Are men more depressed than women in Norway?
Validity of the Hospital Anxiety and Depression Scale

Monica W. Nortvedt\textsuperscript{a,*}, Trond Riise\textsuperscript{b}, Bjarte Sanne\textsuperscript{c}

\textsuperscript{a}Department of Research and Development, Faculty of Health and Social Science, Bergen University College, Norway
\textsuperscript{b}Section for Occupational Medicine, University of Bergen, Norway
\textsuperscript{c}Centre for Child and Adolescent Mental Health, University of Bergen, Norway

Received 17 November 2004; received in revised form 5 July 2005; accepted 5 July 2005

Abstract

\textbf{Objective and methods:} The aim was to evaluate the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) by comparing the gender-specific scores with another self-rated measure of mental health (MH), the SF-12 Health Survey, in a large general population (\textit{N}=16,116). Results: Using a score of 8 as a cut-off point for depression, the odds ratio for depression among men versus women was 1.29 [\textit{P}<.0001; 95\% confidence interval (CI): 1.16–1.43]. This contrasted with the results from the SF-12, in which the women had markedly and significantly poorer scores than men did. The women also reported a higher use of medicine for depression. Two of the seven HADS-D questions, related to interest in personal appearance and the ability to enjoy television, radio, and books, explained 70\% of the higher depression scores among men. Conclusion: The results suggest that the validity related to gender differences in HADS-D is highly questionable.

\textcopyright 2006 Published by Elsevier Inc.

**Keywords:** Depression; Gender; Hospital Anxiety and Depression Scale; SF-12 Health Survey

Introduction

Population-based studies show rather consistently that the symptom load and prevalence of depression are higher in women than in men [1–4]. The self-rated depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) is widely used in population-based surveys [5,6]. The studies reporting gender differences using the HADS-D, however, show unclear results. Some studies report no difference or less depression among women [7–9], whereas one study reports more depression among women [10]. This study aimed to evaluate the HADS-D by comparing the gender-specific scores with those from another self-rated measure of mental health (MH), the SF-12 Health Survey, and the use of medicine for depression in a large general population.

Methods

\textit{Study population}

The Hordaland Health Study 1997–1999 (HUSK) was conducted as a collaboration among the National Health Screening Service, the University of Bergen, and local health services [11]. The study population included all 29,400 individuals born between 1953 and 1957 and a random sample of those born in 1950 and 1951 who resided in Hordaland County on December 31, 1997. A total of 10,263 men and 12,054 women participated, with a participation rate of 65\%. This gave a total study population of 22,317, with a mean age of 42.8 years (range, 40–47) at the time of the study. The study design was cross-sectional and included questionnaires and a clinical examination carried out from October 1997 to June 1999. A first questionnaire including the SF-12 was sent to the participants by mail before the clinical examination and collected at the examination. At the clinical examination, a second
questionnaire, including the HADS questions and a question on the use of medicine for depression, was distributed and returned by prepaid mail. A total of 16,116 individuals (72%) responded to this second questionnaire.

**Measurements**

Depression was measured using HADS-D [12,13], the mental summary score of the SF-12 [14], and the two items on MH in the SF-12.

The HADS [13] is a 14-item, self-administered questionnaire: 7 items constitute the anxiety scale (HADS-A) and 7 constitute the depression scale (HADS-D). HADS-D especially covers anhedonia (reduced pleasure response). In addition, items on psychomotor retardation and depressed mood are included. Each item is scored on a Likert scale from 0 to 3, and the item scores are added, giving subscale scores from 0 (minimum symptom level) to 21 (maximum symptom level) [5]. Missing substitution was performed for individuals who had responded to five or six of the seven HADS-D questions. This was done by multiplying the obtained score by 7/5 if five of the seven questions were answered and by 7/6 if six questions were answered. Such missing substitution was done for 12.3% of the individuals for the HADS-D scale. A total of 59 individuals had answered between one and five questions and were therefore excluded. Caseness is usually defined by a score of 8 or above on HADS-D. This cut-off level has been shown to excluded. Caseness is usually defined by a score of 8 or above on HADS-D. This cut-off level has been shown to be a Gender difference measured in standard deviations. Negative values indicate poorer mental health among women.

**Table 1**

Mean HADS scores according to gender in a general population of 7295 men and 8821 women

<table>
<thead>
<tr>
<th></th>
<th>HADS anxiety scale</th>
<th>HADS depression scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean score CI</td>
<td>Percentage (%) ≥8</td>
</tr>
<tr>
<td>Men</td>
<td>4.45 4.37–4.52</td>
<td>15.7</td>
</tr>
<tr>
<td>Women</td>
<td>4.90 4.83–4.97</td>
<td>20.1</td>
</tr>
<tr>
<td>Differencea</td>
<td>−0.14</td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001 (&lt;t test&gt;)</td>
<td></td>
</tr>
</tbody>
</table>

Population. Furthermore, 2 of the 12 questions directly measure MH, and the results for each of these are reported. One question (SF-12 MH1) asks: “How much of the time during the past four weeks have you felt calm and peaceful?” with answers ranging from “none of the time” to “all the time”, corresponding to scores 1 to 6. The other question (SF-12 MH2) asks “How much of the time during the past four weeks have you felt downhearted and blue?” with the same range of answers such that a score of 1 corresponds to low MH and a score of 6 corresponds to best score of MH. The mental scale of the full SF-36 has been used for the screening for clinical depression [18], and validation of the SF-12 has shown that it reflects most of the information (95%) of the SF-36 [14,17]. Nevertheless, SF-12 cannot be considered as a gold standard for measuring clinical depression.

The study also included a question on use of medicine for depression: “During the past year, have you used medicine for depression daily or almost daily?”

Only individuals who had answered both the HADS-D and the SF-12 questions were included in the analysis.

**Results**

A score for the HADS-D scale was calculated for 15,910 of the 16,116 individuals (98.7%) who sent in the second questionnaire. The mean score for the HADS-D was 3.56 [95% confidence interval (CI): 3.49–3.63] for men and 3.05 (95% CI: 2.99–3.11) for women (Table 1). This constitutes a difference of 0.16 S.D. The percentages with scores 8 and above were 11.0% for men and 8.8% for population. Furthermore, 2 of the 12 questions directly measure MH, and the results for each of these are reported. One question (SF-12 MH1) asks: “How much of the time during the past four weeks have you felt calm and peaceful?” with answers ranging from “none of the time” to “all the time”, corresponding to scores 1 to 6. The other question (SF-12 MH2) asks “How much of the time during the past four weeks have you felt downhearted and blue?” with the same range of answers such that a score of 1 corresponds to low MH and a score of 6 corresponds to best score of MH. The mental scale of the full SF-36 has been used for the screening for clinical depression [18], and validation of the SF-12 has shown that it reflects most of the information (95%) of the SF-36 [14,17]. Nevertheless, SF-12 cannot be considered as a gold standard for measuring clinical depression.

The study also included a question on use of medicine for depression: “During the past year, have you used medicine for depression daily or almost daily?”

Only individuals who had answered both the HADS-D and the SF-12 questions were included in the analysis.

**Table 2**

Mean scores of the SF-12 MCS and of the two separate SF-12 MH questions, MH1 and MH2, according to gender in a general population of 7295 men and 8821 women

<table>
<thead>
<tr>
<th></th>
<th>SF-12 MH1</th>
<th>SF-12 MH2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>Men</td>
<td>50.8 50.6–51.0</td>
<td>4.65 4.63–4.68</td>
</tr>
<tr>
<td>Women</td>
<td>49.3 49.1–49.5</td>
<td>4.44 4.41–4.46</td>
</tr>
<tr>
<td>Differencea</td>
<td>−0.15</td>
<td>−0.23</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

* Gender difference measured in standard deviations. Negative values indicate poorer mental health among women.

**Table 3**

Mean scores of each of the seven HADS-D items according to gender

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enjoy</td>
<td>Laugh</td>
<td>Cheerful</td>
<td>Slow</td>
<td>Appearance</td>
<td>Forward</td>
<td>Radio</td>
</tr>
<tr>
<td>Men</td>
<td>0.51</td>
<td>0.28</td>
<td>0.64</td>
<td>0.56</td>
<td>0.46</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>Women</td>
<td>0.41</td>
<td>0.28</td>
<td>0.60</td>
<td>0.66</td>
<td>0.27</td>
<td>0.42</td>
<td>0.38</td>
</tr>
<tr>
<td>Difference</td>
<td>0.11</td>
<td>0.00</td>
<td>0.04</td>
<td>−0.09</td>
<td>0.19</td>
<td>0.10</td>
<td>0.17</td>
</tr>
</tbody>
</table>
women. The odds ratio for depression among men versus women was 1.29 (95% CI: 1.16–1.43). Using scores of 11 and above as the cut-off point, the corresponding odds ratio was 1.24 (95% CI: 1.03–1.49).

This finding contrasts with the results from the SF-12 MCS and from each MH item of the SF-12, on which the women had markedly and significantly poorer scores than men did (Table 2). Women’s MCS was 0.15 S.D. lower than that of men and 0.23 and 0.19 S.D. lower on the two specific MH questions.

Women had poorer scores than men did on only one of the HADS-D items: D4 (“I feel as if I am slowed down”; Table 3). For all other items, the men’s scores were the same or poorer. The greatest difference was found for item D5 (“I have lost interest in my appearance”) and D7 (“I can enjoy a good book or radio or TV program”). These two questions explained about 70% of the total difference in the summary score. These questions also showed the lowest correlation with the two MH items of the SF-12, about .20, whereas the other five items had correlation coefficients from .32 to .57 (Table 4). Item D5 did not even have a linear relationship: The mean score of the SF-12 items for those with the lowest score on the D5 (“definitely”) was higher than the mean score for those who scored second lowest (“I don’t take so much care as I should”).

Of the 15,910 study participants who had responded to HADS-D, 209 men (3.5%) and 497 women (6.7%) had used medicine for depression. The mean HADS-D score for those who had used medicine for depression was 6.1 and 3.1, respectively, for those who had not used medicine for depression.

### Discussion

The women in this large population-based study of people aged 40–47 years reported less depression than men did on the HADS-D. The men had a 29% higher risk of probable depression using a cut-off point of 8 and above. This finding contrasts with the results from the MCS and the two MH items of the SF-12, on which the women reported significantly poorer MH. Women also reported a significantly higher use of medicine for depression (odds ratio=1.98), which is compatible with a higher rate of depression among women. Furthermore, the results using the HADS-D contradict those of population-based surveys using other measures of depression, on which women consistently report higher rates of depression.

The HUNT study, a large population-based study in Norway using the HADS-D, also found less depression among women [7]. For the age group 40–49 years, the odds ratio for women versus men was 0.88 (95% CI: 0.78–0.98). The corresponding odds ratio in HUSK was similar: 0.78 (95% CI: 0.70–0.86). The authors suggested that this rather unexpected finding could reflect the fact that the HADS-D excludes questions related to “somatic” depression [7]. They refer to Silverstein [19], who found that women in the National Comorbidity Survey exhibited a higher prevalence of “somatic” depression (depression including fatigue and appetite and sleep disturbance) than men did but not a higher prevalence of “pure” depression. According to this, the item in the SF-12 that also measures pure depression, MH2 (“How much of the time during the past four weeks have you felt downhearted and blue?”), would have shown similar results as the HADS-D. This was not found but does not automatically argue against the possibility that at least a part of higher depression scores in women could be explained by the lack of vegetative symptoms in the HADS-D.

The gender difference in depression is known to become more pronounced with increasing severity [20,21]. To evaluate whether our findings reflect the possibility that HADS-D mainly differentiates in the milder spectrum, we calculated the gender differences also using 11 as a cut-off point, thereby including only individuals with “probable” depression as cases. The odds ratio using this cut-off point was 1.24, only slightly lower than the 1.29 using 8 as the cut-off point, showing that this cannot explain the unexpected gender difference found using HADS-D in this study.

Because both studies finding significantly lower depression rates among women (HUSK and HUNT) used the Norwegian translation of the HADS, the translation could be a possible explanation. Nevertheless, a study in Sweden found a gender difference for the HADS-D scale of 0.51 points, the same difference as in the present study [11]. The difference measured in standard deviation was 0.15 in Sweden versus 0.16 in our study. Nevertheless, the difference in Sweden was not statistically significant because the study population was smaller. Furthermore, studies of 5338 people in Germany [13] and 6165 people in the Netherlands [22] using HADS reported no increased prevalence of depression among women. In addition to these findings, the results showing poorer scores among women using (1) the MCS of the SF-12, (2) the single item MH2 measuring pure depression, and (3) the use of medicine for depression argues against another hypothesis that the results could reflect cultural gender differences.

Items D5 and D7 in HADS-D account for most (70%) of the gender difference. In addition, these questions showed the lowest correlation with the other scales of MH. The mean scores at the four levels of these questions did not follow the expected increasing trend when compared with the other measures. This suggests that these questions measure depression poorly or at least that their validity related to gender differences is highly questionable.
The study in Sweden examining the factor structure of the HADS found the appearance item only in the depression factor for women and not in the depression factor for men [9]. This supports the hypothesis that this item does not measure depression in men and that many men have a high score on this item without being depressed. Being less concerned about their appearance is not necessarily a sign of depression in men. The higher HADS-D among men in the present study could therefore partly reflect this.

The HADS has been found to perform well in assessing symptom load and possible cases of anxiety and depressive disorders [6]. However, we question the results of the present study showing that men are more depressed than women are in Norway. We argue that the HADS-D overestimates the prevalence of depression among men and/or underestimates it among women. This bias is mainly associated with the two items in the questionnaire related to interest in personal appearance and the ability to enjoy television, radio, and books.

References