and reported lower health status (EuroQOL) and health-related quality of life (SF-36) than did the DS– group (P 's < .001). Strong associations between PHQ-9 depression severity and convergent variables were found (with BDI $r = .73$, with GHQ-12 $r = .59$).

Conclusion: The results support the construct validity of the PHQ depression scale, which seems to be a useful tool to recognize not only major depression but also subthreshold depressive disorder in the general population.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Depression; Diagnosis; Epidemiology; Health status; Screening

1. Introduction

Depressive disorders are fairly prevalent conditions in the general population and especially in primary care and general hospital settings. Depression is associated with severe impairment in physical, social and role functioning, and with higher health care utilization [1,2]. Depressive symptoms are frequent coexisting problems in many medical illnesses and have been found to increase the risk of mortality and morbidity, such as in coronary heart diseases [3,4]. On the other hand, patients suffering from depressive disorders often do not seek help for psychological problems, but instead present somatic symptoms to their physicians, and their depression often goes unrecognized [5,6]. According to the WHO Psychological Problems in

General Health Care study, only 42% of primary care patients with major depression were recognized by the physician [7]. Therefore it has been emphasized that it is a key challenge in the health care system to identify depressive disorders early. Screening questionnaires have been advocated as an aid to the detection of cases and clinical decision making.

Previous self-report instruments used for case finding or screening of mental disorders yield indices of severity rather than categorical psychiatric diagnoses. The Patient Health Questionnaire (PHQ) is based on *DSM-IV* criteria, its disorders divided into threshold disorders according to *DSM-IV* and subthreshold conditions. It has been developed as a fully self-administered version of the original PRIME-MD by Spitzer et al. [8]. PRIME-MD is a two-stage system consisting of a patient-screening questionnaire and the clinician evaluation guide to detect the most common groups of mental disorders in primary care. The two components of the original PRIME-MD instruments were combined in a

* Corresponding author. Tel.: +49 6421 282 3656 (secretary 282 3657); fax: +49 6421 2828904.

E-mail address: martin@staff.uni-marburg.de (A. Martin).

self-report questionnaire. Two versions are available: the PHQ with the complete diagnostic part (four pages) and the Brief-PHQ (two pages) covering mood and panic disorders.

The PHQ has already been studied in different medical settings, e.g., in primary care patients [8,9], in general hospital inpatients [10] and in obstetrics-gynecological patients [11]. Among 3000 primary care patients, a prevalence rate of 28% of any mental disorder and 10% with major depression and 6% with other depressive disorders was reported [8]. The German version of the PHQ depression module (PHQ-9) has been validated twice in primary care settings: Henkel et al. [9] determined a sensitivity of 78% and a specificity of 85% in the depression module, with specificity and positive predictive value to be better than in the other screening questionnaires GHQ-12 and WHO-5. Löwe et al. [12] reported the PHQ's operating characteristics for major depression to be significantly superior to two other screening instruments (WBI-5 and HADS). Aspects of convergent validity of the PHQ have also been reported in previous studies conducted in medical settings, showing strong associations between PHQ psychiatric diagnosis and functional impairment as well as disability days [8,11]. Further, strong associations of the PHQ-9 depression severity score as a continuous variable with the different aspects of health-related quality of life (SF-20) were found [13]. In sum, these results supply strong evidence for the PHQ as a valid screening instrument in medical settings. Validity of the PHQ in the general population is unknown.

1.1. Aims of the study

Therefore the main subject of the present study was to assess aspects of construct validity of the PHQ-9 in a population-based sample. To compare the PHQ-9 with convergent variables, two screening instruments were chosen to assess psychological distress (GHQ-12) and depression (Brief-BDI). In addition, the relation of diagnostic results of the PHQ-9 to subjective health perception and health-related quality of life was assessed.

We hypothesized that the PHQ-9 diagnostic groups would have different scores on these clinical variables, with lowest disability scores in the DS– and highest scores in the two depressive groups (major depression and other depressive disorder). Further, we expected strong associations between depression severity and the other variables assessing depression, psychological distress, general health and disability.

2. Material and methods

2.1. Subjects

A representative sample of the general population of Germany was selected with the assistance of a demography consulting company (USUMA, Berlin). The sample selection was based on the political election register in 1998, and 201 sample points were used. The sample was

selected to be representative in terms of age, gender and education. Inclusion criteria were age above 13 years and German as a native language. A first attempt was made for 3194 addresses following a random-route procedure. The household respondent was selected by chance. All subjects were visited by an interviewer, informed about the investigation and presented with questionnaires. From the 3194 selected addresses, 3108 were valid. A total of 28.1% disagreed to participate, and 4.7% failed to be contacted (despite four attempts). A total of 2066 persons in the age range between 14 and 93 years agreed to participate. Mean age was 48.8 years (S.D. 18.1); 53% were female. Half of the participants were married (50%), 11% were divorced, 14% widowed, 38% were not married. Fifty-two percent had higher education.

2.2. Assessment instruments

2.2.1. Brief-PHQ

The German version of the Brief-PHQ [14] was used. Translation of the German version followed state-of-the-art procedures in cross-cultural assessment [15]. The depression subscale of the German version showed high internal consistency (Cronbach's alpha .89), and normative data for the depression total scores were reported [16]. The Brief-PHQ covers scales for mood disorders (PHQ-9) and for panic syndrome.

2.2.2. Mood subscale (PHQ-9)

Subjects indicated for each of the nine depressive symptoms (corresponding to the criteria of *DSM-IV*) whether, during the previous 2 weeks, the symptom has bothered them: 0="not at all", 1="several days", 2="more than half of the days" or 3="nearly every day". Subjects were assigned to four diagnostic groups according to the PHQ Office Coding Algorithm to assess the relationship between the PHQ-9 depressive diagnoses and the other variables [8]:

Major depressive syndrome (MDS): is diagnosed if five or more of the nine depressive symptoms are reported to be present at least at "more than half the days" (≥ 2), and one of the symptoms is depressed mood or anhedonia. One of the nine symptom criteria (item "i": "thoughts that you would be better off dead ...") counts if present at all (≥ 1).

Other depressive syndrome (ODS): is diagnosed if only two, three or four of the depressive symptoms are indicated at least at "more than half the days" (with item "i" counted if present at all), and one of the symptoms is depressed mood or anhedonia.

Depression symptom-screen positive (DS+): at least one of the required screening symptoms is fulfilled, but the total symptom score is below the subthreshold diagnosis.

Depression symptom-screen negative (DS–): none of the required screening symptoms is present at "more than half of the days" (scores < 2).

2.2.3. PHQ-9 depression severity

The continuous variable is the sum of scores of the PHQ-9 items (range 0–27).

2.2.4. Panic syndrome

Response categories of five anxiety symptoms are dichotomous (yes/no). Panic syndrome is diagnosed if all five anxiety symptoms are reported to be present (during the last 4 weeks).

2.2.5. GHQ-12

The GHQ [17] is one of the most widely used screening tests for mental disorders internationally. Validity coefficients for the GHQ-12 were found to be generally high [18]. The overall score indicates the “severity of psychological disturbance.” The emphasis is on how the respondent’s present state differs from his/her usual state. The 12-item version with Likert scoring (0–1–2–3) was used in this study (response categories ranging from “less than usual” to “much more than usual”; sum score ranges 0–36).

2.2.6. Brief-BDI

Another instrument used to measure the severity of depression was a shortened version of the Beck Depression Inventory (BDI) [19]. The number of items in the original version [20] was reduced from 84 to 20 items (one item for each symptom; “losing weight” as a symptom was dropped). The frequency of each symptom was measured using six-point rating scales (from 0=*never* to 5=*almost always*). Schmitt and Maes [19] reported the internal consistency (.90) to be good and reported support of the construct validity. The Brief-BDI and the original BDI converge very well on the level of sum scores ($r=.91$). The scores of the 20 items are summed to yield a “total depression score” (range 0–100).

2.2.7. EuroQOL

The visual analogue scale of the German version of the European Quality of Life (EuroQOL) Questionnaire [21,22] was used to assess the present subjective health status (from 0=*worst health status* to 100=*best health status*).

2.2.8. SF-36

The 36-Item Short Form Health Survey (German version [23]) is a multidimensional questionnaire measuring health-related quality of life (based on the WHO definition of health). The SF-36 includes eight subscales (ranging from 2 to 10 items): general health, physical functioning, bodily pain, vitality, mental health, social functioning, physical role limitation and emotional role limitation. Response alternatives vary between two-point and six-point Likert scales. Subscale scores are transformed into a 0–100 scale, with higher scores representing better health.

2.3. Statistical analyses

To assess the differences between the four PHQ-9 diagnostic groups, mean scores of criteria variables were

compared by the Welch test. This procedure is comparable to ANOVA, but takes into account that population variances differ significantly and sample sizes are not equal. If the overall group effect was significant, multiple paired post hoc comparisons of groups were computed with the test of Games–Howell to control for the probability of increasing type I error. This test does not need homogeneity of variances and is recommended in cases of unequal sample sizes. On a descriptive level, data were compared if they fit a linear trend.

PHQ-9 depression severity was correlated with the other construct variables (Pearson). Correlation coefficients were tested for significance level according to hypotheses (with Bonferroni’s correction to adjust for multiple comparisons). All analyses were conducted with SPSS statistical package.

3. Results

3.1. Diagnostic results

Table 1 presents the prevalence of mood disorders and panic syndrome diagnosed by the Brief-PHQ in the representative sample ($n=2060$; in six cases diagnostic information was incomplete/missing). Nearly 10% of the subjects ($n=204$) had at least one of the possible diagnoses. The majority of these subjects fulfilled the criteria of a current mood disorder: 3.8% having major depression and 5.4% having the subthreshold diagnosis of any other depressive disorder. Only 1.9% fulfilled the criteria of panic syndrome and only 18 subjects (0.9%) got both diagnoses.

The prevalence of any mood disorder was significantly higher in females than in males [12.3% vs. 5.7%; $\chi^2(1)=31.5$; $P<.001$]. Depressed subjects were significantly older than nondepressed groups [MDS/ODS mean age=53.4, S.D.=19.0 vs. DS+/DS–=48.4, S.D.=18.0; $t(2085)=3.7$, $P<.001$]. The average family income was significantly lower in the depressed groups than in the two nondepressed groups [$\chi^2(2)=18.2$; $P<.001$].

Table 1
Prevalence rates of current psychiatric disorders detected by Brief-PHQ in a German representative sample ($n=2060$)

	Whole sample		Male		Female	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Any Brief-PHQ diagnosis	204	9.9	58	6.0	146	13.4
Any mood disorder	190	9.2	55	5.7	135	12.3
Major depressive disorder	78	3.8	24	2.5	54	4.9
Other depressive disorder	112	5.4	31	3.2	81	7.4
DS+	27	1.3	13	1.3	14	1.3
DS–	1843	89.5	901	93	942	86.3
Panic syndrome	38	1.9	7	0.7	31	3.0

3.2. Validity of PHQ-9: comparison of diagnostic groups

Subjects were assigned to four different groups according to the degree of criteria fulfillment for major depression assessed by the PHQ-9. Accordingly, the highest score of PHQ-9 depression severity was found in the major depression group (mean=15.92, S.D.=3.07), followed by other depressive disorders (mean=9.62, S.D.=2.19), DS+ (mean=5.54, S.D.=1.82), and the lowest score was in DS– (mean=2.63, SD=2.87). As expected, the group main effect [Welch test $W_j=759$, $df=3$, 86.8, $P<.001$] and all paired comparisons were highly significant ($P<.001$).

Table 2 presents the mean scores on GHQ-12, Brief-BDI and the subjective health perception scale (EuroQOL) in the four diagnostic groups. The group main effects were all highly significant ($P<.001$). The same pattern of increasing scores for depression (BDI), psychological disturbances (GHQ-12) and decreasing perception of subjective health (EuroQOL) was seen from the DS– group, DS+ group to the groups with other depressive disorder and major depression. Scores across groups were increasing/decreasing following a linear trend.

Accordingly, paired comparisons were carried out (see Table 3): In GHQ-12 scores, the DS– group differed significantly from the other three groups (compared with DS+ at $P<.05$, compared to ODS and MDS at $P<.001$). The DS+ group differed significantly from the group with other depressive disorders ($P<.05$) and from major depression ($P=.001$). Only the mean difference between other depressive disorder and major depression was not significant ($P=.372$).

Referring to mean scores of the Brief-BDI version, all groups differed from each other ($P<.001$), except DS–, which did not differ from DS+.

The ratings of subjective health perception (EuroQOL) showed that DS– and DS+ differed significantly from other depressive disorder and major depression ($P<.001$), but did not reveal significant differences between the two screening groups, and between other depressive disorder and major depression.

Table 2 shows the mean scores of SF-36 dimensions, which cover different aspects of health-related quality of life. Again, a similar pattern in all eight subscales was seen with decreasing scores from the DS– and DS+ groups to the group with other depressive disorders, and with lowest scores in the major depression group. The group main effects were all highly significant. Additionally, the test of linear trend was significant for all variables.

Results of the paired comparisons (see Table 3) showed that the scores of the DS– group were significantly higher in all SF-36 subscales compared to both groups with depressive disorders (ODS and MDS), whereas only in a few scales did DS– group differ significantly from the DS+ group. Comparisons of the DS+ group with the major depression group were significant in seven of the eight subscales, and with the other depressive disorders group in

Table 2

Mean scores of GHQ-12, Brief-BDI, subjective health status (EuroQOL) and health-related quality of life (SF-36) by PHQ-9 diagnostic groups

	<i>n</i>	Mean	S.D.	CI (95%)
GHQ12				
DS–	1811	8.96	4.24	(8.77–9.16)
DS+	26	11.81	5.04	(9.77–13.84)
Other depressive disorder (ODS)	109	15.36	6.44	(14.14–16.58)
Major depressive disorder (MDS)	77	16.92	6.50	(15.44–18.39)
Depression (Brief-BDI)				
DS–	1811	15.68	12.31	(15.12–16.25)
DS+	26	20.55	12.57	(15.47–25.62)
Other depressive disorder (ODS)	109	34.26	14.28	(31.54–36.97)
Major depressive disorder (MDS)	77	45.08	16.32	(41.37–48.78)
Health status (EuroQOL)				
DS–	1811	79.17	15.97	(78.43–79.90)
DS+	26	74.96	13.01	(69.71–80.22)
Other depressive disorder (ODS)	109	61.15	22.01	(56.97–65.33)
Major depressive disorder (MDS)	77	55.96	25.37	(50.20–61.72)
SF-36 scales				
Physical functioning				
DS–	1811	89.43	17.72	(88.62–90.25)
DS+	26	83.85	25.03	(73.74–93.96)
Other depressive disorder (ODS)	109	69.94	28.28	(64.57–75.31)
Major depressive disorder (MDS)	77	67.92	31.75	(60.72–75.13)
Physical role functioning				
DS–	1811	84.86	20.44	(83.92–85.80)
DS+	26	74.04	29.46	(62.14–85.94)
Other depressive disorder (ODS)	109	58.14	25.08	(53.38–62.90)
Major depressive disorder (MDS)	77	55.28	31.58	(48.11–62.44)
Pain				
DS–	1811	83.02	22.37	(81.99–84.05)
DS+	26	71.23	31.21	(58.62–83.84)
Other depressive disorder (ODS)	109	52.83	27.26	(47.65–58.00)
Major depressive disorder (MDS)	77	47.56	30.68	(40.60–54.52)
General health				
DS–	1811	69.45	19.54	(68.55–70.35)
DS+	26	55.10	17.33	(48.10–62.09)
Other depressive disorder (ODS)	109	44.77	18.44	(41.27–48.27)
Major depressive disorder (MDS)	77	39.79	18.11	(35.68–43.90)
Vitality				
DS–	1811	68.07	16.14	(67.32–68.81)
DS+	26	57.45	20.31	(49.25–65.66)
Other depressive disorder (ODS)	109	43.29	16.26	(40.20–46.38)
Major depressive disorder (MDS)	77	36.17	17.86	(32.12–40.23)
Social functioning				
DS–	1811	90.90	15.49	(90.19–91.62)
DS+	26	79.81	21.82	(71.00–88.62)
Other depressive disorder (ODS)	109	59.75	20.51	(55.85–63.64)
Major depressive disorder (MDS)	77	52.11	25.52	(46.32–57.90)
Emotional role functioning				
DS–	1811	88.02	19.75	(87.11–88.93)
DS+	26	72.76	24.56	(62.84–82.68)
Other depressive disorder (ODS)	109	60.93	24.21	(56.34–65.53)
Major depressive disorder (MDS)	77	50.65	29.30	(44.00–57.30)
Mental health				
DS–	1811	75.23	14.70	(74.55–75.91)
DS+	26	65.77	14.81	(59.79–71.75)
Other depressive disorder (ODS)	109	49.54	15.12	(46.67–52.41)
Major depressive disorder (MDS)	77	41.69	16.60	(37.92–45.45)

five of the eight scales. Comparison of the two depression groups revealed only two significant effects (in mental health and vitality); thus in most of the subscales they did not differ substantially.

Table 3

Construct validity of PHQ-9 depression subgroups as assessed by the GHQ-12, Brief-BDI, EuroQOL and SF-36 (statistics of overall and paired comparisons)

	Overall comparison				Paired comparison					
	Welch test	df1	df2	P	DS-/DS+	DS-/ODS	DS-/MDS	DS+/ODS	DS+/MDS	ODS/MDS
GHQ12	72.1	3	82.3	<.001	*	***	***	*	**	
Brief-BDI	134.3	3	83.2	<.001		***	***	***	***	***
Health status (EuroQOL)	43.6	3	81.8	<.001		***	***	**	***	
SF-36										
Physical functioning	27.9	3	81.8	<.001		***	***			
Physical role functioning	60.2	3	82.3	<.001		***	***		*	
Bodily pain	73.9	3	82.5	<.001		***	***	*	**	
General health	121.9	3	84.8	<.001	**	***	***	*	**	
Vitality	151.1	3	83.4	<.001		***	***	*	***	*
Social functioning	135.8	3	82.1	<.001		***	***	**	***	
Emotional role functioning	83.8	3	82.5	<.001	*	***	***		**	
Mental health	191.6	3	83.8	<.001	*	***	***	***	***	**

Post hoc comparison (Games–Howell): ***significant at $P < .001$, **significant at $P < .01$, *significant at $P < 0.05$.

3.3. Relationship between PHQ-9 depression severity and indices of convergent validity

PHQ-9 depression severity was strongly associated with the scores of the short version of BDI ($r = .73$; $P < .0001$) and GHQ-12 ($r = .59$; $P < .0001$). The correlation of depression severity with subjective health status (EuroQOL) was lower, but still highly significant ($r = -.50$; $P < .0001$), even when controlling for age (partial $r = -.48$, $P < .0001$).

Further, correlation coefficients of the construct variables were tested to see whether they were significantly different from each other. As hypothesized, the correlation between PHQ-9 depression severity and BDI was higher than between PHQ-9 depression severity and GHQ-12 ($T = 10.48$, $df = 2029$, $P < .001$) and health status, respectively ($T = 14.66$, $df = 2029$, $P < .001$).

The correlations of the PHQ-9 depression severity score with the subscales of SF-36 were moderate to high (ranging from $r = -.45$ for physical functioning to $r = -.71$ for social functioning and mental health). Coefficients are shown in Table 4. As expected, the association of PHQ-9 depression severity was significantly higher with mental health ($r = -.71$) than the association with subscales measuring body functions and disabilities (physical functioning

$T = 14.84$, physical role functioning $T = 11.43$, pain $T = 8.09$, general health $T = 8.30$, $df = 2051$, $P < .001$).

Additionally, the association between depression severity assessed by the PHQ-9 and SF-36 variables was compared to the associations of the two other screening instruments (Brief-BDI, GHQ-12) with the SF-36. Across all three instruments, a comparable pattern of correlations with SF-36 subscales was found. Associations with mental health were the highest, while associations with body functions were lowest. The correlations of PHQ-9 depression severity with all eight SF-36 scales were significantly higher than those of GHQ-12 with the SF-36 (T 's = 6.66–12.5, $df = 2051$, P 's < .001) and those of Brief-BDI with the SF-36 (T 's = 2.62–8.83, $df = 2051$) with one exception of no significant difference in mental health.

4. Discussion

The main aim of this study was to assess the construct validity of the PHQ depression scale PHQ-9 in the general population. While there is already a body of evidence that supports its validity in medical settings [8,10,11], no data are available on its validity in the general population. The diagnoses of the PHQ-9 are based on the criteria for major depression according to *DSM-IV* [24]. An advantage of a screening instrument based on these operational definitions of mental disorders is that it offers the possibility of improving standardization and comparisons across studies.

This study consistently found a strong positive association between depression and disability, and a strong negative association between depression and functional health status across the PHQ-9 diagnostic groups. The two depression groups (major depressive disorder and other depressive disorder) had higher depression scores in the Brief-BDI, higher scores in the GHQ-12 and reported lower subjective health status (EuroQOL) as well as a lower level of functioning in health-related quality of life aspects (SF-36) than did subjects of the DS– group. These results

Table 4

Association of PHQ-9 depression severity, Brief-BDI and GHQ-12 with SF-36 subscales

SF-36	PHQ-9 depression severity	Brief-BDI	GHQ-12
Physical functioning	-.45***	-.39***	-.27***
Physical role functioning	-.55***	-.49***	-.34***
Bodily pain	-.59***	-.51***	-.39***
General health	-.60***	-.56***	-.43***
Vitality	-.68***	-.65***	-.52***
Social functioning	-.71***	-.61***	-.54***
Emotional role functioning	-.58***	-.53***	-.44***
Mental health	-.71***	-.70***	-.62***

$n = 2054$ (only cases with complete data in all subscales were included); Pearson correlation: ***significant at $P < .001$.

provide strong support for the ability of the PHQ-9 to discriminate depressed from nondepressed individuals.

However, the paired comparisons of the two depression groups (major depressive disorder and other depressive disorders) revealed significant differences only in depression scores (Brief-BDI) and in the SF-36 scales mental health and vitality. It is to be expected that these variables would be more closely related to the diagnosis of depression than the other measures that depict a broader concept of disability, therefore discriminating the two diagnostic groups better than the other variables.

Previous results of studies in medical settings provided evidence that the functional level was lowest in the group of “threshold disorder,” followed by “subthreshold disorder,” “symptom screen-positive” and “symptom screen-negative” [8,10,11]. We replicated this finding by showing lower functional level in depressed vs. nondepressed individuals in a general population sample, whereas the two depression groups displayed differences only in a few variables. We did not find differing scores for general health, bodily pain and functioning between major depressive disorder and other depressive disorder.

It has been argued that *subthreshold* or *minor depression* is a variant of depressive disorders that should be considered seriously both as a target for preventive intervention and for treatment [25]. Wells et al. [26] have shown in the Medical Outcome Study that not only patients with major depression but also those with subsyndromal depressive symptoms have additive disability in patients with chronic medical illnesses. Kessler et al. [25] used data from the National Comorbidity Survey to study correlates and course associated with minor depression, major depression with five to six symptoms and major depression with more symptoms. They found a monotonic increase in average number of episodes, impairment and comorbidity across the three groups. There are some parallel findings in our study, i.e., that even the subthreshold depressive syndrome showed almost similar impairment compared to the group with major depression.

The PHQ-9 provides the ability to assess *depression severity* as a continuous measure in addition to the categorical diagnostic groups. Results of this study indicated a strong association between PHQ-9 depression severity and all convergent variables, especially with another indicator of depression (Brief-BDI) and with mental health (SF-36). An additional hint of validity results from the finding that the association between PHQ-9 depression severity with the Brief-BDI was significantly stronger than its association with the GHQ-12.

In the PHQ Primary Care and Obstetrics-Gynecology Study, strong associations between increasing PHQ-9 depression severity and worsening function on the different aspects of health-related quality of life (all six SF-20 scales) were reported, with the strongest negative association between depression severity and mental health, and the lowest in physical functioning and bodily pain [13]. This overall pattern of correlations was confirmed in the present

study. In addition, the associations of PHQ-9 depression severity with the dimensions of the SF-36 were even higher than the associations of two other screening instruments (GHQ-12 and Brief-BDI) with the SF-36.

Data from the representative sample of 2060 Germans indicated the prevalence rate of 9.2% of a current mood disorder (major depression 3.8%, subthreshold depressive disorder 5.4%). In community samples, the point prevalence of major depressive disorder has varied from 5% to 9% for women and 2% to 3% for men [24]. Wittchen et al. [27] reported results of the German National Health Interview and Examination Survey — Mental Health Supplement. In this nationwide field survey, 4181 subjects were interviewed with the full Composite International Diagnostic Interview. They found a 4-week prevalence rate of 6.3% for affective disorders. Compared with the interview-based results, prevalence rates assessed by the self-administered questionnaire did not seem to be generally higher. The PHQ depressive module does not seem to be over-inclusive. Additionally, its results in terms of the socio-demographic characteristics of depressed and nondepressed individuals are similar to former results, e.g., indicating women to be affected by a major depressive disorder almost twice as often as men, and the mean age of depressed subjects to be higher than in nondepressed individuals.

A shortcoming of the present study is that subjects were not assessed with an additional clinical interview to confirm diagnoses. However, there are a number of studies comparing PHQ diagnosis with diagnosis of a structured interview that reported good sensitivity and specificity of the PHQ in medical settings [8,10,12]. The study sample is representative of the general population of Germany. This provides the possibility to compare results with western European and white American populations, but data are not representative of countries with high cultural heterogeneity (e.g., low rates of African Americans and Hispanics in our sample).

The majority of individuals with mental disorders in society are never seen by a mental health professional [28]. The implementation of screening methods for mental disorders in the health care system is suggested to improve detection of affected individuals, although routine administration of screening questionnaires might not be sufficient to increase the overall rate of recognition of mental disorders [29]. Further, recognition of depression alone cannot be expected to be sufficient to improve outcome, but it is a necessary prerequisite. The PHQ with its depressive module seems to be a valid questionnaire to be used for that purpose. Because of its criteria-based structure, it may even be an interesting diagnostic tool in the research of population-based samples where face-to-face diagnostic interviews are not available.

Acknowledgment

This study was supported by grants from the German Ministry of Education and Research BMBF and from Pfizer

Germany, Karlsruhe. The first author has previously published with the surname “Nanke.”

References

- [1] Katon W, von Korff M, Lin E, Lipscomb P, Russo J, Wagner E. Distressed high utilizers of medical care: DSM-III-R diagnosis and treatment needs. *Gen Hosp Psychiatry* 1990;12:355–62.
- [2] Simon G, Ormel J, VonKorff M, Barlow W. Health care costs associated with depressive and anxiety disorders in primary care. *Am J Psychiatry* 1995;152:352–7.
- [3] Carney RM, Freedland KE, Miller GE, Jaffe AS. Depression as a risk factor for cardiac mortality and morbidity. A review of potential mechanisms. *J Psychosom Res* 2002;53:897–902.
- [4] Katon W, Ciechanowski P. Impact of major depression on chronic medical illness. *J Psychosom Res* 2002;53:859–63.
- [5] Roness A, Mykletun A, Dahl AA. Help-seeking behaviour in patients with anxiety disorder and depression. *Acta Psychiatr Scand* 2005;111:51–8.
- [6] Wittchen HU, Lieb R, Wunderlich U, Schuster P. Comorbidity in primary care: presentation and consequences. *J Clin Psychiatry* 1999;60:29–36.
- [7] Simon GE, Goldberg D, Tiemens BG, Ustun TB. Outcomes of recognized and unrecognized depression in an international primary care study. *Gen Hosp Psychiatry* 1999;21:97–105.
- [8] Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999;282:1737–44.
- [9] Henkel V, Mergl R, Kohnen 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;326:200–1.
- [10] Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer R. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosom Med* 2001;63:679–86.
- [11] Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol* 2000;183:759–69.
- [12] Löwe B, Spitzer RL, Gräfe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004;78:131–40.
- [13] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. Validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
- [14] Löwe B, Spitzer RL, Zipfel S, Herzog W. PHQ-D. Gesundheitsfragebogen fuer Patienten [German version of the patient health questionnaire]. Karlsruhe: Pfizer; 2002.
- [15] Bracken B, Barona A. State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *Sch Psychol Int* 1991;12:119–32.
- [16] Rief W, Nanke A, Klaiberg A, Braehler E. Base rates for panic and depression according to the Brief Patient Health Questionnaire: a population-based study. *J Affect Disord* 2004;82:271–6.
- [17] Goldberg D, Williams P. A user's guide to the GHQ. Windsor: NFER-Nelson; 1988.
- [18] Goldberg D, Gater R, Satorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997;27:191–7.
- [19] Schmitt M, Maes J. Vorschlag zur Vereinfachung des Beck-Depressions-Inventars (BDI). *Diagnostica* 2000;46:38–46.
- [20] Beck AT, Steer RA. Beck depression inventory. Manual. San Antonio: The Psychological Cooperation; 1987.
- [21] vd Schulenburg JM, Claes C, Greiner W, Uber A. Die deutsche Version des EuroQuol-Fragebogens. *Z GesundhWiss* 1998;3:210–31.
- [22] EuroQOL-Group. EuroQuol — a new facility to the measurement of health related quality of life. *Health Policy* 1990;16:199–208.
- [23] Bullinger M, Kirchberger I. Der SF-36 Fragebogen zum Gesundheitszustand-Handanweisung. Goettingen: Hogrefe; 1998.
- [24] APA. Diagnostic and statistical manual for mental disorders, fourth edition, text revision. Washington (DC): APA; 2000.
- [25] Kessler RC, Zhao S, Blazer DG, Swartz M. Prevalence, correlates, and course of minor depression and major depression in the national comorbidity survey. *J Affect Disord* 1997;45:19–30.
- [26] Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed patients. Results from the medical outcomes study. *JAMA* 1989;262:914–9.
- [27] Wittchen HU, Müller N, Pfister H, Winter S, Schmidt-kunz B. Affektive, somatoforme und Angststörungen in Deutschland-Erste Ergebnisse des Bundesweiten Zusatzsurveys “Psychische Störungen”. *Gesundheitswesen* 1999;61:216–22.
- [28] Rief W, Martin A, Klaiberg A, Brähler E. Specific effects of depression, panic, and somatic symptoms on illness behavior. *Psychosom Med* [in press].
- [29] Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ* 2001;322:406–9.