Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population

Bernd Löwe, MD, PhD,* Oliver Decker, PhD,† Stefanie Müller, MS,‡ Elmar Brähler, PhD,† Dieter Schellberg, PhD,‡ Wolfgang Herzog, MD,§ and Philipp Yorck Herzberg, PhD†

Background: The 7-item Generalized Anxiety Disorder Scale (GAD-7) is a practical self-report anxiety questionnaire that proved valid in primary care. However, the GAD-7 was not yet validated in the general population and thus far, normative data are not available.

Objectives: To investigate reliability, construct validity, and factorial validity of the GAD-7 in the general population and to generate normative data.

Research Design: Nationally representative face-to-face household survey conducted in Germany between May 5 and June 8, 2006.

Subjects: Five thousand thirty subjects (53.6% female) with a mean age (SD) of 48.4 (18.0) years.

Measures: The survey questionnaire included the GAD-7, the 2-item depression module from the Patient Health Questionnaire (PHQ-2), the Rosenberg Self-Esteem Scale, and demographic characteristics.

Results: Confirmatory factor analyses substantiated the 1-dimensional structure of the GAD-7 and its factorial invariance for gender and age. Internal consistency was identical across all subgroups (α = 0.89). Intercorrelations with the PHQ-2 and the Rosenberg Self-Esteem Scale were r = 0.64 (P < 0.001) and r = −0.43 (P < 0.001), respectively. As expected, women had significantly higher mean (SD) GAD-7 anxiety scores compared with men [3.2 (3.5) vs. 2.7 (3.2); P < 0.001]. Normative data for the GAD-7 were generated for both genders and different age levels. Approximately 5% of subjects had GAD-7 scores of 10 or greater, and 1% had GAD-7 scores of 15 or greater.

Conclusions: Evidence supports reliability and validity of the GAD-7 as a measure of anxiety in the general population. The normative data provided in this study can be used to compare a subject’s GAD-7 score with those determined from a general population reference group.

Key Words: anxiety, outcome assessment, psychometrics, reference standards, validation studies

The National Comorbidity Survey Replication (NCS-R) has shown that anxiety disorders are the most prevalent class of mental disorders in the general population.1,2 In a recent primary care-based anxiety study,3 19.5% of patients were diagnosed with at least 1 of the 4 most common anxiety disorders (ie, generalized anxiety disorder, panic disorder, social anxiety disorder, and posttraumatic stress disorder).4–6 Despite the substantial association of these 4 anxiety disorders with comorbid depression, somatic symptom burden, functional impairment, and high use of health care services,3,7,8 and despite the availability of effective treatment,9,10 only a minority of patients with anxiety are recognized in primary care11,12 and 41% of patients with anxiety disorders report no current treatment.13 Compared with the efforts to improve diagnosis and treatment for depression in primary care, anxiety disorders are sometimes regarded as the neglected stepchild of primary care-based mental health care.3 As a consequence, screening for anxiety disorders was suggested as necessary first step in improving outcomes in patients with anxiety disorders.13

Whereas numerous brief and validated self-report measures for depression,14,15 such as the 9-item Patient Health Questionnaire (PHQ-9),16–18 exist, there is a paucity of brief, validated, and freely available measures for anxiety.19 To overcome this paucity, the 7-item Generalized Anxiety Disorder Scale (GAD-7) was recently developed in a large primary care patient sample.3,19 This self-report questionnaire evidenced good reliability, as well as criterion, construct, factorial, and procedural validity. With respect to the criterion standard diagnostic interview conducted in 965 patients, sensitivity of the GAD-7 (cut point ≥10) for diagnosing generalized anxiety disorder was 89% and specificity was 82%.19 Another article shows that the GAD-7 not only performs well as a screening tool for detecting generalized anxiety disorder, but also performed almost as well for detecting panic disorder, social anxiety disorder, and posttraumatic stress disorder.3 Nevertheless, given that no study has investigated the GAD-7 outside the primary care setting, it is necessary to investigate reliability and validity of this...
new instrument in other settings. In addition, normative data from the general population are necessary to interpret the clinical significance of individual GAD-7 scores.

Therefore, we aimed to investigate reliability and validity of the GAD-7 in a large and representative sample from the general population and to provide normative data for the GAD-7. Specifically, we investigated item characteristics, reliability, and factorial validity including factorial invariance for gender and age. Second, construct validity of the GAD-7 was assessed in the general population by investigating associations between GAD-7 scores, other self-report measures, and known risk factors of generalized anxiety disorder. Third, with the hypotheses that GAD-7 scores would be lower in the general population compared with GAD-7 scores both in primary care patients and patients with generalized anxiety disorder; we compared GAD-7 scores across these subject groups. Finally, to provide comparative data for the application of the GAD-7, we generated age- and gender-specific normative data for the GAD-7.

METHODS

Study Design and Participants

This study was part of a nationally representative face-to-face household survey conducted in Germany. Within this survey, the study participants were interviewed using a structured self-report questionnaire. The survey was carried out in 2 waves between May 5 and June 8, 2006, by a total of 231 (first wave) and 237 (second wave) professional interviewers of a demographic consultation company (USUMA, Berlin). Within each wave, a representative sample of the German population aged 14 years or older was approached using 258 sample points. Addresses were selected following the random-route procedure. Of the 8398 selected addresses, 8106 were valid. The interviewers were controlled by sending prestamped postcards to 31% of the participants requesting to confirm the conduct of the interview. The actual conduct of the interview was confirmed by 100% of the returned postcards.

For the comparison of GAD-7 general population scores with GAD-7 scores from primary care patients and patients with generalized anxiety disorder, we used data from the original GAD-7 primary care validation study. From this study, we used the total primary care patient group and the patient subgroup with generalized anxiety disorder, as diagnosed with a criterion standard psychiatric interview, as comparison groups for our general population sample. The GAD-7 primary care validation study is described in detail elsewhere.

Measures

Participants of the survey completed a self-report questionnaire that included the GAD-7. The GAD-7, which was designed to identify probable cases of generalized anxiety disorder and to assess symptom severity, evidenced high reliability and validity in primary care patients. The GAD-7 items describe the most prominent diagnostic features of the DSM-IV diagnostic criteria A, B, and C for generalized anxiety disorder. On the GAD-7, subjects are asked how often, during the last 2 weeks, they have been bothered by each of the 7 core symptoms of generalized anxiety disorder. Response options are “not at all,” “several days,” “more than half the days,” and “nearly every day,” scored as 0, 1, 2, and 3, respectively. Therefore, GAD-7 scores range from 0 to 21, with scores of ≥5, ≥10, and ≥15 representing mild, moderate, and severe anxiety symptom levels, respectively. These cut-points were chosen based on receiver operating characteristic analyses in the GAD-7 primary care validation study. The German version of the GAD-7 used here was compiled with a total of 7 steps of translation and blind back-translation conducted by 4 independent translators following state-of-the-art procedures for test translation. The translation process was directed and supervised by the first author of this article, who is also one of the authors of the original English version. The American and German versions of the GAD-7 are freely available.

The survey questionnaire also included questions about age, gender, education, marital status, employment status, net household income, nationality, place of residence, and church membership. As an indicator of depression, the survey included the validated German version of the 2-item depression module from the Patient Health Questionnaire. In addition, validated German versions of the Rosenberg Self-Esteem Scale, of the Questionnaire on Life Satisfaction, and of the Resilience Scale were part of the survey.

Data Analysis

To assess generalizability of our results to the German population and to the American population, we compared demographic characteristics of our study population with federal data from the total German population and data from the US NCS-R. These comparisons were performed descriptively, because, as a result of the large sample sizes included in the federal data of the German population and the NCS-R, even very small differences would reach the level of statistical significance.

Item characteristics of the GAD-7 items were investigated including item mean scores, item-intercorrelations, and corrected item-total correlations. For reliability, internal consistency of the GAD-7 was assessed for the total general population sample, both genders, and 4 different age groups. The factor structure of the GAD-7 was examined with confirmatory factor analysis. Taking the ordinal nature of the data into account, we based the analyses on a 2-stage estimation approach developed by Lee et al. In step 1, polychoric correlations were computed. The model was then analyzed with the correct weight matrix obtained in step 1. This approach ensures unbiased parameter estimates and standard errors for ordinal data. We used 4 criteria to assess how well a model fitted the data. Two of these criteria are measures of absolute model fit: the root mean square error of approximation (RMSEA) and the 90% confidence interval for RMSEA. The other 2 criteria are measures of relative model fit: the comparative fit index (CFI) and the Tucker Lewis index (TLI). RMSEA < 0.05 represents a “close fit”; RMSEA between 0.05 and 0.08 represents a “reasonably close fit”; and RMSEA > 0.10 represents an “unacceptable model.” CFI and TLI indicate how well a model fits the data. These two criteria are measures of absolute model fit: the root mean square error of approximation (RMSEA) and the 90% confidence interval for RMSEA. The other 2 criteria are measures of relative model fit: the comparative fit index (CFI) and the Tucker Lewis index (TLI). RMSEA < 0.05 represents a “close fit”; RMSEA between 0.05 and 0.08 represents a “reasonably close fit”; and RMSEA > 0.10 represents an “unacceptable model.” CFI and TLI indicate how well a model fits the data.
data relative to a “null” model, which assumes that sampling error alone explains the covariation among the observed measures. Hu and Bentler\textsuperscript{34} have suggested that measurement models should have a CFI and TLI of at least 0.95.

For construct validity,\textsuperscript{38} we investigated GAD-7 scale intercorrelations with the PHQ-2,\textsuperscript{23,24} the Rosenberg Self Esteem Scale,\textsuperscript{25,26} the Questionnaire on Life Satisfaction,\textsuperscript{27} and the Resilience Scale.\textsuperscript{28} In addition, we investigated associations between GAD-7 scores and sociodemographic characteristics that are known as risk factors associated with generalized anxiety disorder. Based on results from previous anxiety studies, we expected that women would have higher anxiety levels compared with men,\textsuperscript{2,39,40} that levels of anxiety increase with age in the respondents aged between 45 and 60 years,\textsuperscript{30,39–41} that subjects living with a partner would have lower anxiety scores compared with subjects not living with a partner,\textsuperscript{1,40} that anxiety scores would be lower in subjects with higher educational levels compared with subjects with lower educational levels,\textsuperscript{40,42} that employed subjects would have lower anxiety scores compared with unemployed subjects,\textsuperscript{40} and that anxiety levels would be lower in subjects with higher household income compared with subjects with lower household income.\textsuperscript{42,43}

For other demographic variables (ie, nationality, place of residence, urbanicity of domicile, and church membership) we had no directional hypotheses and performed exploratory group comparisons. Univariate comparisons with the GAD-7 score as dependent variable were performed using \textit{t} test and analysis of variance. In addition, to identify the sociodemographic variables with the strongest association with the anxiety sum score, we performed a multiple linear regression analysis with the GAD-7 score as dependent variable and demographics as independent variables. However, given the relatively low prevalence of anxiety in the general population, the confounding of demographic variables such as education and income, and the unspecific influence of demographics on anxiety, we expected associations of small to moderate size between anxiety and demographic characteristics.

Construct validity was also assessed by comparing the GAD-7 general population scores to the GAD-7 scores from primary care and patients with diagnosis of generalized anxiety disorder.\textsuperscript{44} For this purpose, we used analyses of covariance adjusted for age and gender. Because the GAD-7 score as dependent variable displayed a skewed (but unimodal) distribution, we also reran the models using the rank transformation of the dependent variable. Bonferroni-adjustment was performed to account for multiple testing.

Finally, to provide normative data for the GAD-7, we generated age- and gender-specific percentiles for each GAD-7 score. Percentiles were used because of their independence from the distribution of the scale scores. Our sample size was sufficient to create gender-specific age groups comprising 10 years each. Given the large sample size, an \( \alpha \)-level of 5\% (2-sided), and a \( \beta \)-level of 20\%, our study was sufficiently powered to detect small effects in all analyses. Statistical analyses were performed using SPSS (Version 14.0; Chicago, IL), EQS (Version 6.1; Encino, CA), and AMOS (Version 5; Chicago, IL).

### RESULTS

#### Sample Characteristics

From the 8106 valid addresses, 1199 persons (14.8\%) were not in at the time of the 3 visits of the interviewers, 1806 persons refused to participate (22.3\%), and 65 persons (0.8\%) were not able to complete the study questionnaire because of severe illness. A total of 5036 persons agreed to participate, provided verbal informed consent, and completed the study questionnaire. Response rate among all subjects met by the interviewers was 72.9\% (5036/6906), whereas participation rate among all eligible subjects was 62.1\% (5036/8106). The demographic characteristics of our study population are summarized in Table 1. Mean (SD) age of the participants was 48.4 (18.0) years, 53.6\% were women. The characteristics of the study sample closely match those of the total German population\textsuperscript{29} and the US NCS-R\textsuperscript{42,43} on gender (women: 53.6\%, 51.7\%, and 55.5\%), employment status (unemployed: 5.8\%, 7.1\%, and 3.9\%), marital status (married: 54.9\%, 57.2\%, and 57.2\%), and educational level. In addition, the mean age in our study sample was similar to the mean age in the German general population aged 14 years or older (48.4 vs. 46.9 years). However, compared with the German general population, subjects with non-German nationality were somewhat underrepresented in our study sample (3.2\% vs. 8.8\%).

The comparison group from the GAD-7 primary care validation study included 965 subjects with a mean (SD) age of 47.1 (15.1) years and 69\% female subjects.\textsuperscript{3,19} The patient subsample with generalized anxiety disorder included 73 subjects, aged 45.7 (13.2) years, of which 82\% were women. Demographic characteristics of the comparison groups were previously described in more detail elsewhere.\textsuperscript{3,39}

#### Item Characteristics, Internal Consistency, and Factorial Validity

Characteristics of the GAD-7 items and the total scale are summarized in Table 2. The mean (SD) score of the total GAD-7 scale is \( M = 2.95 \) (3.41) and the mean item scores range from 0.30 to 0.54. Corrected item-total correlations (ie, correlations between the respective item and the total sum score, without the respective item) were \( r = 0.63 \) or greater. Taking the brevity of the scale into account, internal consistency of the GAD-7 was acceptable (\( \alpha = 0.89 \)). This held true for all subsamples tested above (ie, Cronbach’s \( \alpha \) remained unchanged for all subsamples). Intercorrelations between the GAD-7 items ranged from \( r = 0.45 \) to \( r = 0.65 \).

For testing the unidimensionality of the GAD-7, all 7 items were specified as indicators of a single factor. This model fitted the data well, as indicated by the robust \( \chi^2 \) goodness-of-fit index\textsuperscript{25} \( \chi^2_{S-B} = 314.10 \), df = 14, \( P < 0.001 \), and CFI = 0.990, TLI = 0.990, RMSEA = 0.065, and the 90\% confidence interval for RMSEA = 0.059–0.072. As shown in Table 2, factor loadings were high, ranging from 0.76 to 0.90 with a mean of 0.83.

Multigroup confirmatory factor analysis was used to test the appropriateness of the unidimensional model across
TABLE 1. Demographic Characteristics of Study Sample

<table>
<thead>
<tr>
<th></th>
<th>Total Sample (N = 5030)</th>
<th>Men (n = 2332)</th>
<th>Women (n = 2698)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group, yr (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14–24</td>
<td>11.1</td>
<td>12.7</td>
<td>9.6</td>
</tr>
<tr>
<td>25–34</td>
<td>13.6</td>
<td>12.7</td>
<td>14.3</td>
</tr>
<tr>
<td>35–44</td>
<td>19.2</td>
<td>17.1</td>
<td>21.0</td>
</tr>
<tr>
<td>45–54</td>
<td>17.2</td>
<td>17.4</td>
<td>16.9</td>
</tr>
<tr>
<td>55–64</td>
<td>16.0</td>
<td>17.2</td>
<td>15.0</td>
</tr>
<tr>
<td>65–74</td>
<td>15.5</td>
<td>16.9</td>
<td>14.4</td>
</tr>
<tr>
<td>≥75</td>
<td>7.5</td>
<td>6.0</td>
<td>8.7</td>
</tr>
<tr>
<td>Living with a partner (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>59.9</td>
<td>62.7</td>
<td>57.4</td>
</tr>
<tr>
<td>No</td>
<td>40.1</td>
<td>37.3</td>
<td>42.6</td>
</tr>
<tr>
<td>Education, yr (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–8</td>
<td>1.1</td>
<td>0.9</td>
<td>1.3</td>
</tr>
<tr>
<td>9</td>
<td>44.2</td>
<td>45.2</td>
<td>43.3</td>
</tr>
<tr>
<td>10–11</td>
<td>34.3</td>
<td>31.0</td>
<td>37.0</td>
</tr>
<tr>
<td>12</td>
<td>2.9</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>13–15</td>
<td>7.6</td>
<td>7.7</td>
<td>7.6</td>
</tr>
<tr>
<td>≥16</td>
<td>6.5</td>
<td>8.0</td>
<td>5.2</td>
</tr>
<tr>
<td>Currently student</td>
<td>3.5</td>
<td>4.2</td>
<td>2.8</td>
</tr>
<tr>
<td>Employment status (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In training</td>
<td>8.5</td>
<td>10.3</td>
<td>7.0</td>
</tr>
<tr>
<td>Full-time employment</td>
<td>35.9</td>
<td>48.5</td>
<td>25.1</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>10.4</td>
<td>1.8</td>
<td>17.8</td>
</tr>
<tr>
<td>Military or civilian service, maternity leave</td>
<td>0.9</td>
<td>0.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Unemployed/short-time work</td>
<td>5.8</td>
<td>6.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Homemaker</td>
<td>8.3</td>
<td>0.2</td>
<td>15.3</td>
</tr>
<tr>
<td>Retired</td>
<td>30.1</td>
<td>32.3</td>
<td>28.2</td>
</tr>
<tr>
<td>Net household income (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;750 €/mo</td>
<td>4.1</td>
<td>4.3</td>
<td>3.9</td>
</tr>
<tr>
<td>750–1249 €/mo</td>
<td>18.3</td>
<td>14.7</td>
<td>21.5</td>
</tr>
<tr>
<td>1250–1999 €/mo</td>
<td>36.1</td>
<td>37.0</td>
<td>35.4</td>
</tr>
<tr>
<td>≥2000 €/mo</td>
<td>41.5</td>
<td>44.1</td>
<td>39.2</td>
</tr>
<tr>
<td>Nationality (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>96.8</td>
<td>96.5</td>
<td>97.0</td>
</tr>
<tr>
<td>Other</td>
<td>3.2</td>
<td>3.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Place of residence in Germany (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>80.0</td>
<td>80.0</td>
<td>80.1</td>
</tr>
<tr>
<td>East</td>
<td>20.0</td>
<td>20.0</td>
<td>19.9</td>
</tr>
<tr>
<td>Urbanicity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban area</td>
<td>86.8</td>
<td>86.8</td>
<td>86.8</td>
</tr>
<tr>
<td>Rural area</td>
<td>13.2</td>
<td>13.2</td>
<td>13.2</td>
</tr>
<tr>
<td>Church member (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>76.6</td>
<td>72.9</td>
<td>79.8</td>
</tr>
<tr>
<td>No</td>
<td>23.4</td>
<td>27.1</td>
<td>20.2</td>
</tr>
</tbody>
</table>

*Complete values were available for all variables except household income (N = 4767) and church membership (N = 5020).

gender and age (Table 3). Two nested models were tested for each group. The first model assumed the same item-factor assignment across the groups as described earlier. The second model assumed the same item-factor assignment, and the factor loadings were constrained to be equivalent across groups. Given the large sample size, the χ² values were not treated as absolute indicators of fit but rather used to evaluate the appropriateness of increasingly restricted models by considering changes in their values. For men (n = 2334) and women (n = 2702), these nested models had adequate goodness-of-fit (Table 3). Testing differences among both models revealed nonsignificant χ² values for changes in model fit. Evaluating all of the fit indices indicated that the model with equal factor loadings fitted the data best. Thus, the results indicate gender invariance of the GAD-7 factor structure.

To assess unidimensionality of the GAD-7 across different age groups, participants were assigned to 4 age groups, beginning with the first group who ranged from 14 to 25 years (n = 634). The second group ranged from 26 to 45 years (n = 1651), the third group from 46 to 65 years (n = 1687), and the final group consisted of people older than 65 years (n = 1064). [Other partitions of age did not change the interpretation of the results (results are available on request from PYH). Rationale for group assignment was to create meaningful groups of roughly equal sizes.] Again, both models had adequate goodness-of-fit (Table 3). Testing differences among all the models revealed significant χ² values for changes in model fit. Evaluating all of the fit indices indicated that the model with equal factor loadings fitted the data best. Thus, the results indicate age invariance of the GAD-7 factor structure at the level of a congeneric measure.

Construct Validity

The intercorrelations between the GAD-7 anxiety scale and the PHQ-2 depression scale,23,24 the Rosenberg Self Esteem Scale,25 the Questionnaire on Life Satisfaction,27 and the Resilience Scale were r = 0.64 (95% CI: 0.63–0.66), r = −0.43 (95% CI: −0.41 to −0.46), r = −0.34 (95% CI: −0.32 to −0.37), and r = −0.29 (95% CI: −0.26 to −0.31), respectively. Table 4 shows the association of GAD-7 sum scores with diverse demographic characteristics. As hypothesized, women had higher GAD-7 anxiety scores compared with men, the GAD-7 sum score increased with age with the highest levels in respondents aged between 45 and 65 years, and subjects living with a partner had lower GAD-7 scores compared with subjects not living with a partner. Also in accordance with the hypotheses, GAD-7 scores were lower in subjects with higher educational levels compared with subjects with lower educational levels, employed subjects had lower GAD-7 scores compared with unemployed subjects, and GAD-7 scores were lower in subjects with higher household income compared with subjects with lower household income. Of note, there was no difference in GAD-7 scores between German and non-German participants.

In the multiple linear regression analyses, unemployment (standardized β = 0.16; P < 0.001), household income (standardized β = −0.11; P < 0.001), gender (standardized β = 0.07; P < 0.001), urbanicity (standardized β = −0.04; P = 0.002), church membership (standardized β = −0.04; P = 0.02), and age (standardized β = 0.03; P = 0.04) were significant predictors of anxiety as measured with the GAD-7 sum score. Total variance explained was 6.3%. All results
were similar when models were rerun using the rank transformation of the GAD-7 sum score.

GAD-7 Scores in Diverse Samples

GAD-7 mean scores (95% CI), adjusted for age and gender, in the general population (N = 5030), in primary care (N = 956), and in patients with generalized anxiety disorder (N = 73) were 2.97 (95% CI: 2.86–3.07), 5.57 (95% CI: 5.33–5.81), and 14.18 (95% CI: 13.31–15.05), respectively. Analyses of covariance, controlled for age and gender, showed significant subgroup differences between the 3 samples (df = 4, 6054; F = 266; P < 0.001). In the general population sample, percentages (95% CI) of subjects with moderate (GAD-7 score: 10–14) and severe anxiety symptom levels (GAD-7 score: 15–21) were 4.1% (95% CI: 3.5–4.6) and 1.0% (95% CI: 0.7–1.3), respectively. In the primary care sample these, percentages increased to 14.4% (95% CI: 12.2–16.7) and 8.9% (95% CI: 7.1–10.7), respectively. Finally, in patients with the diagnosis of generalized anxiety disorder, percentages with moderate and severe levels of anxiety were 41.5% (95% CI: 39.8–43.2) and 48.1% (95% CI: 36.9–59.2), respectively.

GAD-7 Normative Data

The normative data for the different age levels and both genders are presented in Table 5. Percentiles from this table can be used to compare an individual subject’s GAD-7 score with those determined from the general population reference group based on age and gender. It could be said that for all subgroups, approximately 5% had GAD-7 scores of 10 or greater; corresponding to moderate and severe levels of anxiety, and approximately 1% had GAD-7 scores of 15 or greater, corresponding to a severe level of anxiety.

DISCUSSION

This study, including more than 5000 subjects, gives evidence that the GAD-7 is a reliable and valid self-report measure for anxiety in the general population. Specifically, results from the confirmatory factor analyses affirm the unidimensional structure of the GAD-7 in different groups of age and gender. The conclusion that the unidimensional structure of the GAD-7 appropriately reflects anxiety in different subject groups is supported by the fact that internal consistency was virtually identical for all subject groups. The intercorrelations of the GAD-7 with the PHQ-2 depression scale (r = 0.64), the Rosenberg Self-Esteem Scale (r = −0.43), the Questionnaire on Life Satisfaction (r = −0.34), and the Resilience Scale (r = −0.29) are similar to intercorrelations between these concepts in other studies and suggesting construct validity of the GAD-7. Construct validity is also supported by the extent to which the GAD-7 sum score was associated with known risk factors of generalized anxiety disorders, such as gender, age, educational level, partnership,
TABLE 4. Association of GAD-7 Scores With Demographic Characteristics (N = 5030)

<table>
<thead>
<tr>
<th></th>
<th>GAD-7 Score, M (SD)</th>
<th>Group Differences, P&lt;001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.66 (3.24)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.20 (3.52)</td>
<td></td>
</tr>
<tr>
<td>Age group, yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14–24</td>
<td>2.76 (3.49)</td>
<td>0.017</td>
</tr>
<tr>
<td>25–34</td>
<td>2.81 (3.33)</td>
<td></td>
</tr>
<tr>
<td>35–44</td>
<td>2.82 (3.34)</td>
<td></td>
</tr>
<tr>
<td>45–54</td>
<td>3.14 (3.43)</td>
<td></td>
</tr>
<tr>
<td>55–64</td>
<td>3.25 (3.60)</td>
<td></td>
</tr>
<tr>
<td>65–74</td>
<td>2.79 (3.23)</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td>3.05 (3.38)</td>
<td></td>
</tr>
<tr>
<td>Living with a partner</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>2.76 (3.22)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.23 (3.65)</td>
<td></td>
</tr>
<tr>
<td>Education, yr</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0–8</td>
<td>4.46 (4.92)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.13 (3.52)</td>
<td></td>
</tr>
<tr>
<td>10–11</td>
<td>2.85 (3.35)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.04 (3.17)</td>
<td></td>
</tr>
<tr>
<td>13–15</td>
<td>2.52 (2.95)</td>
<td></td>
</tr>
<tr>
<td>≥16</td>
<td>2.63 (2.98)</td>
<td></td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>In training</td>
<td>2.70 (3.35)</td>
<td></td>
</tr>
<tr>
<td>Full-time employment</td>
<td>2.43 (2.85)</td>
<td></td>
</tr>
<tr>
<td>Part-time employment</td>
<td>3.00 (3.07)</td>
<td></td>
</tr>
<tr>
<td>Military or civilian service</td>
<td>2.13 (2.36)</td>
<td></td>
</tr>
<tr>
<td>Unemployed/short-time work</td>
<td>5.65 (4.75)</td>
<td></td>
</tr>
<tr>
<td>Homemaker</td>
<td>3.47 (4.05)</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>2.98 (3.40)</td>
<td></td>
</tr>
<tr>
<td>Net household income</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&lt;750 €/mo</td>
<td>4.81 (4.50)</td>
<td></td>
</tr>
<tr>
<td>750–1249 €/mo</td>
<td>3.83 (3.83)</td>
<td></td>
</tr>
<tr>
<td>1250–1999 €/mo</td>
<td>2.80 (3.31)</td>
<td></td>
</tr>
<tr>
<td>≥2000 €/mo</td>
<td>2.51 (2.99)</td>
<td></td>
</tr>
<tr>
<td>Nationality</td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>German</td>
<td>2.95 (3.40)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2.90 (3.44)</td>
<td></td>
</tr>
<tr>
<td>Region in Germany</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>2.84 (3.31)</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>3.37 (3.73)</td>
<td></td>
</tr>
<tr>
<td>Urbanicity</td>
<td></td>
<td>0.002</td>
</tr>
<tr>
<td>Urban area</td>
<td>3.01 (3.43)</td>
<td></td>
</tr>
<tr>
<td>Rural area</td>
<td>2.57 (3.20)</td>
<td></td>
</tr>
<tr>
<td>Church member</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>2.84 (3.32)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.31 (3.61)</td>
<td></td>
</tr>
</tbody>
</table>

*Group differences were performed using t test and analysis of variance, respectively (N = 5030). Group differences that remained significant after Bonferroni-adjustment for multiple testing (P < 0.005) are printed in bold.

Another main result of this study is the standardization of the GAD-7 with the provision of normative data. Given that age and gender-specific comparative data were generated based on subgroups consisting of 141,566 subjects each, the sample size was sufficient to provide normative data for all subgroups. The homogeneity of GAD-7 scores across the subgroups and the expected associations of GAD-7 scores with gender and age, support the conclusion that the GAD-7 is applicable in men and women as well as in older and younger subjects.

The normative data enable users of the GAD-7 to discern whether an individual’s anxiety score is normal, or mildly, moderately, or severely elevated. With the rule of thumb that approximately 5% of subjects from the general population have GAD-7 scores of 10 or greater and that approximately 1% have GAD-7 scores of 15 or greater, GAD-7 scores are easily interpretable. For use in clinical practice, we recommend considering GAD-7 scores of ≥10 as “yellow flags” and GAD-7 scores of ≥10 as “red flags” for the presence of anxiety disorders. This procedure is supported by the primary care validation study, in which GAD-7 sum scores of ≥10 and ≥15 were associated with positive likelihood ratios for the presence of generalized anxiety disorder of 5.1 and 8.7, respectively. Of note, elevated GAD-7 scores are also associated with panic disorder, social anxiety disorder, posttraumatic stress disorder, and any anxiety disorder with the positive likelihood ratio for the presence of any anxiety disorder for GAD-7 scores of ≥10 being 5.5. Therefore, GAD-7 sum scores of ≥10 should be followed by an extended diagnostic evaluation for any anxiety disorder and the initiation of treatment as appropriate.

An important question, however, concerns the representativeness of the study sample and the generalizability of the study results to other countries. There are 2 reasons supporting the assumption that the results regarding validity and normative data of the GAD-7 may also apply to other developed countries, especially the United States. First, the demographic characteristics of our study population closely mirror the characteristics of both the German general population and those of the American general population. Therefore, the study population may be considered representative for the German general population and comparable to the American general population. Second, for a related measure, the PHQ-9, studies revealed similar distributions and cut-off scores for the original American and the German version of the instrument. Given the similar distribution of PHQ-9 depression scores in both countries, it seems unlikely that both countries substantially differ with respect to the distribution of GAD-7 anxiety scores.

This study has a number of strengths including its large sample size and the representativeness of the study sample. A potential limitation of our study is the response rate of 72.9%. However, response rates in general population studies are
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16.4</td>
<td>19.8</td>
<td>18.7</td>
<td>18.8</td>
<td>14.5</td>
<td>15.9</td>
<td>21.1</td>
<td>20.2</td>
<td></td>
<td>19.2</td>
<td>15.4</td>
<td>15.7</td>
<td>13.9</td>
<td>13.8</td>
<td>13.8</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>38.7</td>
<td>46.6</td>
<td>42.3</td>
<td>44.5</td>
<td>35.1</td>
<td>38.4</td>
<td>47.7</td>
<td>47.2</td>
<td></td>
<td>41.9</td>
<td>36.6</td>
<td>36.6</td>
<td>33.0</td>
<td>32.9</td>
<td>34.8</td>
<td>33.1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>50.4</td>
<td>59.8</td>
<td>54.2</td>
<td>57.3</td>
<td>47.9</td>
<td>51.0</td>
<td>57.9</td>
<td>59.9</td>
<td></td>
<td>50.2</td>
<td>48.1</td>
<td>46.9</td>
<td>45.1</td>
<td>44.6</td>
<td>47.3</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>61.2</td>
<td>70.1</td>
<td>67.0</td>
<td>68.3</td>
<td>59.7</td>
<td>61.4</td>
<td>66.9</td>
<td>69.1</td>
<td></td>
<td>60.0</td>
<td>59.6</td>
<td>58.5</td>
<td>56.9</td>
<td>55.3</td>
<td>57.9</td>
<td>54.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>70.5</td>
<td>77.5</td>
<td>76.4</td>
<td>77.5</td>
<td>69.2</td>
<td>69.1</td>
<td>74.7</td>
<td>76.6</td>
<td></td>
<td>67.7</td>
<td>70.2</td>
<td>69.5</td>
<td>67.2</td>
<td>64.8</td>
<td>67.3</td>
<td>62.8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>77.5</td>
<td>83.1</td>
<td>83.2</td>
<td>83.0</td>
<td>77.2</td>
<td>74.9</td>
<td>80.5</td>
<td>81.9</td>
<td></td>
<td>72.3</td>
<td>77.8</td>
<td>77.5</td>
<td>75.8</td>
<td>73.8</td>
<td>75.0</td>
<td>70.7</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>82.7</td>
<td>87.0</td>
<td>88.9</td>
<td>86.1</td>
<td>83.3</td>
<td>78.9</td>
<td>84.5</td>
<td>85.1</td>
<td></td>
<td>77.5</td>
<td>83.2</td>
<td>82.6</td>
<td>81.5</td>
<td>80.5</td>
<td>81.6</td>
<td>77.4</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>87.9</td>
<td>89.7</td>
<td>93.4</td>
<td>90.8</td>
<td>88.5</td>
<td>84.1</td>
<td>89.7</td>
<td>90.1</td>
<td></td>
<td>84.4</td>
<td>88.1</td>
<td>87.7</td>
<td>87.0</td>
<td>86.5</td>
<td>87.2</td>
<td>83.8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>91.9</td>
<td>92.2</td>
<td>95.6</td>
<td>94.7</td>
<td>92.7</td>
<td>88.8</td>
<td>94.2</td>
<td>93.6</td>
<td></td>
<td>89.0</td>
<td>91.0</td>
<td>91.9</td>
<td>91.6</td>
<td>90.8</td>
<td>91.0</td>
<td>90.2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>94.0</td>
<td>94.4</td>
<td>96.8</td>
<td>96.1</td>
<td>94.3</td>
<td>90.9</td>
<td>96.3</td>
<td>93.6</td>
<td></td>
<td>91.9</td>
<td>92.6</td>
<td>93.7</td>
<td>94.4</td>
<td>93.1</td>
<td>93.4</td>
<td>93.4</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>95.7</td>
<td>96.1</td>
<td>96.8</td>
<td>96.1</td>
<td>94.3</td>
<td>90.9</td>
<td>96.3</td>
<td>95.0</td>
<td></td>
<td>94.6</td>
<td>94.8</td>
<td>95.1</td>
<td>96.1</td>
<td>94.7</td>
<td>95.7</td>
<td>95.1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>96.8</td>
<td>97.5</td>
<td>98.1</td>
<td>98.5</td>
<td>96.6</td>
<td>95.5</td>
<td>98.2</td>
<td>96.1</td>
<td></td>
<td>96.2</td>
<td>96.1</td>
<td>95.9</td>
<td>96.7</td>
<td>95.9</td>
<td>96.9</td>
<td>96.8</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>97.6</td>
<td>98.1</td>
<td>98.1</td>
<td>98.9</td>
<td>97.2</td>
<td>97.0</td>
<td>98.7</td>
<td>96.8</td>
<td></td>
<td>97.5</td>
<td>96.8</td>
<td>96.7</td>
<td>97.6</td>
<td>97.0</td>
<td>97.7</td>
<td>97.6</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>98.2</td>
<td>98.5</td>
<td>99.0</td>
<td>98.9</td>
<td>97.7</td>
<td>98.3</td>
<td>99.2</td>
<td>97.9</td>
<td></td>
<td>98.3</td>
<td>97.3</td>
<td>97.3</td>
<td>98.5</td>
<td>97.7</td>
<td>98.3</td>
<td>98.5</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>98.7</td>
<td>98.8</td>
<td>99.7</td>
<td>99.4</td>
<td>98.3</td>
<td>98.3</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>98.7</td>
<td>97.5</td>
<td>98.1</td>
<td>98.5</td>
<td>98.4</td>
<td>98.8</td>
<td>99.4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>99.1</td>
<td>98.8</td>
<td>99.7</td>
<td>99.4</td>
<td>99.1</td>
<td>98.8</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.0</td>
<td>97.8</td>
<td>98.6</td>
<td>98.5</td>
<td>99.1</td>
<td>98.8</td>
<td>99.4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>99.3</td>
<td>99.2</td>
<td>99.7</td>
<td>99.9</td>
<td>99.1</td>
<td>99.3</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.4</td>
<td>98.4</td>
<td>98.9</td>
<td>98.5</td>
<td>99.4</td>
<td>99.4</td>
<td>99.4</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>99.5</td>
<td>99.5</td>
<td>99.7</td>
<td>99.9</td>
<td>99.1</td>
<td>99.3</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.4</td>
<td>99.1</td>
<td>99.2</td>
<td>99.0</td>
<td>99.4</td>
<td>99.4</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>99.6</td>
<td>99.5</td>
<td>99.7</td>
<td>99.9</td>
<td>99.1</td>
<td>99.6</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.8</td>
<td>99.1</td>
<td>99.2</td>
<td>99.0</td>
<td>99.4</td>
<td>99.6</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>99.7</td>
<td>99.5</td>
<td>99.7</td>
<td>99.9</td>
<td>99.6</td>
<td>99.9</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.8</td>
<td>99.5</td>
<td>99.6</td>
<td>99.2</td>
<td>99.6</td>
<td>99.6</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>99.8</td>
<td>99.5</td>
<td>99.7</td>
<td>99.9</td>
<td>99.6</td>
<td>99.9</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.8</td>
<td>99.9</td>
<td>99.8</td>
<td>99.7</td>
<td>99.9</td>
<td>99.9</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>99.9</td>
<td>99.8</td>
<td>99.7</td>
<td>99.9</td>
<td>99.9</td>
<td>99.9</td>
<td>99.7</td>
<td>99.3</td>
<td></td>
<td>99.8</td>
<td>99.9</td>
<td>99.8</td>
<td>99.7</td>
<td>99.9</td>
<td>99.9</td>
<td>99.8</td>
<td></td>
</tr>
</tbody>
</table>
generally lower compared with clinical studies and our response rate is similar to the response rate of 70.9% in the American NCS-R.\(^2\) In addition, given the agreement between the demographic characteristics of the study sample with those from the general population, selection bias seems unlikely. The only demographic variable that differed substantially between the participants and the German general population was the percentage of subjects with non-German nationality (3.2% vs. 8.8%). However, participants with and without German nationality did not differ significantly with respect to their GAD-7 anxiety score, making it improbable that our results regarding validity and normative data were influenced by this slight imbalance.

Given this large sample size, some readers may ask why the analyses presented here were not based on item response theory (IRT) analysis. IRT is another latent variable approach for examining measurement equivalence. Our favoring of the covariance approach is based on findings, which have shown that such analyses yields robust results, virtually identical to results obtained with IRT-based approaches.\(^3,4,9,50\)

Finally, validity testing of an instrument needs to demonstrate both convergent and discriminant validity. Although convergent and concurrent validity for the GAD-7 was already assessed in primary care,\(^1,19\) this large general population study did not include standard criterion measures to assess convergent and discriminant validity. Nevertheless, the size of the associations of the GAD-7 with measures of depression, self-esteem, life satisfaction, resilience, and the demographic risk factors of generalized anxiety disorder, suggest construct validity of the GAD-7 in the general population.

For the interpretation of the study results, it is important to consider that validity and reliability are not immutable characteristics of a measure. Our results, primarily generalizable to the general population, may, however, change when the GAD-7 is applied to different samples. Nevertheless, the results from this large general population study correspond well with results from earlier primary care studies in terms of reliability, factorial validity, and construct validity of the GAD-7.\(^3,19\) Therefore, there is growing evidence supporting the conclusion that the GAD-7 is an efficient and valid self-report anxiety measure for subjects from clinical and nonclinical settings.

ACKNOWLEDGMENTS

The authors thank all subjects for participating in our study.

REFERENCES


© 2008 Lippincott Williams & Wilkins