A psychometric validation of the Hospital Anxiety and Depression Scale (HADS) in community-dwelling older adults.

The HUNT study, Norway.

Heidi Emly Sivertsen1, Anne-Sofie Helvik1,2, Linda Gjøra2,3, Gørill Haugan1,4

Author details

1 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

2 The Norwegian National Centre for Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway.

3 Department of Psychiatry, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway.

4 Nord University, Faculty of Nursing and Health Sciences, Levanger, Norway.

Corresponding author

Heidi Emly Sivertsen, Sjølystvegen 39 C, 7562 Saksvik, Norway.

Telephone: +47 93600836 E-mail: heidi.sivertsen@ntnu.no

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Objectives

The Hospital Anxiety and Depression scale (HADS) is commonly used to measure anxiety and depression, but the number of studies validating psychometric properties in older adults are limited. To our knowledge, no previous studies have utilized confirmative factor analyses in community-dwelling older adults, regardless of health conditions. Thus, this study aimed to examine the psychometric properties of HADS in older adults living at home in a large Norwegian city.

Methods

In total, 1190 inhabitants ≥70 (range 70 – 96) years completed the HADS inventory in the population-based Trøndelag Health Study (HUNT); termed “HUNT4 70+” in Trondheim, Norway. Confirmatory factor analyses (CFA) were performed to test the dimensionality, reliability, and construct validity.

Results

The original two-factor-solution (Model-1) revealed only partly a good fit to the present data; however, including a cross-loading for item 6D (“I feel cheerful”) along with a correlated error term between item 2D (“I still enjoy the things I used to enjoy”) and 12D (“I look forward with enjoyment to things”), Model-3 with a two-factor solution demonstrated an acceptable/good fit. Good to acceptable measurement reliability was demonstrated, and construct validity was supported.
Conclusions

The HADS involves some items which are not reliable and valid indicators for the depression construct in this population. Especially item 6 which is problematic. To improve the reliability and validity of HADS rewording of some items is recommended.

Keywords:
Anxiety, Community-dwelling, Confirmatory factor analysis, Construct validity, Depression, Dimensionality, HADS, HUNT, Older adults, Psychometric, Reliability.
Introduction

Older adults (65+) living at home may have multiple diseases, impaired physical function (1-6) and impaired cognitive function (6-8). Compared with younger age groups, less characteristic symptoms of depression often appear in old ages, and depression may be detected and diagnosed less frequently among older adults (3, 8-12). In addition, many community-dwelling older adults who meet diagnostic criteria for depression do not seek health care for their symptoms; less than half have contact with the health service and barely 10% receive effective treatment (8). However, major depressive disorder (MDD) and clinically significant depressive symptoms (CSDS) are associated with decreased quality of life (QOL), increased comorbidity with physical illness, reduced physical functioning in daily activities, increased risk of dementia (13) and increased need for help and risk of mortality (7, 10, 11, 14).

Depression is prevalent among older adults worldwide with an estimate of depressive disorders of 5.4%, including MDD and CSDS (aged >70 years) (15). Furthermore, a recent review based on 20 studies found the pooled prevalence of MDD to be 13.3% (16). The pooled prevalence of CSDS has been reported to be higher than for MDD. Two recent systematic reviews, largely based on studies of community-dwelling older adults, found the pooled prevalence to be 28.4% and 31.7%, respectively. Moreover, these two reviews found a large variety of the prevalence rates ranging between 7.7 – 81.1%; the authors explained this variety
by cultural differences, along with differences in sample characteristics, methodology, and screening tools (17, 18).

Depression often co-exists with anxiety and anxiety symptoms. Compared to younger age groups, older adults report a higher frequency of anxiety (8, 19). In 2019, the estimated prevalence of anxiety disorders (a combined estimate of all subtypes) among adults aged >70 years was 4.4% (15). A recent review among community-dwelling adults > 55 years old revealed a pooled prevalence of anxiety disorders and anxiety symptoms of 5.4% and 7.9%, respectively (20). Generalized anxiety disorder is the most prevalent anxiety disorder and leads to high mortality risk, which is even higher when accompanied by depression (21).

Similar to MDD and CSDS and compared to younger age groups, older adults report less characteristic symptoms of anxiety disorders and anxiety symptoms (20); for instance are worries about health, disturbances of sleep and less reassurance seeking behaviors common symptoms of anxiety among older adults (19). Consequently, anxiety disorder and anxiety symptoms may be detected and diagnosed less frequently among older adults. It is also known that older adults are less likely to seek health care for their symptoms than younger adults, because older adults are less likely to report their symptoms, have less knowledge regarding anxiety disorders and available treatments, with barriers to treatment seeking which include stigma, cost, transportation, and mobility (22).
Another explanation for this could be that common instruments are developed for younger age groups and might be less suitable for older adults (20).

The Hospital Anxiety and Depression Scale (HADS) is a screening tool commonly used in epidemiological research to estimate the prevalence of anxiety symptoms and CSDS among older adults (17). Originally, the HADS aimed to measure states of anxiety and depression in adults receiving treatment for physical health problems in general hospitals (23). HADS consists of 14 items and includes subscales for anxiety (HADS-A; seven items) and depression (HADS-D; seven items). The items are scored on a four-point scale ranging from totally disagree to totally agree. Each item is rated from 0-3, where higher scores indicate more anxiety and/or depression. The maximum score on each subscale is 21 ranging from 0-7 (normal), 8-10 (mild disorder), 11-14 (moderate disorder), to 15-21 (severe disorder) (23). HADS has been widely used (24, 25), suggested with a lower cut-off for older adults (26) and tested with satisfactory psychometric properties (24). However, to a less extent among older adults.

To increase acceptability and avoid individuals feeling as though they are being tested for mental disorders, symptoms of severe psychopathology have been excluded, this makes HADS more sensitive to milder psychopathology (27, 28). According to the International Classification of Diseases (ICD-10) five of seven items in HADS-D focus on lack of positive feelings and cover only two of three
main criteria for depression; physical symptoms as loss of energy, sleep- and
appetite disturbances are not covered.

The Norwegian version of the HADS is psychometrically evaluated among
older adults without cognitive impairment in nursing homes (28), older adults
admitted to somatic hospitals (29), older community-dwelling adults (30) and
among adults consulting the general practice (31). However, with one exception
(28) Norwegian studies have evaluated the psychometrical features of HADS by
means of principal component analysis (PCA) (27, 29, 31) without utilizing the
confirmative approach. Internationally, a few studies have studied features of
HADS using confirmative factor analysis (CFA) in community-dwelling, non-clinical
samples of older adults (24, 32, 33). These studies give support to the original two-
factor structure showing good internal consistency. Even so, problems with cross-
loadings and ceiling effect are reported (24). Utilizing CFA a psychometrical
evaluation of HADS in a clinical sample of veterans reported a three-factor
structure to fit better than the original two-factor solution (34).

Up to date, the factor structure, the internal consistency, construct validity
and homogeneity of the HADS have not been assessed by means of CFA among
community-dwelling older adults ≥70 years in Norway. Therefore, the aim of the
present study was to examine the psychometric properties of the HADS scale
among community-dwelling older adults in Norway by investigating its
dimensionality, reliability, and construct validity. Dimensionality concerns the
homogeneity of a scale's items (35), indicating if the included items match the defined construct. Depression and anxiety have been seen to correlate strongly but are still considered different constructs. Reliability encompasses a scale's consistency and lack of error (36). To assess internal consistence of the items the reliability coefficients of Cronbach's alpha ($\alpha$) and composite reliability ($\rho_c$) were utilized. Finally, construct validity implies various aspects such as convergent, discriminant and content validity. In this study, convergent and discriminant validity denotes if HADS relates with other constructs as expected, while content validity embraces whether the theoretical content of the Anxiety and Depression constructs involved in HADS are adequately represented by the 14 items. This is, whether the included items cover the theoretical definition they are aimed to represent (37). If the wording of items is too similar, Cronbach's alpha, content validity, and dimensionality will be falsely improved. In consequence, the average correlation among items increases, and therefore also coefficient alpha; however, without adding substantially to the content validity of the scale. Obviously, to tap into the same construct some similarity/correlation is needed. Nevertheless, items simply representing a rephrasing of other items are redundant.

Methods

Sample and procedures
During 2017-2019 persons aged 70+ were recruited from the city of Trondheim, county of Trøndelag in Central Norway as a part of the fourth wave of The Trøndelag Health Study (HUNT). HUNT is a population-based cross-sectional study (38). The “HUNT 4 70+ Study” includes the entire population in the county; 9930 (64%) consented to participate (39). The HUNT study comprises questionnaires, clinical measurements, and collections of biological samples.

Totally, 5087 inhabitants in the city of Trondheim were invited to participate. The inclusion criteria were age of ≥ 70 years and permanent residency in two randomly selected districts of the city. Out of the 5078, a total of 1747 (response rate of 34.3%) individuals participated (57.9% women and 42.1% men). The participants underwent a comprehensive clinical evaluation at the field-station and received a visit at home (some lived in a nursing home). In addition, they filled in two questionnaires (Q1 and Q2), either at the field station or at home. The questionnaire was sent to the field station by mail. HADS was included in Q2.

In this study, the inclusion criterion was to be community-living, while the exclusion criteria included: (1) lack of responses in HADS (n=392); (2) nursing home residents (n=65); and (3) lack of other essential health variables (n=100 participants). In total, 557 were excluded leaving an effective sample comprising 1190 participants.

Statistical analysis
The analysis were conducted using the IBM Statistical Package for the Social Sciences (SPSS) Version 28 software (40) and Stata 17 software package (41). Confirmatory Factor Analysis (CFA) represents a more accurate evaluation of the psychometric properties of scales used. In this study, the model fit adequacy was assessed by $\chi^2$-statistics and conventional fit indices: $\chi^2$-statistics, the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) with values <0.10 are acceptable and values $\leq$0.05 indicates a good fit (42, 43). Further, the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) with acceptable fit at 0.95 and good fit at 0.97 (42-45). Skewness and kurtosis were significant. Therefore, the Satorra-Bentler corrected $\chi^2$ which is the correct asymptotic mean even under non-normality is reported (46).

Results

Sample characteristics

The sample of 1190 adults were aged between 70-96 years with a mean of 76.5 (SD = 5.3); 644 (54.1%) were women (mean age 76.7) while 546 (45.9%) were men (mean age 76.2) (Table 1). Furthermore, 325 (27.3%) had completed higher education (≥4 years of university/college). About 50% had physical or mental long-term illness, injury, or a disorder that impaired their daily functioning (47.1%), only a few had any kind of in-home care (4.7 %), nursing care (4.5 %) and/or had been admitted to a nursing home for a period during the last year (3.3%).
Looking at those who were excluded, these were significantly older (mean age 83.3; SD=8.0 years), more often female (65.7%) and less high education (≥4 years of university/college; 7.2 %). Dementia diagnosis was also more prevalent among the excluded participants (56.2% versus 6.3%).

**HADS item score statistics**

The mean anxiety and depression scores were 3.4 (SD = 2.9) and 3.0 (SD = 2.5), respectively (Table 2.) The internal consistence of the anxiety and depression constructs (Table 2) was good ($\alpha_{\text{anxiety}}=.79=)$ or acceptable ($\alpha_{\text{depression}}=.66$).

Composite reliability ($\rho_c$) displayed values between 0.65-0.92 (Table 3); values $\geq 0.60$ are acceptable, whereas values $\geq 0.70$ are good (42, 47).

**Table 1.** Demographic characteristics of the study population by gender.

<table>
<thead>
<tr>
<th></th>
<th>Total (N=1190)</th>
<th>Women (n = 644, 54%)</th>
<th>Men (n=546, 46%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, M (SD)</td>
<td>76.5 (5.3)</td>
<td>76.7 (5.4)</td>
<td>76.2 (5.1)</td>
<td>0.004$^a$</td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>709 (59.5)</td>
<td>299 (25.2)</td>
<td>410 (34.5)</td>
<td>&lt; 0.001$^b$</td>
</tr>
<tr>
<td>Education n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>115 (9.7)</td>
<td>87 (13.5)</td>
<td>28 (5.1)</td>
<td></td>
</tr>
<tr>
<td>High School 1-2 years</td>
<td>190 (16)</td>
<td>138 (21.4)</td>
<td>52 (9.5)</td>
<td></td>
</tr>
<tr>
<td>High School 3 years</td>
<td>125 (10.5)</td>
<td>77 (12)</td>
<td>48 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Apprenticeship certificate</td>
<td>132 (11.1)</td>
<td>56 (8.7)</td>
<td>76 (13.9)</td>
<td></td>
</tr>
<tr>
<td>University/college &lt; 4 years</td>
<td>303 (25.5)</td>
<td>148 (23)</td>
<td>155 (28.4)</td>
<td></td>
</tr>
<tr>
<td>University/college &gt; 4 years</td>
<td>325 (27.3)</td>
<td>138 (21.4)</td>
<td>187 (34.2)</td>
<td></td>
</tr>
<tr>
<td>Morbidity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>125 (10.5)</td>
<td>84 (13)</td>
<td>41 (7.5)</td>
<td>0.001$^b$</td>
</tr>
<tr>
<td>Diabetes</td>
<td>92 (7.7)</td>
<td>39 (6.1)</td>
<td>53 (9.7)</td>
<td>&lt; 0.005$^b$</td>
</tr>
<tr>
<td>Dementia</td>
<td>75 (6.3)</td>
<td>39 (6.1)</td>
<td>36 (6.5)</td>
<td>n.s.</td>
</tr>
<tr>
<td>Heart attack</td>
<td>102 (8.6)</td>
<td>36 (5.6)</td>
<td>66 (12.1)</td>
<td>&lt; 0.001$^b$</td>
</tr>
<tr>
<td>Impaired function in daily life &gt; one year</td>
<td>560 (47.1)</td>
<td>325 (58)</td>
<td>235 (42)</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>Home care and hospitalized, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home care last year</td>
<td>56 (4.7)</td>
<td>37 (5.7)</td>
<td>19 (3.5)</td>
<td>n.s.$^b$</td>
</tr>
<tr>
<td>Nursing care last year</td>
<td>53 (4.5)</td>
<td>35 (5.4)</td>
<td>18 (3.3)</td>
<td>n.s.$^b$</td>
</tr>
</tbody>
</table>
Table 2. Means, Standard deviation (SD) and Cronbach's alpha for the Norwegian version of the Hospital Anxiety and Depression Scale (HADS).

<table>
<thead>
<tr>
<th>Items</th>
<th>Response *</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Cronbach's alpha (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A I feel tense or 'wound up'</td>
<td>911</td>
<td>245</td>
<td>28</td>
<td>6</td>
<td>1190 .27 .52</td>
</tr>
<tr>
<td>2D I still enjoy the things I used to enjoy</td>
<td>704</td>
<td>450</td>
<td>28</td>
<td>8</td>
<td>1190 .45 .58</td>
</tr>
<tr>
<td>3A I get a sort of frightened feeling as if something awful is about to happen</td>
<td>641</td>
<td>370</td>
<td>130</td>
<td>49</td>
<td>1190 .65 .83</td>
</tr>
<tr>
<td>4D I can laugh and see the funny side of things</td>
<td>924</td>
<td>236</td>
<td>27</td>
<td>3</td>
<td>1190 .25 .50</td>
</tr>
<tr>
<td>5A Worrying thoughts go through my mind</td>
<td>778</td>
<td>299</td>
<td>90</td>
<td>23</td>
<td>1190 .46 .72</td>
</tr>
<tr>
<td>6D I feel cheerful*</td>
<td>828</td>
<td>302</td>
<td>56</td>
<td>4</td>
<td>1190 .36 .59</td>
</tr>
<tr>
<td>7A I can sit at ease and feel relaxed</td>
<td>661</td>
<td>480</td>
<td>48</td>
<td>1</td>
<td>1190 .49 .58</td>
</tr>
<tr>
<td>8D I feel as if I'm slowed down</td>
<td>382</td>
<td>668</td>
<td>99</td>
<td>41</td>
<td>1190 .83 .72</td>
</tr>
<tr>
<td>9A I get a sort of frightened feeling like 'butterflies' in the stomach</td>
<td>702</td>
<td>464</td>
<td>21</td>
<td>3</td>
<td>1190 .43 .54</td>
</tr>
<tr>
<td>10D I have lost interest in my appearance</td>
<td>860</td>
<td>255</td>
<td>50</td>
<td>25</td>
<td>1190 .36 .66</td>
</tr>
<tr>
<td>11A I feel restless as if I must be on the move*</td>
<td>451</td>
<td>568</td>
<td>152</td>
<td>19</td>
<td>1190 .78 .72</td>
</tr>
<tr>
<td>12D I look forward with enjoyment to things</td>
<td>733</td>
<td>338</td>
<td>107</td>
<td>12</td>
<td>1190 .49 .70</td>
</tr>
<tr>
<td>13A I get sudden feelings of panic*</td>
<td>909</td>
<td>246</td>
<td>26</td>
<td>9</td>
<td>1190 .27 .54</td>
</tr>
<tr>
<td>14D I can enjoy a good book or radio/TV program</td>
<td>990</td>
<td>167</td>
<td>22</td>
<td>11</td>
<td>1190 .21 .51</td>
</tr>
<tr>
<td>Total A</td>
<td>1190</td>
<td></td>
<td></td>
<td></td>
<td>3.35 .67 .79</td>
</tr>
<tr>
<td>Total D</td>
<td>1190</td>
<td></td>
<td></td>
<td></td>
<td>2.95 .54 .66</td>
</tr>
<tr>
<td>Total A+D</td>
<td>1190</td>
<td></td>
<td></td>
<td></td>
<td>6.30 .46 .80</td>
</tr>
</tbody>
</table>

Note. N=1190. A= anxiety and D= depression. * Items were scored on a four-point scale ranging from totally disagree to totally agree. A-response: 0 = 'Not at all', 1 = 'Not very often', 2 = 'Quite often' or 3 = 'Very often' and D-response: 0 = 'Most of the time', 1 = 'Sometimes', 2 = 'Not often' or 3 = 'Not at all'. N= 1190. The standard scoring algorithm was used for A = sum of items 1*, 3*, 5*, 7, 9, 11*, 13*; and for D = sum of items 2, 4, 6*, 8*, 10*, 12, 14. * Items starred are reverse scored.

Confirmatory Factor Analysis (CFA)

We tested the original two-factor solution including 14 items and termed this solution Model-1; the factor loadings (λ) ranged between 0.32 and 0.72, followed by multiple squared correlations (R²) from 0.10 to 0.52. Three items belonging to the depression construct; 8D (“I feel as I'm slowed down”), 10D (“I have lost interest in my appearance”) and 14D (“I can enjoy a good book or a TV program”) revealed low loadings 0.34, 0.32 and 0.37, respectively; explaining 12%, 11% and 13% of the
variance of the depression construct. The fit indices indicated misspecification: \( \chi^2=395.010, \ p=0.00001, \ \text{df}=76, \ \chi^2/\text{df}=5.20, \ \text{RMSEA}=0.059, \ \text{p-close}=0.04, \ \\
\text{SRMR}=0.052, \ \text{CFI}=0.91, \ \text{TLI}=0.89 \) (Table 3). The RMSEA, which is an estimate of approximate fit was acceptable, while the \( \chi^2 \) was much too high. For an acceptable fit the \( \chi^2/\text{df} \) should be \( \leq 3.0 \), and \( \leq 2 \) for a good fit. Further, the CFI and TLI were too low, all of which indicating misspecification. Exploring the normalized residuals, 23 residuals were significant with item 6\(_D\) ("I feel cheerful") involved in several highly significant estimates. Hence, we scrutinized the modification indices (MI) presenting some extremely high values; item 6\(_D\) exposed an extremely high MI=77.29 with the anxiety factor, and an MI=30.374 with item 5\(_A\) ("Worrying thoughts go through my mind"). Also, item 2\(_D\) ("I still enjoy the things I used to enjoy") and 12\(_D\) ("I look forward with enjoyment to things") demonstrated an exceptionally high MI=44.383. In total, 16 MIs were \( \geq 10 \).

Looking at one factor at a time, both anxiety and depression including 7 items each demonstrated a too high \( \chi^2 \), while the other indices were good to acceptable (\textit{Anxiety}: \( \chi^2=61.695, \ p=0.00001, \ \text{df}=14, \ \chi^2/\text{df}=4.41, \ \text{RMSEA}=0.054, \ \text{p-close}=0.320, \ \text{SRMR}=0.029, \ \text{CFI}=0.97, \ \text{TLI}=0.96; \textit{Depression}: \chi^2=84.264, p=0.00001, \ \text{df}=14, \ \chi^2/\text{df}=6.02, \ \text{RMSEA}= 0.065, \ \text{p-close}=0.30, \ \text{SRMR}=0.038, \ \text{CFI}=0.94, \ \text{TLI}=0.92)\) in Table 3. Composite reliability was good, showing estimates of \( \rho_{\text{Anxiety}}=0.78 \) and \( \rho_{\text{Depression}}=0.70 \).
It is plausible that if one is feeling cheerful, one is improbable to feel anxiety at the same time, and vice versa. Thus, it is theoretically meaningful that feeling cheerful (item 6_D) and feeling anxiety correlate negatively. Accordingly, in Model-2 we included a path (cross-loading) from item 6_D (“I feel cheerful”) to the Anxiety construct, which improved the fit considerably: $\chi^2=317.951$, $p=0.0001$, $df=75$, $\chi^2/df=4.24$, RMSEA=0.046, p-close=0.262, SRMR=0.042, CFI=0.93, TLI=0.92. However, still the fit was not good.

An extremely high MI between the items 2_D (“I still enjoy the things I used to enjoy”) and 12_D (“I look forward with enjoyment to things“) was uncovered. It is rational that still enjoying things and looking forward to things with enjoyment correlate. Therefore, we included a correlated error term between item 2_D and 12_D which further improved the fit in Model-3: $\chi^2=296.919$, $p=0.0001$, $df=74$, $\chi^2/df=4.01$, RMSEA=0.050, p-close=0.452, SRMR=0.040, CFI=0.94, TLI=0.92. For Model-3, composite reliability ($\rho_c$) was 0.78 for the anxiety subscale and 0.65 for depression subscale. Still, 14 normalized residuals were significant, asking for several cross-loadings and correlated errors.

Table 3. Goodness-of-fit indices for HADS measurement models: Model-1, Model-2, Model-3.

<table>
<thead>
<tr>
<th>Fit Measure</th>
<th>Model-1 2-factors</th>
<th>Model-2 2-factors</th>
<th>Model-3 2-factors</th>
<th>Model Anxiety</th>
<th>Model Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ Satorra Bentler</td>
<td>395.010</td>
<td>317.951</td>
<td>296.919</td>
<td>61.695</td>
<td>84.264</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\chi^2/df$ Satorra Bentler</td>
<td>5.20</td>
<td>4.18</td>
<td>4.01</td>
<td>4.41</td>
<td>4.52</td>
</tr>
<tr>
<td>Df=76</td>
<td>Df=75</td>
<td>Df=74</td>
<td>Df=14</td>
<td>Df=14</td>
<td></td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.059</td>
<td>0.052</td>
<td>0.050</td>
<td>0.054</td>
<td>0.065</td>
</tr>
</tbody>
</table>
Therefore, we checked if a unidimensional solution termed **Model-4** would fit better, though revealing an exceedingly bad fit: \( \chi^2=840.194, \ p=0.0001, \ df=77, \chi^2/df=10.91, \ RMSEA=0.091, \ p-close=0.0001, \ SRMR=0.070, \ CFI=0.78, \ TLI=0.74. \)

However, internal consistency was good \( \rho_{HADS}=0.81. \) Hence, possibly the misspecification was caused by error covariances. Consequently, we turned to Model-3, again scrutinizing the MI values. The items 8\( _D \), 12\( _D \) and 14\( _D \) displayed MIs>15. These items also displayed low loadings indicating poor reliability as indicators for the depression construct. Possibly, removing some of these items would improve the model fit. Nonetheless, composite reliability was \( \rho_{Depression}=0.65; \) hence, removing items would cause a week construct with even lower reliability.

All the tested models revealed a chi-square indicating misspecification. However, it is well known that chi-square as a model fit index has limitations. First and foremost, chi-square is sensitive to sample size (48). The present sample is large (N=1190). For that reason, we randomly split the data in two equally sized parts with n=595, representing a sample size more suitable for SEM (structural
equation modeling) (47-49). We termed the two parts as Sample1 and Sample2 and tested the original HADS version (Model-1) in both samples. Except the chi-square showing better values, the fit indices demonstrated a similar pattern:

**Sample1** \(\chi^2=225.391, p=0.0001, df=76, \chi^2/df=2.97, \text{RMSEA}=0.058, \text{p-close}=0.074, \text{SRMR}=0.057, \text{CFI}=0.90, \text{TLI}=0.88, \rho_{\text{Anxiety}}=0.76\) and \(\rho_{\text{Depression}}=0.67\). **Sample2** \(\chi^2=257.228, p=0.0001, df=76, \chi^2/df=3.38, \text{RMSEA}=0.063, \text{p-close}=0.005, \text{SRMR}=0.053, \text{CFI}=0.91, \text{TLI}=0.90, \rho_{\text{Anxiety}}=0.79\) and \(\rho_{\text{Depression}}=0.73\). Loadings ranged between 0.31 and 0.67 for Sample1 and between 0.29 and 0.77 for Sample2.

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Model 3 – the best fitting measurement model.
Finally, Spearman’s rank correlation was computed to assess the relationship between HADS and overall global QOL (“Thinking about your life at the moment, would you say that you mostly are satisfied with life, or are you mostly dissatisfied?”). There was a negative correlation between high HADS-A score and overall good global QOL, \((r=-0.408, p < 0.001)\) and high HADS-D score and overall good global QOL, \((r = -0.450, p < 0.001)\).

**Discussion**

According to the European Commission’s Green paper on mental health (50), depression is one of the most prevalent mental health problems facing European citizens today. The incidence of depression with increasing age is stated (15), simultaneously the number of adults over 80 years is globally expected to dramatically increase in the coming decades (51). Hence, access to a valid and reliable scale assessing anxiety and depression among older community-dwelling adults is highly warranted. Therefore, the aim of the present study was to evaluate the psychometric properties of the HADS among community-dwelling older Norwegians \(\geq 70\) years. The present sample included 1190 older adults, with a mean age of 76.5 years. To the authors’ knowledge, no previous studies have examined the psychometric properties of the HADS in a Norwegian population among community-dwelling older adults by means of CFA.
The CFA approach eliminates the need to summate scales because the SEM programs, such as STATA, compute latent construct scores for each respondent. This process allows relationships in the model tested to be automatically corrected for error variance which is a fundamental strength of CFA to construct validation. Thus, the resulting estimates are adjusted for measurement error (47, 49). In this study, the original version of the HADS (Model-1) showed only partly a good fit. Especially, the chi-square demonstrated extremely high values indicating misspecification. However, utilizing the chi-square as a model fit index relates to some limitations. As already stated, chi-square is sensitive to sample size: a misfit may be trivial, still with larger samples the p-value decreases followed by higher estimates (48). This means, that in practice, the chi-square test is “not always the final word in assessing fit” (52). The present sample size is large (N=1190) revealing extraordinarily high estimates for the chi-square. When splitting the file into two parts giving a sample size N=595, the chi-square improved substantially, and the RMSEA was still acceptable. Hence, reflecting on the chi-square statistic in the light of the large sample size, a wide variety of other indices were included to assess model adequacy. The SEM literature states that as a minimum, RMSEA, CFI, and SRMR should be reported in combination with chi-square (46). The use of multiple fit indices provides a more holistic view of goodness of fit, accounting for sample size, model complexity, and other considerations relevant to the study.
Conversely, the RMSEA estimate has demonstrated lower values with large sample sizes (53, 54). For an acceptable fit RMSEA should be ≤ 0.080 (46, 47, 49) or ≤0.10 (42), while estimates ≤ 0.050 suggest a good fit. Looking at Model-1, the RMSEA along with SRMR were acceptable and almost good (0.059, 0.052, respectively), while the CFI and TLI were too low. Concerning CFI and TLI, including a cross-loading item (6D) along with a correlated error term between the items 2D and 12D improved these fit indices as well as the total model fit. Consequently, low reliability and content validity seemed to cause the low values for CFI and TLI.

**Dimensionality**

Concerning the dimensionality of the HADS, the two-factor-model undoubtedly showed the best fit to the present data; the dimensionality of the HADS questionnaire stood out to be unquestionable. The two factors were properly correlated.

**Reliability**

All items were significant. Largely, the items revealed good loadings accompanied by good multiple squared correlations (R²) demonstrating good reliability. Nevertheless, particularly three items belonging to the Depression construct (8D, 10D, 14D) demonstrated low factor loadings and thus poor reliability, explaining
very little of the variance in the construct. These three items caused a low reliability coefficient for depression, while anxiety displayed good reliability.

**Construct validity**

Construct validity concerns if the set of measured items reflects the theoretical latent construct those items are designed to measure. Hence, it deals with accuracy of measurement involving psychometric evidence of convergent and discriminant validity (55). A measure is said to process convergent validity if independent measures of the same construct converge or are highly correlated (35). Normally, researchers do not have data on two different e.g., depression scales scored by the same sample: this represents a frequent problem connected with convergent validity. However, measures which theoretically are predicted to significantly correlate with depression might be used. The present study included measures of overall global QOL to test for convergent validity, which was supported by a significant correlation in an expected direction.

Discriminant validity entails that a measure must not correlate too highly with other measures from which it is expected to diverge (35). The anxiety and depression factors performed like distinct concepts, supporting the discriminant validity. Simultaneously, the factor correlation between anxiety and depression was highly significant supporting convergent validity (35). The convergent and
discriminant validity were further supported by significantly correlations in the predicted direction for anxiety and depression towards QOL.

**Content validity – a vital aspect of construct validity**

Content validity is a central aspect of construct validity. Reliability and content validity represent interrelated measurement properties. In fact, despite a good reliability content validity might be poor. Contrariwise, validity cannot be good if reliability is low (35). The item 8D concerns “I feel as I’m slowed down” demonstrated low reliability and thereby poor validity. In the present sample with a mean age of 76.5 years, most individuals have for a long time been outside an active work-life having lots of time to adjust to a slower pace of life. Possibly, ‘feeling slowed down’ does not correspond well to older home-living adults’ daily experiences in relation to depression. This item did not perform to be a valid nor reliable indicator of depression in this population. What is more, about 50% of the participants reported physical or mental long-term illness, injury, or loss of function in daily life. Relevantly, a slower pace of life might seem natural and not necessarily an indicator of depression (2). Likewise, item 10D “I have lost interest in my appearance” did not communicate well with these older adults, indicating low reliability and content validity. Losing interest in one’s appearance did not act like a valid indicator of depression in this population. Loose of interest in one’s appearance may be reasoned in the inevitable age-related changes they experience rather
than as a depressive symptom. Moreover, item 14D “I can enjoy a good book or TV program” also stood out as an unreliable indicator of depression. Plausibly, being old, enjoying a good book or watching TV do not relate with depression. Living in your seventies-eighties-nineties, passive leisure activities are common activities which is useful as restoration time after active leisure activities and related to QOL (56). Reading books might be more demanding due to a declined sight as well as fatigue. Consequently, item 14D did not explain any variance in the depression construct and thus misbehaved as a valid indicator for the depression construct. These findings are consistent with previous studies among nursing home residents without cognitive impairment (28) and in hospitalized older adults (29) where the same three items were troublesome among older adults in Norwegian care facilities. In older ages, for the first time in their life retired adults can slow down. Also, due to a decline in age-related reserve capacity and fear of falling which is the most common fear in older adults (57), many older adults may be forced to a slower pace. Doing passive activities such as watching TV or reading may also be a consequence of having chronic medical condition and multimorbidity which is associated with anxiety and depression (58). Hence, to improve reliability the wording of the items 8D, 10D and 14D should be carefully considered. Furthermore, the former validation study among older adults in nursing homes (28) also involved a cross-loading for item 6D to anxiety, as well as highly significant error variances between items 2D and 12D. Surprisingly, home-
dwelling older adults (the present study), nursing home residents (two different samples giving an approximate N=500; mean age 84.5 and 86 years) (28) and hospitalized older adults (N=484; mean age 80.7 years) (29) respond similarly findings of the HADS-D items.

Summarized, construct validity and reliability of anxiety were good. Conversely, the depression construct revealed low validity and reliability which are interrelated measurement properties. Exclusively, content validity includes to which extent elements of a measurement scale are appropriate to and characteristic of the specific construct for a certain assessment purpose (35). In this study, content validity concerns if the 14 HADS items and the two-factor dimensionality precisely represent anxiety and depression in this population. Besides, evidence of face validity can be considered as one aspect of content validity (35). High face validity of an instrument increases its use in practical situations via ease of use, proper reading level, clarity, and appropriate response formats. Thus, to improve content validity and thereby also reliability for the depression factor, qualitative studies could be applied to get closer to the actual content of depression, investigating what might be the most essential indicators for depression among community-dwelling older adults. Based in such novel evidence, the three troublesome items could be formulated in a more valid format.
Strengths and limitations

A notable strength of this research is the empirical examination of the HADS which has not been tested previously in a community-dwelling population in Norway using CFA. Also, the large sample size is a strength, giving the possibility to randomly split the sample into two different samples including 595 community-dwelling older adults each.

Although the older adults were selected randomly in two subsamples, we cannot state that the sample is representative to the community-dwelling older adults in the actual city since 3340 of 5087 declined participation. In addition, those who participated in HUNT4 Trondheim 70+ but excluded from this present study had significantly more often dementia diagnosis, were more often females, had higher age and lower level of education. Hence, we assume that the present sample may be disrupted in that respect not representing all community-dwelling older adults.

Conclusion

This study showed that the two-factor structure assessing symptoms of anxiety and depression is unquestionable. In conclusion, when we included a cross-loading item (6D) along with a correlated error term between item 2D and 12D (Model-3), a good to acceptable measurement reliability was demonstrated, and construct validity was supported.
However, concerning the internal consistency, the original version of HADS (Model-1) revealed a good reliability coefficient for anxiety, but a poor estimate for depression; items 8_D, 10_D and 14_D stood out to be unreliable and invalid indicators for depression in this population. Therefore, to be valid indicators of depression among community-dwelling older individuals, these items need to be re-worded informed by qualitative studies exploring symptoms of depression among older adults living at home.

**Abbreviations**

CFA: Confirmative factor analysis; CFI: Comparative Fit Index; CSDS: Clinically significant depressive symptoms; HUNT: The Trøndelag Health Study; MDD: Major depressive disorder; RMSEA: Root Mean Square Error of Approximation; SEM: structural equation modeling; SRMR: Standardized Root Mean Square Residual; MI: Modification indices; TLI: Tucker-Lewis Index.

**Acknowledgments**

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The authors wish to acknowledge the HUNT Research Centre, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology and National Advisory Unit for Aging for project management in
the HUNT4 Trondheim 70+ survey. Further, we wish to acknowledge Trondheim municipality, students from Norwegian University of Science and Technology for collaborating in the data collection, as well as the older adults who voluntarily participated in the study.

Authors contribution

Heidi Emly Sivertsen: Data collection - Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. Anne-Sofie Helvik: Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing. Linda Gjøra: Data collection - Writing - review & editing. Gørill Haugan: Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Writing - review & editing.

Role of funding source

The funding sources were the researchers' affiliations, providing salaries and providing financial support to the research project by Faculty of Medicine and Health Sciences Department of Public Health and Nursing.

Availability of data and materials

The Trøndelag Health Study (HUNT) has invited persons aged 13 - 100 years to four surveys between 1984 and 2019. Comprehensive data from more than 140,000 persons having participated at least once and biological material from 78,000 persons is collected. The data are stored in HUNT databank and biological material in HUNT biobank. HUNT
Research Centre has permission from the Norwegian Data Inspectorate to store and handle these data. The key identification in the data base is the personal identification number given to all Norwegians at birth or immigration, whilst de-identified data are sent to researchers upon approval of a research protocol by the Regional Ethical Committee and HUNT Research Centre. To protect participants’ privacy, HUNT Research Centre aims to limit storage of data outside HUNT databank and cannot deposit data in open repositories. HUNT databank has precise information on all data exported to different projects and can reproduce these on request. There are no restrictions regarding data export given approval of applications to HUNT Research Centre. For more Information: Åsvold BO, Langhammer A, Rehn TA et al., Cohort profile update: The HUNT Study, Norway. Int J Epidemiol. Published 17. May 2022.

**Declarations**

**Ethics approval and consent to participate**

All participants received oral and written information and gave informed written consent prior the study. The project was submitted to the Regional Ethics Committee for Medical and Health Research (REC) in Mid-Norway, reference REK 2021/348836, and was approved. This study was conducted in accordance with the ethical principles stated in the Declaration of Helsinki.

**Competing interests**

The authors declare that they have no competing interest.
References


41. Stata Corp. Stata Statistical Software. 17 ed. College Station, TX: Stata Corp LLC; 2021.