THE VALIDITY OF THE PHQ-9 AND THE GAD-7 FOR SCREENING DEPRESSIVE AND ANXIETY DISORDERS IN SICK-LISTED WORKERS

|Wietske Homans 0472956 | Master thesis Clinical and Health Psychology|

|Accompanied by S. Doosje PhD. (UU), M.C. Vlasveld Msc. and D. Volker Msc. (Ti)|

|Utrecht University | Trimbos-institute | June 30, 2012|

ABSTRACT

The aim of this study was to validate the Patient Health Questionnaire-9 for major depressive disorder and the Generalized Anxiety Disorder Scale-7 for any anxiety disorder within a population of workers with long-term sickness absence from work. Participants were recruited in collaboration with the occupational health service. Workers on sickness absence between 4 and 26 weeks were asked to complete both the PHQ-9 and the GAD-7 and were then evaluated with the MINI-interview to assess DSM-IV disorders. Reliability (Cronbach's α) of the PHQ-9 was .92 and of the GAD-7 it was .94. Criterion validity was calculated in terms of sensitivity and specificity. For the PHQ-9 a cut-off value of 10 provided a sensitivity of .86 and a specificity of .78. For the GAD-7 a cut-off value of 9 provided a sensitivity of .82 and a specificity of .76. Based on Receiver Operating Characteristics (ROC) analysis, the area under the curve (AUC) for the PHQ-9 for major depressive disorder was .90. The AUC for the GAD-7 for any anxiety disorder was .85. In conclusion, both tests show good characteristics of reliability and criterion validity.

SAMENVATTING

Dit artikel beschrijft het onderzoek naar de betrouwbaarheid en de validiteit van de Patient Health Questionnaire-9 voor de depressieve stoornis en de Generalized Anxiety Disorder Scale-7 voor angststoornissen voor een populatie van langdurig ziek gemelde werknemers. Participanten werden geworven in samenwerking met de arbodienst. Werknemers die 4 tot 26 weken waren ziek gemeld werden gevraagd om zowel de PHQ-9 als de GAD-7 in te vullen, vervolgens werd hen telefonisch het MINI-interview afgenomen en werden zij eventueel gediagnosticeerd volgens de DSM-IV stoornissen. De betrouwbaarheid (Cronbach's α) van de PHQ-9 was .92 en van de GAD-7 .94. Criterium validiteit werd berekend in termen van sensitiviteit en specificiteit. Een cut-off waarde van 10 leverde voor de PHQ-9 een sensitiviteit op van .86 en een specificiteit van .78. Een cut-off waarde van 9 leverde voor de GAD-7 een sensitiviteit op van .82 en een specificiteit van .76. Gebaseerd op de Receiver Operating Characteristics (ROC) analyse, werd er een Area Under the Curve (AUC) gevonden van de PHQ-9 voor de depressieve stoornis van .90. De AUC van de GAD-7 voor angststoornissen was .85. Concluderend, beide testen laten goede karakteristieken van zowel betrouwbaarheid als criterium validiteit zien.

1. INTRODUCTION

Depressive and anxiety disorders have a significant influence on work functioning (Plaisier et al., 2010). As a consequence, depression and anxiety are strongly associated with long-term sickness absence from work (Terluin, Van Rhenen, Anema & Taris, 2011). It is therefore of great importance to be able to identify these disorders at an early stage, allowing workers to be treated and longer absence to be prevented. The current study examines the validity of the Dutch versions of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder Scale-7 in screening depression and anxiety in a population of workers with long-term sickness absence.

The lifetime prevalence of major depressive disorder in the general population of the Netherlands is 15.4% (Bijl, Ravelli & Van Zessen, 1998). The lifetime prevalence of anxiety disorders in the same population is 19.3% (Bijl et al., 1998). The year prevalence of major depressive disorder in the working population of the Netherlands is 5.1% (Laitinen-Krispijn & Bijl, 2002) and the year prevalence of anxiety disorders in that population is 10.5%. That means that there is a high prevalence of these disorders in the working population and moreover, workers in the Netherlands with depressive or anxiety disorders are at greater risk to long-term sickness absence than workers without these disorders (Plaisier et al., 2010; Vlasveld et al., 2012).

Long-term sickness absence is responsible for enormous costs for patients, companies and society as a whole. The loss in productivity and the payments on disability benefits place a substantial burden on the economies of many developed countries (Henderson, Harvey, Overland, Mykletun & Hotopf, 2011). Furthermore, long-term sickness absence from work is called the major public health problem in the western world (Henderson, Glozier & Holland-Elliot, 2005). Because depression and anxiety disorders are highly associated with sickness absence, these disorders are not only responsible for personal suffering, but also for high societal costs (Smit et al., 2006). Total costs of sickness absence in the Netherlands as a consequence of psychological disorders are estimated at 2.7 billion Euros (De Graaf, Tuithof, Van Dorsselaer & Ten Have, 2011). Despite their high prevalence, depressive and anxiety disorders often remain unrecognized (Ellen, Norman & Burrows, 1997). However, because of the increased risk of long term sickness absence for workers with these disorders, it is important for occupational physicians to be able to recognize depression and anxiety and to start treatment in time. Therefore, the availability of good screening instruments for depression and anxiety among workers with long-term sickness absence is important. For the occupational health care practice, these instruments must be brief and easy to use and

besides that, also reliable and valid within the population of interest (Street, Gold & McDowell, 1994).

The Patient Health Questionnaire is a self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD). The PHQ-9 is a subscale of the PHQ and has proven to be a reliable and valid instrument for screening depression compared with other brief, validated and freely available instruments for screening major depressive disorder according to DSM-IV diagnosis in primary care (Kroenke, Spitzer & Williams, 2001). The self-report GAD-7 or the Generalized Anxiety Disorder Scale was developed for screening generalized anxiety disorder, according to DSM-IV diagnosis, but also functions in screening any anxiety disorder (Spitzer, Kroenke, Williams & Lowe, 2006). Several studies have reported good psychometric qualities of the PHQ-9 and the GAD-7 in primary care settings as well as in the general population of the Netherlands (Donker, Van Straten, Marks & Cuijpers, 2011). But among workers with long-term sickness absence other dynamics take place. Working is for many people an important aspect of daily life and absence of work is associated with social isolation or loss of daily routines (Plaisier et al., 2010). Besides that, sick-listed workers often have other physical disorders or conditions that may overlap with the psychological symptoms of depressive and anxiety disorders. This overlap may be cause of higher scores on the PHQ-9 and the GAD-7 in this population of sick-listed workers. It is therefore possible that to correctly identify major depressive disorder or generalized anxiety disorder in this population, higher cut-off values are necessary than in the general population.

The current study examines the reliability and the criterion validity of the Dutch versions of the PHQ-9 and the GAD-7 in a population of employees with long-term sickness absence. Therefore, DSM-IV diagnosis according to the Dutch version of the M.I.N.I. PLUS interview is used as the golden standard. The MINI-interview has in many studies been found to be a reliable and valid diagnostic interview to conduct within a short time frame (Sheehan et al., 1997; Lecrubier et al., 1997). The criterion validity of the PHQ-9 and the GAD-7 according to the M.I.N.I. PLUS will be analyzed in terms of sensitivity and specificity. Besides that, receiver operating characteristic analyses will be performed to look at the distinctive character of the tests.

2. METHODS

2.1 PARTICIPANTS

In this study, data were used that were collected in the recruitment phase from a randomized controlled trial. This ‘Return to Work’–study is a study into the cost effectiveness of an e-health module embedded in collaborative occupational health care for common mental disorders. Sick-listed workers (due to any cause) were recruited as participants in collaboration with the occupational health service. Thus, workers with only physical problems were also included. Excluded from participation were workers who were not sick-listed between 4 and 26 weeks, those who did not have sufficient command of the Dutch language and those who were pregnant.

2.2 PROCEDURE

Workers on sickness absence between 4 and 26 weeks received written information about the study from the occupational health service, together with an information leaflet from the Trimbos-institute, a first informed consent form and a screener. The screener contained the PHQ-9 and the GAD-7. The workers were asked if they were willing to participate in the study investigating sickness guidance for sick-listed workers with common mental disorders. If they agreed, they were asked to sign the informed consent form and to return it together with the completed screener to the researchers. In the information leaflet it was emphasized that participation in the study is voluntary and that refusal to participate would have no consequences for their (future) guidance and sickness certification. Workers who were willing to participate were then contacted by telephone within 30 days for a diagnostic interview, the Mini International Neuropsychiatric Interview (Sheehan et al., 1998). Interviewers were blinded for the results of the screener. At the end of the interview, participants who reached a cut-off value of 10 on one of the scales of the screener, were invited by the research assistant to be included in the 'Return to Work' – study.

2.3 MEASUREMENT INSTRUMENTS

2.3.1 Patient Health Questionnaire-9

The Patient Health Questionnaire (PHQ) is a self-administered version of the PRIME-MD diagnostic instrument for common mental disorders (Kroenke et al., 2001). The PHQ-9 is the scale for depression, which scores each of the 9 DSM-IV criteria from 0 to 3. A score of 0 indicates 'not at all' and a score of 3 indicates 'nearly every day'. As such, the

PHQ-9 score can range from 0 to 27. It is common use to diagnose MDD at a score of ≥10 (Kroenke et al., 2001).

2.3.2 Generalized Anxiety Disorder Scale-7

The Generalized Anxiety Disorder Scale-7 consists of 7 items that reflect most of the DSM-IV symptom criteria for generalized anxiety disorder (Spitzer et al., 2006). The questionnaire asks patients how often, during the last two weeks, they were bothered by these symptoms. Scoring options were 'not at all', 'several days', 'more than half the days' and 'nearly every day', scored as 0, 1, 2 and 3, respectively. Because each of the 7 items is scored from 0 to 3, the GAD-7 scale score ranges from 0 to 21. A score of ≥10 represents a reasonable cut-off value for identifying cases of GAD (Spitzer et al., 2006). The current study reviews the GAD-7 for screening any anxiety disorder.

2.3.3 MINI-International Neuropsychiatric Interview PLUS

The MINI-International Neuropsychiatric Interview (M.I.N.I.) is a short structured diagnostic interview, developed jointly by psychiatrists and clinicians, for diagnosis of the most common DSM-IV and ICD-10 psychiatric disorders (Sheehan et al, 1998). For the current study, a Dutch version of the interview was used (Van Vliet & De Beurs, 2007). The M.I.N.I.-Plus includes 23 disorders, but for the current study only the modules for depressive and anxiety disorders were used. All interviewers were trained in carrying out the interview and were able to consult the researchers when uncertain about a diagnosis.

2.4 STATISTICAL ANALYSIS

In the current study, internal consistency reliability (Cronbach's alpha) of the PHQ-9 and the GAD-7 for the population of sick-listed workers was calculated. Furthermore, the criterion validity of the PHQ-9 and the GAD-7 was analyzed in terms of sensitivity, specificity, positive and negative predictive value and the efficiency for different cut-off values of the PHQ-9 and the GAD-7 in detecting depressive and anxiety disorders (Bossuyt et al., 2003). Table 1 shows the calculations of the validity parameters.

Sensitivity refers to the probability that the test will detect the presence of the disorder in a person who actually suffers from the disorder, according to the MINI. It can also be explained as the number of true positives (TP) within all positive MINI-diagnoses (TP+FN). Specificity refers to the probability that the test will detect the absence of the disorder in a person who does not suffer from the disorder, according to

the MINI. It can also be explained as the number of true negatives (TN) within all negative MINI-diagnoses (FP+TN). The positive predictive value (PPV) is the probability of a positive diagnosis after a positive screening and the negative predictive value (NPV) is the probability of a negative diagnosis after a negative screening. The efficiency is the percentage of all the results from the screeners that correctly identified the positive and negative MINI-diagnoses.

Then, we assessed the ultimate cut-off value of the PHQ-9 and the GAD-7. A common method for deriving the best cut-off value is to find the cut-off value which generates the most optimal levels of sensitivity and specificity. However, there is no consensus on what levels of sensitivity and specificity are optimal, as they depend on the test's aim, costs and benefits. In the current study, it is decided that the ultimate cut-off value is where the sum of sensitivity and specificity reaches its maximum, striving for a minimum of .75 for both coefficients (Smits, Smit & Cuijpers, 2007).

Table 1. Calculations of sensitivity, specificity, positive and negative predictive values and efficiency

	MINI +	**MINI -**	**Total**
PHQ-9/GAD-7 +	TP	FP	TP+FP
PHQ-9/GAD-7 -	FN	TN	FN+TN
Total	TP+FN	FP+TN	TP+FP+FN+TN

Sensitivity = TP/(TP+FN) ; Specificity = TN/(FP+TN) ; Efficiency = (TP+TN)/(TP+FP+FN+TN) ;

Positive predictive value = TP/(TP+FP) ; Negative predictive value = TN/(FN+TN)

In the end, Receiver Operating Characteristic (ROC) analyses were performed which calculated an Area Under the Curve (AUC) for both the PHQ-9 and the GAD-7. The AUC can be interpreted as the distinctive character of the tests, or the probability that a randomly chosen participant would be correctly distinguished based on their screening score (Hanley & McNeil, 1983).

3. RESULTS

The mean age of the final study sample (N=170) was 45.40 years (SD=10.85). Age ranged from 21 to 66 years. The gender of the participants was divided equally (50.00%). The average number of days between completion of the screener and administration of the MINI was 13.65 days (SD=7.16). Figure 1 and 2 show flowcharts of the participants in this study for the PHQ-9 and the GAD-7, respectively.

Figure 1. Flowchart participants PHQ-9

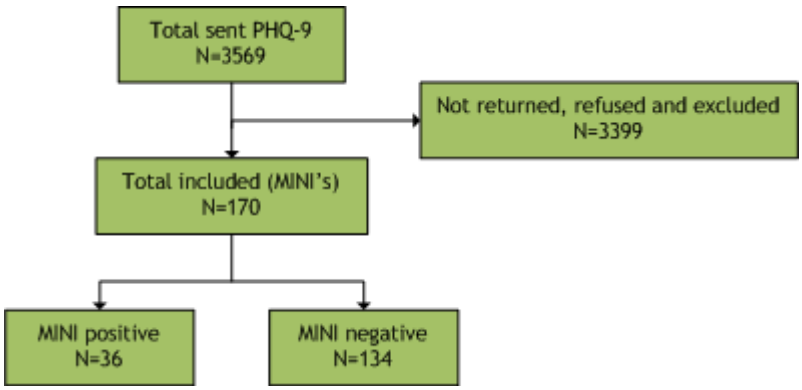
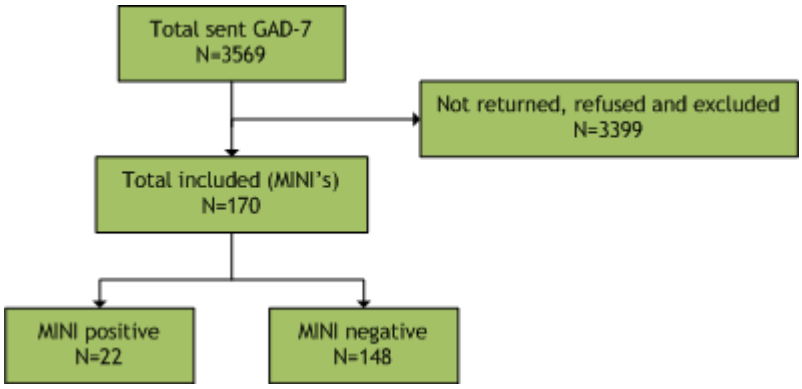


Figure 2. Flowchart participants GAD-7



The mean score on the PHQ-9 was 8.01 (SD=7.14) and the mean score on the GAD-7 was 6.09 (SD=6.04). From a total of 170 MINIs, 36 scored positive for MDD (prevalence of 21.18%) and 22 scored positive for anxiety (prevalence of 12.94%). The presence of MDD and anxiety according to a cut-off value of ≥10 are shown in Table 2 and 3.

Table 2. Presence of major depressive disorder according to PHQ-9, cut-off value of ≥10

	MINI positive MDD	**MINI negative MDD**	**Total**
PHQ-9 (>10)	31	29	60
PHQ-9 (≤10)	5	105	110
Total	36	134	170

Scale analysis showed that the internal consistency reliability (Cronbach's α) of the PHQ-9 in our sample was .92. The reliability of the GAD-7 in our sample was .94.

Table 3. Presence of anxiety disorders according to GAD-7, cut-off value of ≥10

	MINI positive anxiety	**MINI negative anxiety**	**Total**
GAD-7 (>10)	16	29	45
GAD-7 (≤10)	6	119	125
Total	22	148	170

Table 4 and 5 show the sensitivity, specificity, and predictive values for both positive and negative test results (PPV and NPV) for different cut-off values of the PHQ-9 and the GAD 7. These were calculated for all the possible cut-off values, but the ranges that are shown had the most optimal results and are therefore presented here.

Table 4. Validity outcomes for different cut-off values of the PHQ-9

	Score ≥ 9	**Score ≥ 10**	**Score ≥ 11**	**Score ≥ 12**
Sensitivity	.86	.86	.78	.75
Specificity	.75	.78	.82	.85
PPV	.48	.52	.54	.54
NPV	.95	.95	.93	.93
Efficiency	.62	.65	.69	.69

In the present study we look at the most optimal interaction between sensitivity and specificity for both the PHQ-9 and the GAD-7. Table 4 shows that a cut-off value of ≥10 results in the most optimal interaction, with a sensitivity of .86 and a specificity of .78. Predictive values for the test-positive results were .52, for the test-negative results .95 and an efficiency of .65. Table 5 shows that a cut-off value of ≥9 results in the most optimal interaction, with a sensitivity of .82 and a specificity of .76.

Table 5. Validity outcomes for different cut-off values of the GAD-7

	Score ≥ 8	**Score ≥ 9**	**Score ≥ 10**	**Score ≥ 11**
Sensitivity	.91	.82	.73	.73
Specificity	.74	.76	.80	.82
PPV	.34	.33	.36	.37
NPV	.98	.97	.95	.95
Efficiency	.76	.76	.79	.81

Predictive values for the test-positive results were .33, for the test-negative results .97 and an efficiency of .76.

The ROC curves, calculated for both the PHQ-9 and the GAD-7 are shown in figure 3 and 4. The calculated AUC for the PHQ-9 score versus the MINI was .90 (SE=0.02; CI=0.85-0.94). The calculated AUC for the GAD-7 score versus the MINI was .85 (SE=0.03; CI=0.79-0.92). These values are significant.

Figure 3. ROC-curve PHQ-9 versus MINI

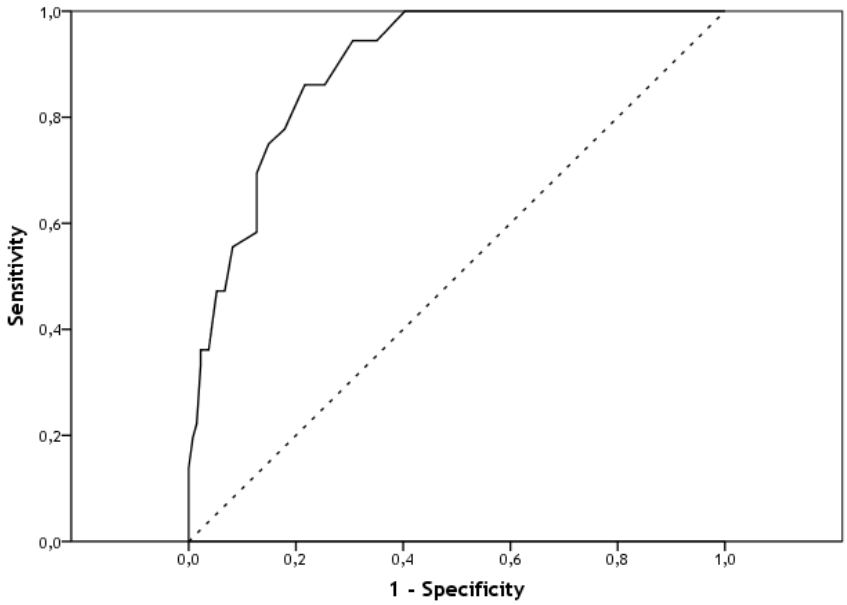
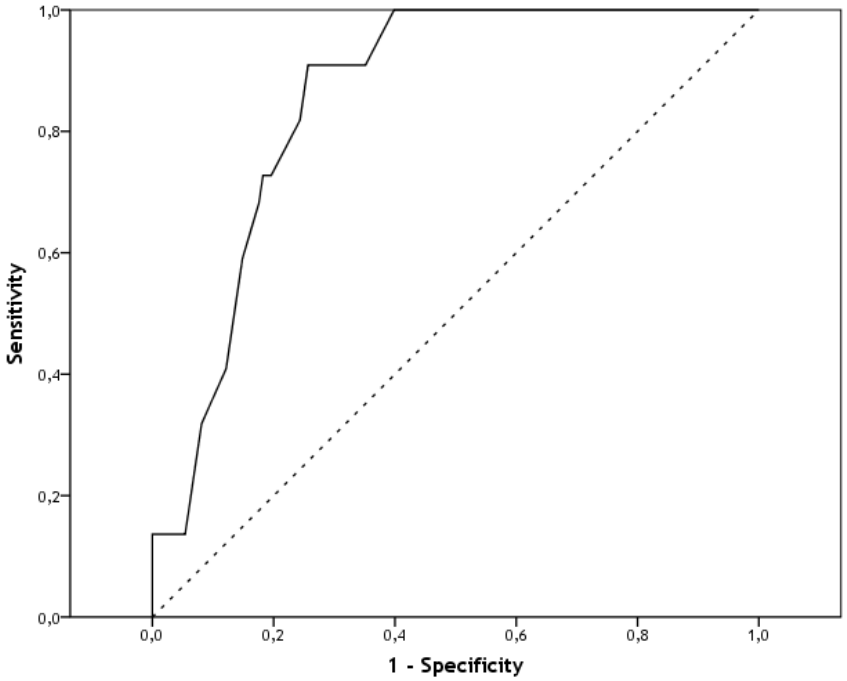


Figure 4. ROC-curve GAD-7 versus MINI



4. DISCUSSION

The current study evaluated the reliability and the criterion validity of the PHQ-9 and the GAD-7 in screening depressive and anxiety disorders among sick-listed workers.

Both the PHQ-9 and the GAD-7 proved to be reliable (Cronbach's α of .92 and .94) in the population of sick-listed workers in the Netherlands. With a cut-off value for the PHQ-9 and the GAD-7 of respectively 10 and 9, both tests also show good criterion validity characteristics. The PHQ-9 has a sensitivity of .86 and a specificity of .78. This means that 86% of sick-listed workers with major depressive disorder according to the MINI, will be detected as such and 78% of sick-listed workers without major depressive disorder will score negative on this screener. The GAD-7 has a sensitivity of .82 and a specificity of .76. This means that 82% of sick-listed workers with any anxiety disorder will be detected as such and 76% of sick-listed workers without any anxiety disorder will score negatively on this screener. The AUC refers to the distinctive character of the tests and is .90 for the PHQ-9 and .85 for the GAD-7. Receiver operating characteristic analyses and calculation of the AUC is often applied in comparisons of two tests that measure the same outcome. That is not the case in the current study, but nonetheless these are signs for valid tests (Smits et al., 2007).

A recent study suggested that the optimal cut-off value is determined by the decisions that are made based on the cut-off value, depending on the context in which the screening instrument is used (Smits et al., 2007). The current study was conducted in the context of the 'Return to Work'–study. Occupational physicians have to decide on the deployment of treatment. It is important for them to save costs by avoiding unnecessary treatment and to deploy treatment correctly for the workers that need it. The test needs to detect the presence of the disorder in workers who actually suffer from the disorder, but also needs to detect the absence of the disorder in a person who does not suffer from the disorder. For this reason, it is found that the optimal cut-off values is where the sum of sensitivity and specificity reaches its maximum, striving for a minimum of .75 for both coefficients. Based on this principle, the optimal cut-off value for the PHQ-9 is 10 and for the GAD-7 it is 9. These cut-off values do not deviate far from the cut-off values that are typically used (Kroenke et al., 2001; Spitzer et al., 2006). For the PHQ-9 it is the same en for the GAD-7 it differs 1 scoring point. That is not as was expected initially. But it is possible that this 1 point lower cut-off values is due to the fact that the GAD-7 is originally specified for the generalized anxiety disorder. The GAD-7 might be less sensitive for other anxiety disorders than the MINI. Other aims, costs and benefits of the decision- making based on these specific screeners in the population of sick-listed workers might result in other expectations of

the sensitivity and specificity. In that case, it is recommendable to look at the corresponding cut-off values.

The strengths of the current study are related to the fact that this was the first validation study within a population of sick-listed workers. Besides that, the interviewers were blinded for the results of the screener. The current study also had its limitations. Because of the high rate of non-response, refusals and exclusion of participants that could not be reached within 30 days for the MINI interview, a first limitation may be selection bias. A second limitation may be the amount of time between completion of the screener and the diagnostic interview. It is possible that the absence or presence of a major depressive disorder or any anxiety disorder at the time of completion of the screener does not match the results of the MINI interview at a later moment in time. Diagnosis based on a telephone MINI interview as the golden standard, may form a third limitation. While conducting the MINI-interview, it is common that more diagnoses are found than in other diagnostic interviews (Van Vliet & De Beurs, 2007). Systematically running through all diagnostic criteria ensures that no diagnoses are being overlooked. But it may also be cause of diagnosing too quickly. Because diagnostic interviews are not common in the general population, it is possible that diagnosing with this interview in comparison, affects the ultimate cut-off values of the screeners that were found. But in the end, this limitation or any other limitations concerning the golden standard are of course inherent in criterion validation studies.

Finally, it is noteworthy that the prevalence of major depressive disorder and anxiety in the group of participants does not differ much from the lifetime prevalence of major depressive disorder and anxiety in the general population of the Netherlands. In addition, standard deviations of 7.14 and 6.04 and internal consistency reliabilities of .92 and .94 suggest that participants scored quite different from each other but very consistent. In the present study, participants were sick-listed due to any cause. For future research, it would be interesting to investigate to what extent the differentiation of the group of participants based on the cause of the sick leave would lead to different results.

REFERENCES

Bijl, R. V., Ravelli, A., & Zessen, G. van (1998). Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology*, *33*, 587–595.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Moher, D., Rennie, D., Vet, H.C.W. de, & Lijmer, J.G. (2003). The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. *Clinical Chemistry*, *49*, 7-18.

Donker, T., Straten, A. van, Marks, I., & Cuijpers, P. (2011). Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Research*, *188*, 58-64.

Ellen, S. R., Norman, T. R., & Burrows, G. D. (1997). MJA Practice Essentials - Mental Health. Assessing anxiety and depression in primary care. *Medical Journal of Australia*, *167*, 328-333.

Graaf, R. de, Tuithof, M., Dorsselaer, S. van, & Have, M. ten (2011). *Verzuim door psychische en somatisch aandoeningen bij werkenden. Resultaten van de 'Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2)*. Utrecht: Trimbos-instituut.

Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*, 839-843.

Henderson, M., Harvey, S. B., Overland, S., Mykletun, A., & Hotopf, M. (2011). Work and common psychiatric disorders. *Journal of the Royal Society of Medicine*, *104*, 198–207.

Henderson, M., Glozier, N., & Holland-Elliot, K. (2005). Long term sickness absence is caused by common conditions and needs managing. *British Medical Journal*, *330*, 802–803.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine, 16, 606-613.

Laitinen-Krispijn, S. & Bijl, R. (2002). *Werk, psyche en ziekteverzuim. Aard en omvang van psychische stoornissen, ziekteverzuim en zorggebruik in de beroepsbevolking*. Utrecht: Trimbos-instituut.

Lecrubier, Y., Sheehan, D.V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI): A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry, 12*, 224-231.

Plaisier, I., Beekman, A. T. F., Graaf, R. de, Smit, J. H., Dyck, R. van, & Penninx, B. W. J. H. (2010). Work functioning in persons with depressive and anxiety disorders: the role of specific psychopathological characteristics. *Journal of Affective Disorders, 125*, 198–206.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G.C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59* (*Suppl. 20*), 22-33.

Sheehan, D.V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M. F., & Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry, 12*, 232-241.

Smit, F., Cuijpers, P., Oostenbrink, J., Batelaan, N., de Graaf, R., & Beekman, A. (2006). Costs of nine common mental disorders: implications for curative and preventive psychiatry. *Journal of Mental Health Policy and Economics, 9*, 193-200.

Smits, N., Smit, F., & Cuijpers, P. (2007). Using decision theory to derive optimal cut-off scores of screening instruments: an illustration explicating costs and benefits of mental health screening. *International Journal of Methods in Psychiatric Research, 16*, 219-229.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*, 1092-1097.

Street, R. L., Gold, W. R., & McDowell, T. (1994). Using health status surveys in medical consultations. *Medical Care, 32*, 732-744.

Terluin, B., Rhenen, W. van, Anema, J. R., & Taris, T. W. (2011). Psychological symptoms and subsequent sickness absence. *International Archives of Occupational and Environmental Health*, *84*, 825–837.

Vlasveld, M.C., Feltz-Cornelis, C.M. van der, Bültmann, U., Beekman, A.T.F., Mechelen, W. van, Hoedeman, R., & Anema, J.R. (2012). Predicting Return to Work in Workers with All-Cause Sickness Absence Greater than 4 Weeks: A Prospective Cohort Study. *Journal of Occupational Rehabilitation*, *22*, 118-126.

Vliet, I.M. van, & Beurs, E. de (2007). Het Mini Internationaal Neuropsychiatrisch Interview (MINI): een kort gestructureerd diagnostisch psychiatrisch interview voor DSM-IV en ICD-10-stoornissen. *Tijdschrift voor Psychiatrie*, *49*, 393-397.