

Secondary Publication



Kliem, Sören; Sachser, Cedric; Lohmann, Anna; u. a.

Psychometric evaluation and community norms of the GAD-7, based on a representative German sample

Date of secondary publication: 25.08.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-109684x

Primary publication

Kliem, Sören; Sachser, Cedric; Lohmann, Anna; u. a. (2025): Psychometric evaluation and community norms of the GAD-7, based on a representative German sample, in: *Frontiers in psychology*, Lausanne: Frontiers Research Foundation, Vol. 16, Nr. 1526181, pp. 1–10, doi: 10.3389/fpsyg.2025.1526181.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



OPEN ACCESS

EDITED BY

Laura Badenes-Ribera,
University of Valencia, Spain

REVIEWED BY

Maria Rita Sergi,
University of G. d'Annunzio, Italy
Cesar Merino-Soto,
Monterrey Institute of Technology and Higher
Education (ITESM), Mexico

*CORRESPONDENCE

Sören Kliem
✉ soeren.kliem@eah-jena.de

†These authors have contributed equally to
this work and share first authorship

RECEIVED 11 November 2024

ACCEPTED 28 February 2025

PUBLISHED 20 March 2025

CITATION

Kliem S, Sachser C, Lohmann A, Baier D,
Brähler E, Fegert JM and Gündel H (2025)
Psychometric evaluation and community
norms of the GAD-7, based on a
representative German sample.
Front. Psychol. 16:1526181.
doi: 10.3389/fpsyg.2025.1526181

COPYRIGHT

© 2025 Kliem, Sachser, Lohmann, Baier,
Brähler, Fegert and Gündel. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Psychometric evaluation and community norms of the GAD-7, based on a representative German sample

Sören Kliem^{1*†}, Cedric Sachser^{2†}, Anna Lohmann^{1†}, Dirk Baier³,
Elmar Brähler^{4,5}, Jörg M. Fegert² and Harald Gündel⁵

¹Department of Social Welfare, Ernst-Abbe-Hochschule Jena - University of Applied Sciences, Jena, Germany, ²Department for Child and Adolescent Psychiatry/Psychotherapy, University Clinic for Psychosomatic Medicine and Psychotherapy Ulm, Ulm, Germany, ³Institute of Delinquency and Crime Prevention, Zurich University of Applied Sciences, Zurich, Switzerland, ⁴Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, ⁵Department of Medical Psychology and Sociology, Leipzig University, Leipzig, Germany

Introduction: The Generalized Anxiety Disorder-7 (GAD-7) scale is widely used to assess generalized anxiety symptoms in both clinical and general populations. However, updated psychometric evaluations and population norms for the German adult population are lacking. This study aims to examine the psychometric properties of the GAD-7 and provide representative population norms.

Methods: A representative sample of the adult population in Germany (N = 2,519) was assessed. Item characteristics (means, standard deviations, inter-item correlations) were examined. Construct validity was evaluated through correlations with the PHQ-9 and BSI-18. Internal consistency was assessed using coefficient omega. Confirmatory factor analysis (CFA) was conducted to test the one-factor model, and measurement invariance across gender and age was examined using multi-group CFA.

Results: The GAD-7 demonstrated strong internal consistency and construct validity. CFA supported the assumed one-factor model. Measurement invariance analyses indicated that the GAD-7 provides comparable measurements across gender and age groups. Updated population-based norms were established for the total sample and specific age groups.

Discussion: Findings confirm the GAD-7 as a psychometrically sound measure for generalized anxiety in the general population. The updated norms enhance its applicability in clinical and epidemiological research, supporting its use for screening and assessment across diverse demographic groups.

KEYWORDS

GAD-7, generalized anxiety, self-report questionnaire, population norms, psychometrics, measurement invariance

1 Introduction

Generalized anxiety disorder (GAD) is among the most prevalent mental disorders (Remes et al., 2016; Somers et al., 2006), yet it is generally thought to be underdiagnosed (Parmentier et al., 2013) and undertreated (Ruscio et al., 2017). Brief and effective screening tools are essential for identifying and assessing anxiety symptoms in both clinical and research settings. The Generalized Anxiety Disorder-7 (GAD-7) scale is widely used for this purpose, having been validated against diagnostic clinical interviews to establish its sensitivity, specificity, and predictive values (Spitzer et al., 2006). Originally developed for primary care settings, the GAD-7 has also been employed in epidemiological studies (e.g.,

Hinz et al., 2017; Johansson et al., 2013; Löwe et al., 2008) and has seen extensive use in public health monitoring, including the National Health Interview Survey (NHIS) (Terlizzi and Zablotsky, 2024) and COVID-19 research (Fancourt et al., 2021; McBride et al., 2021).

The GAD-7 has been translated into numerous languages, including Spanish, French, German, Chinese, Japanese, Korean, Arabic, Portuguese, Russian, and Turkish and validated across diverse patient populations (see Plummer et al., 2016, for an overview). However, despite its widespread use, research on normative scores remains limited (Hinz et al., 2017). Normative values are crucial for contextualizing individual and group distress levels. To date, only one large-scale normative study has been conducted in a general population sample in Germany (Löwe et al., 2008). Other general population studies confirm strong psychometric properties, including high internal consistency and factorial validity, supporting the one-dimensional structure of the scale (e.g., South Korea: Ahn et al., 2019), (Belgium: De Man et al., 2021), (Spain: Garcia-Campayo et al., 2010), (Brazil: Monteiro et al., 2022). Shevlin et al. (2020) established measurement invariance of the GAD-7 across four European countries (UK, Ireland, Spain, and Italy), further supporting its cross-cultural validity. A recent large-scale Japanese study provided valuable prevalence data but did not report psychometric properties or norms (Matsuyama et al., 2024). Moreover, the GAD-7 has demonstrated robust performance in web-based assessments (Donker et al., 2011).

Previous studies examining the psychometric properties of the GAD-7 in the German general population have primarily focused on its factorial structure, internal consistency, and criterion validity. Löwe et al. (2008) originally validated the German version, demonstrating strong internal consistency (Cronbach's $\alpha > 0.85$) and high correlations with related constructs such as depression (PHQ-2, $r = 0.64$, $p < 0.001$) and self-esteem (Rosenberg Self-Esteem Scale, $r = -0.43$, $p < 0.001$). They found CFA support for an one-factor model with high factor loadings ranging from 0.76 to 0.90. Hinz et al. (2017) reports the same high internal consistency and also confirmed the unidimensional structure of the scale via confirmatory factor analysis (CFA). They furthermore supported the scale's measurement invariance across age and gender. Both studies found consistently higher anxiety scores in women.

Population-level shifts in anxiety symptoms underscore the need for regularly updated normative data. Establishing community norms for the GAD-7 is essential for contextualizing individual and group scores, facilitating interpretation relative to the general population. Norms provide a reference framework to distinguish typical from elevated anxiety levels, aiding in both clinical and epidemiological applications (Spitzer et al., 2006; Löwe et al., 2008). They also enable comparisons across demographic groups, contributing to a more nuanced understanding of anxiety

prevalence and disparities (Hinz et al., 2017). Recent findings by Hinz et al. (2023) confirmed longitudinal measurement invariance over six years in Germany and identified a statistically significant increase in anxiety symptoms, emphasizing the necessity for contemporary reference values.

The present study evaluates the psychometric properties of the GAD-7 in a large, representative German community sample and provides updated norms. We test evidence based on internal structure by assessing its reliability and unidimensional factor structure and hypothesize that the GAD-7 demonstrates high internal consistency and configural, metric, and scalar invariance across gender and age. To establish evidence based on relations to other variables, we hypothesize strong positive associations with depressive symptoms and somatization and furthermore expect differences in anxiety levels across gender. Evidence based on test content and response processes is not the focus of this study, as the GAD-7 is a well-established measure with extensive prior validation. By addressing these aspects, this study provides a comprehensive update on the scale's psychometric properties and its interpretability in the general population.

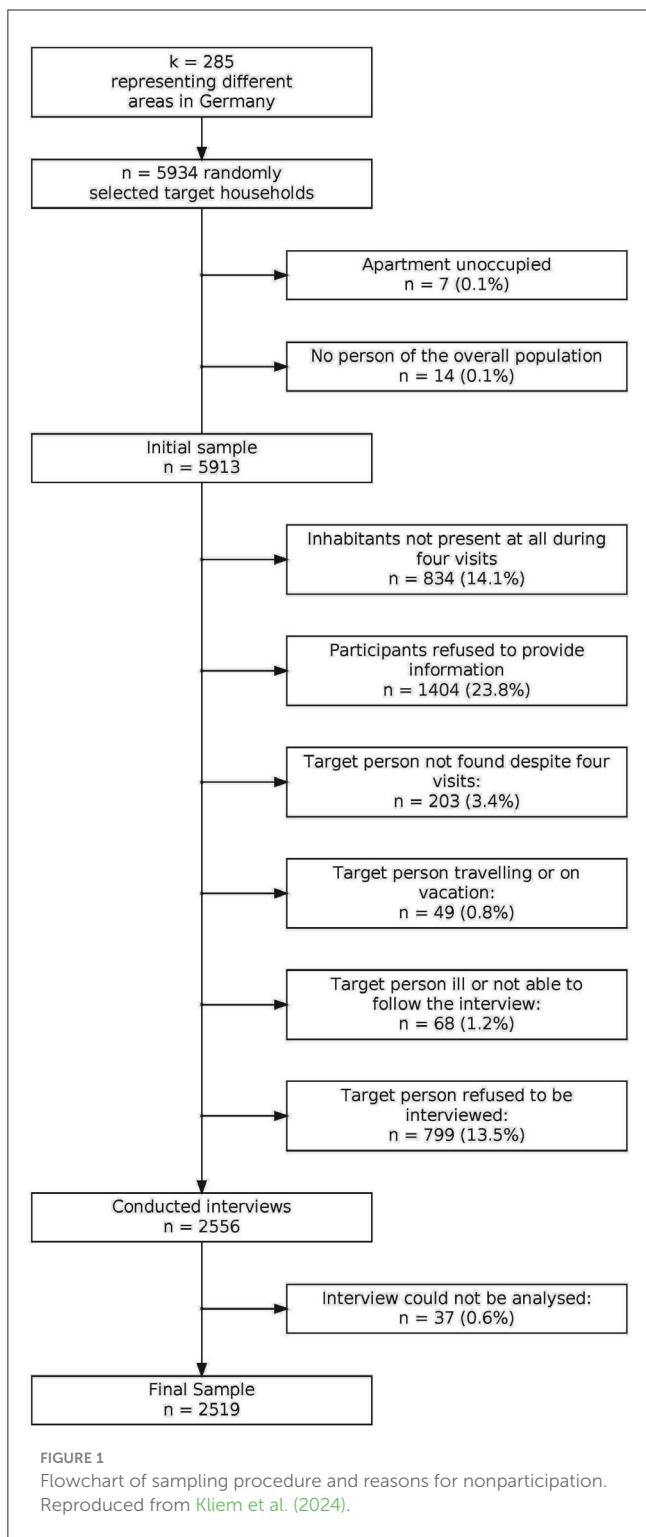
1.1 Procedure

The GAD-7 was presented as part of a large survey conducted by Leipzig University between December 2020 and March 2021. The survey was carried out by the contractor USUMA Markt- und Sozialforschung an independent institute for opinion and social research. The goals of the survey were (a) to assess prevalence rates of a variety of relevant physical or mental disorders and related risk behaviors (descriptive epidemiology), (b) to examine causes and conditions of these disorders (analytic epidemiology), and (c) to analyze psychometric properties and provide German population norms for clinical-psychological instruments. The survey consisted of two parts. The first part was interviewer guided and consisted of demographic as well as household information in accordance with the principles of the German Statistisches Bundesamt (Federal Statistical Office). In the second part the participants filled in paper-based questionnaires themselves in the presence of but independent as well as out of sight of the interviewer. The interviewer was, however, available for questions. Informed consent was obtained from all participants prior to the interview. Minimum age for participation was 16 years. For under-aged participants at least one legal guardian was informed about the sampling procedure and the contents of the survey. All participants were provided a written copy of the confidentiality agreement containing details regarding the handling of their personal information. The study was conducted in accordance with the Declaration of Helsinki. All procedures were approved by the Ethics Committee of the Medical Faculty of the University of Leipzig (Az.: 474/20-ek).

1.2 Sample description

The sampling was conducted according to the ADM sampling system F2F. This procedure consists of three steps. In a first step, the area of the federal republic of Germany is divided into regions of which 258 are sampled with sampling probability proportional

Abbreviations: PHQ-9, Patient Health Questionnaire; GAD-7, Generalized Anxiety scale; BSI GSI, Brief Symptom Inventory Global Severity Index; BSI Somatization, Brief Symptom Inventory Somatization Subscale; BSI Anxiety, Brief Symptom Inventory Anxiety Subscale; BSI Depression, Brief Symptom Inventory Depression Subscale; CFA, confirmatory factor analysis; CI, Confidence Interval; WLSMV, Weighted least square means and variance adjusted estimation; MI, measurement invariance; MGCFAs, multiple group factor analysis.



to the number of households. In a second step, 5,676 households are selected based on a random route procedure. The target person within each household is identified using a Kish selection grid. Details regarding response can be obtained from Figure 1. The following analyses are based on data from $N = 2,519$ participants which corresponds to a response rate of 42 %. Figure 1 presents a detailed overview of the sampling procedure and reasons for non-response. Sample descriptives can be found in Table 1.

1.3 Instruments

As the survey served multiple epidemiological purposes, only those measures that were used in the validation process are discussed in this paper. In addition to extensive demographic information (see Table 1), health related behavior, such as the number of sick days, doctor visits, and hospital stays, were assessed. The following measures were used for the validation of the scales at hand.

1.3.1 The general anxiety disorder scale (GAD-7)

The GAD-7 (Spitzer et al., 2006) is a brief self-report scale with seven items assessing generalized anxiety. Each of the seven items is rated on a scale from 0 (not at all) to 3 (almost every day). The total score of the GAD-7 scale ranges from 0 to 21. When formally validated using diagnostic psychiatric interviews, a GAD-7 score of ≥ 10 has been found to have a sensitivity of 89% and a specificity of 82% for detecting generalized anxiety disorder (Spitzer et al., 2006). The GAD-7 showed high internal consistency in previous general population studies ($\alpha = 0.89$; Löwe et al., 2008).

1.3.2 Patient health questionnaire (PHQ-9)

The PHQ-9 (Kroenke et al., 2001) is a self-report scale, that scores depression symptoms on a four-point scale from 0 (not at all) to 3 (almost every day), providing a total severity score ranging from 0 to 27. In the present study, the German version of the PHQ-9 (Martin et al., 2006) was used. To assess internal consistency within our sample, we computed Cronbach's alpha and McDonald's omega. The PHQ-9 demonstrated high internal consistency in the study at hand ($\alpha = 0.90$, 95% CI [0.89, 0.91]; $\omega = 0.93$, 95% CI [0.92, 0.94]), which aligns with findings from previous general population studies ($\alpha = 0.87$; Kocalevent et al., 2013).

1.3.3 The brief symptom inventory (BSI-18)

The BSI-18 (Derogatis and Fitzpatrick, 2004) is an 18-item short form of the Symptom-Checklist 90-R. It contains three subscales of six items each: somatization (SOMA), depression (DEPR), and anxiety (ANX). The sum score of all 18 items can furthermore be interpreted as a Global Severity Index (GSI). To evaluate internal consistency in our sample, we calculated Cronbach's alpha and McDonald's omega. The BSI-18 showed high reliability in the study at hand ($\alpha = 0.93$, 95% CI [0.92, 0.94]; $\omega = 0.93$, 95% CI [0.93, 0.94]), consistent with previous general population studies [$\alpha = 0.93$ (GSI), 0.82 (SOMA), 0.87 (DEPR), 0.84 (ANX); Franke et al., 2017].

1.4 Statistical analysis

1.4.1 Missing data

To account for missing data, we applied chained equation modeling (see van Buuren and Groothuis-Oudshoorn, 2011) using the following variables: gender, age, nationality, marital status, living with a partner as well as all items from the scales GAD-7, PHQ-9 and BSI-18 to estimate missing data (proportion of missing

TABLE 1 Demographic characteristics of the study sample.

	Male (N 1,193)	Female (N 1,322)	Diverse (N 4)	Total (N 2,519)
Age (years)				
Mean (SD)	50.1 (17.7)	50.5 (18.3)	44.8 (26.5)	50.3 (18.1)
Median [Min, Max]	52.0 [16.0, 96.0]	51.0 [16.0, 96.0]	41.5 [21.0, 75.0]	51.0 [16.0, 96.0]
Age categories				
16–24	102 (8.5%)	125 (9.5%)	2 (50.0%)	229 (9.1%)
25–34	190 (15.9%)	174 (13.2%)	0 (0%)	364 (14.5%)
35–44	174 (14.6%)	225 (17.0%)	0 (0%)	399 (15.8%)
45–54	195 (16.3%)	216 (16.3%)	0 (0%)	411 (16.3%)
55–64	244 (20.5%)	243 (18.4%)	1 (25.0%)	488 (19.4%)
65–74	190 (15.9%)	200 (15.1%)	0 (0%)	390 (15.5%)
75+	98 (8.2%)	139 (10.5%)	1 (25.0%)	238 (9.4%)
Nationality				
German	1,151 (96.5%)	1,271 (96.1%)	3 (75.0%)	2,425 (96.3%)
Not German	42 (3.5%)	48 (3.6%)	1 (25.0%)	91 (3.6%)
Missing	0 (0%)	3 (0.2%)	0 (0%)	3 (0.1%)
Marital status				
Married/living together	547 (45.9%)	527 (39.9%)	2 (50.0%)	1,076 (42.7%)
Married/separated	40 (3.4%)	25 (1.9%)	0 (0%)	65 (2.6%)
Single	398 (33.4%)	357 (27.0%)	2 (50.0%)	757 (30.1%)
Divorced	143 (12.0%)	227 (17.2%)	0 (0%)	370 (14.7%)
Widowed	62 (5.2%)	181 (13.7%)	0 (0%)	243 (9.6%)
Missing	3 (0.3%)	5 (0.4%)	0 (0%)	8 (0.3%)
Living with partner				
Living with partner	729 (61.1%)	737 (55.7%)	2 (50.0%)	1,468 (58.3%)
Not living with partner	444 (37.2%)	565 (42.7%)	2 (50.0%)	1,011 (40.1%)
Missing	20 (1.7%)	20 (1.5%)	0 (0%)	40 (1.6%)

data of the GAD-7 items: 0.30%–0.70%). To avoid implausible item values, the estimated values (\hat{y}) were corrected by predictive mean matching (i.e., the observable values closest to the predicted value were chosen). We used the R package mice (van Buuren and Groothuis-Oudshoorn, 2011) for imputation.

1.4.2 Item characteristics

Mean and standard deviations were obtained for all GAD-7 items in the total sample as well as for the sub samples of male and female gender. Effect sizes of group differences in item means were computed as Cohen's d and Cliff's δ . Item correlations were obtained. To assess item characteristics we examined the skewness and kurtosis, item difficulty (i.e. percentage of participants endorsing each item), item-total correlations, and Cronbach's alpha if the item was deleted. Item characteristics were obtained using the R package psych (William Revelle, 2024).

1.4.3 Construct validity

To assess construct validity of the GAD-7, we correlated the scale with the PHQ-9 and the three BSI-18 subscales (somatization, anxiety and depression) as well as with the BSI global severity index. The following hypotheses were formulated: anxiety levels should be higher in individuals with (a) higher depression scores (e.g. Byrd-Bredbenner et al., 2021; Hinz et al., 2017; Löwe et al., 2008), and (b) higher somatization scores (e.g. Gierk et al., 2014; Kliem et al., 2017a).

1.5 Population norms

Population norms are computed as cumulative percentiles. Female and male subsamples as well as several of age groups are tabulated separately. To avoid spurious jumps we also provide smoothed norms, which were obtained via a shape-constrained additive modeling (SCAM) approach. SCAM utilizes penalized

regression splines with built-in monotonicity constraints, ensuring that the resulting curves are non-decreasing—a critical property for cumulative distributions. The smoothing was implemented using the `scam` package in R (Pya, 2023).

1.5.1 Internal consistency

Considering that coefficient α could be affected by problems stemming from its assumptions not being met (McNeish, 2018), internal consistency of the GAD-7 is reported additionally as McDonald's ω which is computed using the R package `semTools` (Jorgensen et al., 2021). To evaluate the internal consistency of the GAD-7 scale across severity categories, McDonald's omega was computed using polychoric correlations and the `psych` package (William Revelle, 2024). This approach was chosen due to the challenges in obtaining reliable CFA-based omega estimates in subgroup analyses.

1.5.2 Factorial validity and measurement invariance

To test the one-dimensional structure of the GAD-7, confirmatory factor analyses (CFA) were conducted. CFAs were performed using the `lavaan` package (Rosseel, 2012) for R statistics. As suggested for the use with ordered categorical measures, weighted least square means and variance adjusted estimation (WLSMV) was used. Measurement invariance (MI) was tested using multiple group factor analysis (MG-CFA). Following the procedure suggested by Wu and Estabrook (2016) for ordered categorical variables, we used theta parameterization and identified the model by setting the means and variances of the latent factors to 0 and 1, respectively, item intercepts to 0 and residual variances to 1. The following models were subsequently tested: configural invariance (no constraints apart from those necessary for model identification), threshold invariance (constraining all thresholds to be equal), weak invariance (constraint of loadings), strong invariance (constraining of intercepts), and full invariance (constraining residual invariance). See Supplementary Figure A1 for a path model overview of the structural equation models assessed. Chen (2007) suggests the following cut-off criteria: a change of < -0.01 in CFI in addition to a change of ≥ 0.015 in RMSEA indicating non-invariance. We conducted MG-CFAs for the GAD-7 one-factor model across gender, age (below median age vs. above median age), as well as age \times gender. Due to the low number of individuals classifying as neither male nor female, these cases were not included in the MG-CFAs of gender and age \times gender. MI analyses were conducted using the `semTools` package (Jorgensen et al., 2021) for R statistics.

1.5.3 Sensitivity analysis—Careless response patterns

To evaluate the robustness of the results, a sensitivity analysis was conducted by excluding respondents with inconsistent response patterns. Outliers were identified based on G+ scores, a measure of Guttman errors, which occur when respondents endorse harder items but not easier ones, violating the expected hierarchical order of items (Van der Ark, 2012). The G+

scores were calculated using the `check.errors()` function from the `Mokken` package in R (Van der Ark, 2012). Outliers were defined using Tukey's method, with scores exceeding the upper fence ($Q3+3\times IQR$) flagged as discordant (Tukey, 1977). A total of 87 outliers (3.5% of the sample) were excluded, as inconsistent response patterns can bias psychometric results and distort conclusions. Analyses were repeated without these cases to assess the sensitivity of the findings.

2 Results

2.1 Item characteristics

Supplementary Table A3 displays means and standard deviations for the seven items of the GAD-7 in the total sample as well as effect sizes for mean differences regarding gender. On the item-level there was a consistent pattern of female participants exhibiting higher mean scores as well as higher variability on most GAD-7 items. Effect sizes (Cohen's d) of these mean differences ranged from $d = -0.02$ $[-0.1, 0.05]$ to $d = -0.16$ $[-0.24, -0.08]$. Item characteristics can be obtained from Supplementary Table A6. Supplementary Table A7 presents GAD-7 item-level associations with PHQ-9 and BSI-18 subscales. Supplementary Figure A3 displays the distribution of responses across the seven GAD-7 items.

2.2 Construct validity

To determine evidence of construct validity of the GAD-7 correlation coefficients were calculated with related instruments. In line with our hypotheses, there were high positive correlations between the GAD-7 and measures of somatization, anxiety and depression as assessed by BSI-18 sub scales (see Supplementary Table A5). In the same vein, the PHQ-9 assessing depression showed positive correlations with the GAD-7.

Supplementary Table A7 presents the associations between individual GAD-7 items and the PHQ-9, as well as the Global Severity Index, Somatization, Anxiety, and Depression subscales of the BSI-18, using both the Maximum Information Coefficient (MIC) and Spearman's rank correlation (ρ_s). As expected, all items show positive associations with these scales, with the strongest relationships observed with the PHQ-9 rather than the anxiety subscale of the BSI-18. This pattern may reflect the substantial overlap between generalized anxiety and depressive symptoms. Additionally, the stronger association with the PHQ-9 may be partly attributable to the identical response scaling of both measures.

2.3 Population norms

Prevalences according to the severity cut-offs suggested by Spitzer et al. (2006) can be found in Table 2.

Table 3 shows cumulative percentiles of GAD-7 scores for the total sample. Additional norms split by gender as well as age group

can be found in the [Supplementary Tables A9, A11](#). Smoothed norms are also available in the [Supplementary Tables A8, A10, A12](#).

2.4 Internal consistency

Cronbach's alpha of the GAD-7 for the full sample was $\alpha = 0.9$, 95% CI (0.89, 0.91). McDonald's omega of the GAD-7 for the full sample was $\omega = 0.92$, 95% CI (0.91, 0.93).

The reliability estimates suggest that the GAD-7 exhibits strong internal consistency in the minimal ($\omega = 0.78$) and severe ($\omega = 0.72$) categories. However, the reliability is notably lower in the mild ($\omega = 0.53$) and moderate ($\omega = 0.65$) categories. This pattern may

reflect differences in item functioning or response variability across severity groups.

In the minimal and severe groups, participants may exhibit more uniform response patterns, either due to consistently low or consistently high anxiety levels, leading to higher internal consistency. In contrast, the mild and moderate groups might display greater variability in responses, potentially reflecting more heterogeneity in the manifestation of anxiety symptoms within these groups.

2.5 Factorial validity

A CFA was conducted to assess the unidimensional structure of the GAD-7. The fit indices indicated reasonable model fit, with a robust CFI of 0.936, a robust TLI of 0.905, and an SRMR of 0.033. However, the robust RMSEA was 0.176 (90% CI [0.157, 0.196]), suggesting some misfit.

To improve the model, we inspected modification indices and introduced residual correlations between items with overlapping content: #2 (Not able to stop/control worrying) with #3 (Worrying too much). These adjustments improved the fit indices (CFI = 0.961, RMSEA = 0.143).

TABLE 2 Percentage of participants per severity level based on [Spitzer et al. \(2006\)](#) GAD-7 cut-offs.

GAD 7 severity	Total, %	Male, %	Female, %
Minimal (0–4)	81.30	83.40	79.50
Mild (5–9)	14.37	12.82	15.73
Moderate (10–14)	3.29	2.77	3.78
Severe (15–21)	1.03	1.01	0.98

TABLE 3 Population based norms (cumulative percentiles) of the GAD-7 scores (total sample).

GAD-7	Total	Age 16–24	Age 25–34	Age 35–44	Age 45–54	Age 55–64	Age 65–74	Age 75+
0	47.8	53.3	54.9	51.9	43.3	43.0	47.7	42.0
1	60.5	61.6	68.7	63.4	60.1	54.7	61.0	54.2
2	70.1	65.5	75.0	71.7	69.6	66.6	72.3	68.5
3	76.7	71.2	79.1	78.4	76.2	77.0	77.4	73.9
4	81.3	76.0	84.6	83.2	81.3	80.5	82.1	78.6
5	85.9	84.7	87.4	87.0	86.1	84.6	86.9	84.0
6	89.4	88.2	91.8	90.0	90.5	86.9	90.5	87.8
7	93.2	93.0	95.3	92.0	94.6	92.4	94.1	89.9
8	94.7	93.9	97.5	92.7	96.1	94.1	95.4	92.0
9	95.7	95.6	98.1	93.7	97.1	95.3	95.6	93.7
10	96.7	96.5	98.1	95.0	98.3	96.3	96.7	95.4
11	97.5	96.9	99.2	95.5	98.5	97.1	97.7	97.9
12	98.0	96.9	99.5	96.2	98.8	97.1	98.2	99.6
13	98.5	98.7	99.7	97.0	99.0	97.5	98.7	>99.9
14	99.0	99.1	>99.9	97.2	99.8	98.2	99.2	>99.9
15	99.2	99.6	>99.9	98.2	99.8	98.2	99.5	>99.9
16	99.4	>99.9	>99.9	98.5	>99.9	98.4	99.7	>99.9
17	99.6	>99.9	>99.9	99.0	>99.9	98.6	>99.9	>99.9
18	99.7	>99.9	>99.9	99.2	>99.9	99.2	>99.9	>99.9
19	99.8	>99.9	>99.9	99.5	>99.9	99.4	>99.9	>99.9
20	99.8	>99.9	>99.9	99.7	>99.9	99.4	>99.9	>99.9
21	>99.9	>99.9	>99.9	>99.9	>99.9	>99.9	>99.9	>99.9

Despite these modifications, strong standardized factor loadings (0.82–0.91, a path diagram can be found in [Supplementary Figure A2](#)) of the speak in favor of a one factor solution.

2.6 Measurement invariance

The fit measures obtained in the measurement invariance analyses of the GAD-7 are presented in [Supplementary Table A2](#). Adequate fit was observed across all levels of invariance testing, with CFI values consistently ≥ 0.998 and RMSEA values decreasing as model constraints were added. For gender, configural invariance showed a good fit (CFI = 0.998, RMSEA = 0.072), and full invariance was supported (CFI = 0.998, RMSEA = 0.045) with negligible changes in model fit (CFI = 0, RMSEA ≤ -0.009). Similarly, for age groups, configural invariance yielded CFI = 0.998 and RMSEA = 0.077, with full invariance confirmed (CFI = 0.998, RMSEA = 0.051, CFI = 0, RMSEA ≤ -0.01). When testing invariance across both gender and age simultaneously, model fit remained stable (configural: CFI = 0.999, RMSEA = 0.074; full invariance: CFI = 0.998, RMSEA = 0.047, CFI = 0, RMSEA ≤ -0.013).

3 Discussion

The present study investigates the psychometric quality of the GAD-7 using a large and representative sample of the German general population. Based on coefficient ω , the GAD-7 demonstrates high internal consistency, reinforcing its reliability as a screening tool. The lower reliability observed in the mild and moderate severity categories may indicate greater heterogeneity in symptom manifestation within these groups. This aligns with prior research suggesting that individuals with moderate anxiety levels may experience a broader range of symptoms or more nuanced symptom profiles (e.g., [Terlizzi and Zablotzky, 2024](#)). In contrast, participants in the minimal and severe groups likely exhibit more consistent response patterns, reflecting either an absence or a pronounced intensity of symptoms. These findings highlight the importance of further examining item functioning and variability across the severity spectrum to ensure the robustness of the GAD-7 across diverse populations.

Furthermore, measurement invariance testing using MGCFAs confirmed comparable factor structures across gender and age groups. The model fit statistics indicated that the assumed factor structure of the GAD-7 holds across these subgroups, supporting its utility for comparative analyses across demographic variables. This provides validity evidence based on internal structure, following the framework outlined in the “Standards” ([American Educational Research Association et al., 2014](#)). Additionally, the observed correlations between the GAD-7 and other established measures of psychological distress (PHQ-9, BSI-18) align with prior research, further supporting validity evidence based on relationships with other variables ([Löwe et al., 2008](#)). Overall, the results affirm that the GAD-7 is an efficient, reliable, and valid instrument for assessing generalized anxiety in the German population.

A key contribution of this study is the provision of updated norm tables, the primary aim of this research. These percentiles (see [Table 3, Supplementary Tables A8–A11](#)) are stratified by age and available in both gender-specific and gender-neutral formats. A clinical cut-off of 10 points has been suggested in previous research ([Spitzer et al., 2006](#)), which, in our sample, corresponds to the 96th–99th percentiles. This indicates symptom severity well above the population mean, ranging from 1.75 to 2.33 standard deviations above the mean. It is important to note that these normative data are provided as reference values for clinical interpretation and are not intended to serve as definitive clinical norms. Compared to the norms presented by [Hinz et al. \(2017\)](#), which were derived from a sample in Leipzig (a large city in eastern Germany, population: 600,000), the present study benefits from a sample representative of the entire Federal Republic of Germany. Substantially lower item means were observed across all items except for item 7 (“Feeling afraid”), where endorsement rates were similar. Given socio-structural differences between eastern and western Germany, as well as between urban and rural regions, it is conceivable that the previous norms ([Hinz et al., 2017](#)) do not generalize to the entire German population.

The percentiles obtained in this study deviate considerably from prior normative studies ([Erhardt et al., 2022](#); [Hinz et al., 2017](#); [Löwe et al., 2008](#)). A notably high proportion of participants reported minimal generalized anxiety symptoms. This decline in anxiety levels aligns with recent population-level data from [Erhardt et al. \(2022\)](#), based on data from 2014–2019, which reported the following prevalence rates: minimal (0–4): 74.9.

Several large surveys assessing the GAD-7 have been conducted in Germany ([Bäuerle et al., 2020](#); [Skoda et al., 2021](#); [Streit et al., 2022](#)) and other European countries ([Hyland et al., 2021](#); [José et al., 2023](#)). While GAD symptom burden varied across these studies, all reported substantially higher anxiety levels than the present study. Notably, these studies employed large online convenience samples and framed their surveys around pandemic-related mental health concerns. Given the potential for self-selection bias in such studies, the present results, derived from a representative face-to-face survey, provide an important counterbalance to prevailing narratives.

Mean GAD-7 scores in this study ($M = 2.18$, $SD = 3.28$) were substantially lower than those observed in recent nationally representative samples from other countries [e.g., UK: 5.25 (5.68), Ireland: 5.03 (5.52), Spain: 5.86 (5.24), Italy: 5.73 (5.14)] ([Shevlin et al., 2022](#)). This aligns with global epidemiological data from [Ruscio et al. \(2017\)](#), which identified Germany as having a relatively low prevalence of generalized anxiety disorder compared to other nations. Taken together with findings on internal consistency, these results underscore the need for careful monitoring of item wording to ensure that the GAD-7 captures the nuances of anxiety symptoms across different severity levels.

These findings also emphasize the necessity of regularly updating norms and caution against assuming stable prevalence rates over time. For example, [Gottschick et al. \(2023\)](#) found elevated anxiety symptoms in the German population during the onset of the Russo-Ukrainian war compared to the COVID-19 pandemic period. This underscores the importance of continuously revising normative data to ensure clinical relevance. Our study provides a

critical reference point for longitudinal research, offering insights into symptom variability at the population level.

3.1 Practical implications

The findings of this study hold significant practical implications for both clinical and research settings. First, the provision of updated norm tables allows for improved screening and diagnostic decision-making by healthcare professionals. By offering percentile-based reference values, clinicians can better interpret individual GAD-7 scores relative to the broader population, aiding in the identification of individuals at risk for generalized anxiety disorder. We emphasize that while these norms serve as valuable reference data in clinical contexts, they are derived from a representative general population sample and should not be used as definitive clinical norms. Second, the confirmation of measurement invariance across gender and age groups ensures that the GAD-7 can be reliably used in diverse demographic settings, facilitating comparisons across subpopulations. Third, given the observed decline in anxiety symptoms and divergence from prior norms, this study highlights the importance of considering contemporary population trends when applying psychological assessments. Researchers and practitioners should remain cautious about relying on outdated reference values, particularly when tracking prevalence changes over time. Finally, these results contribute to the broader discourse on mental health epidemiology by emphasizing the role of socio-structural factors in shaping anxiety prevalence. Future research should continue to examine how economic, political, and public health events influence mental health trajectories at the population level.

3.2 Limitations

Despite the many strengths of this study (especially the representativeness of the sample), certain limitations must be mentioned. The response rate is only 42.6%. However, lower response rates than in clinical studies are quite common in general population studies, and the response rate of this study was comparable to similar surveys (e.g. Kliem et al., 2015, 2018, 2017b). Although considerable efforts have been made to maximize the representativeness of the sample, a certain level of non-response is inevitable within the current design, raising concerns about potential bias stemming from this non-response. Unfortunately, a systematic evaluation of non-response bias is not feasible due to the lack of demographic information on those who did not respond. Such an assessment would require access to registry data, which is not readily available in Germany without government authorization. Furthermore, the diagnostic efficiency of the GAD-7 could not be examined because no additional clinical interviews were conducted.

3.3 Conclusion

In summary, the German GAD-7 has proven to be reliable and valid instruments for the use in different frameworks. Based on the

(potentially COVID-19 related) change in symptom burden which the norm values reported in this study indicate we suggest to update the norms again in the near future to see whether the pandemic has to be interpreted as an interlude or “the new normal”.

Data availability statement

The datasets presented in this article are not readily available because ethics review board approval did not include data sharing. Requests to access the datasets should be directed to Harald Gündel, harald.guendel@uniklinik-ulm.de.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Medical Faculty of the University of Leipzig (Az.: 474/20-ek). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

SK: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. CS: Conceptualization, Writing – review & editing. AL: Formal analysis, Writing – original draft, Writing – review & editing. DB: Writing – review & editing. EB: Conceptualization, Data curation, Funding acquisition, Project administration, Writing – review & editing. JF: Conceptualization, Data curation, Funding acquisition, Project administration, Writing – review & editing. HG: Conceptualization, Data curation, Funding acquisition, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The research did not receive specific funding, but was performed as part of the employment of the authors. SK and AL Ernst-Abbe Hochschule University of Applied Sciences Jena, CS and JF University Clinic Ulm, DB Zurich University of Applied Sciences, and EB University Mainz. The funders were neither involved in manuscript writing, editing, approval, or decision to publish. Open access funding by Zurich University of Applied Sciences (ZHAW).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1526181/full#supplementary-material>

References

- Ahn, J.-K., Kim, Y., and Choi, K.-H. (2019). The Psychometric Properties and Clinical Utility of the Korean Version of GAD-7 and GAD-2. *Front. Psychiat.* 10:127. doi: 10.3389/fpsyg.2019.00127
- American Educational Research Association American Psychological Association and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Bauerle, A., Teufel, M., Musche, V., Weismüller, B., Kohler, H., Hetkamp, M., et al. (2020). Increased generalized anxiety, depression and distress during the COVID-19 pandemic: a cross-sectional study in Germany. *J. Public Health* 42, 672–678. doi: 10.1093/pubmed/fdaa106
- Byrd-Bredbenner, C., Eck, K., and Quick, V. (2021). GAD-7, GAD-2, and GAD-mini: Psychometric properties and norms of university students in the United States. *Gen. Hosp. Psychiatry* 69, 61–66. doi: 10.1016/j.genhosppsych.2021.01.002
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Model.* 14, 464–504. doi: 10.1080/10705510701301834
- De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., et al. (2021). Are the PHQ-9 and GAD-7 suitable for use in India? a psychometric analysis. *Front. Psychol.* 12:676398. doi: 10.3389/fpsyg.2021.676398
- Derogatis, L. R., and Fitzpatrick, M. (2004). *The SCL-90-r, the brief symptom inventory (BSI), and the BSI-18*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Donker, T., Van Straten, A., Marks, I., and Cuijpers, P. (2011). Quick and easy self-rating of generalized anxiety disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Res.* 188, 58–64. doi: 10.1016/j.psychres.2011.01.016
- Erhardt, A., Gelbrich, G., Klinger-König, J., Streit, F., Kleinedam, L., Riedel-Heller, S. G., et al. (2022). Generalised anxiety and panic symptoms in the German National Cohort (NAKO). *World J. Biol. Psychiat.* 24, 881–896. doi: 10.1080/15622975.2021.2011409
- Fancourt, D., Steptoe, A., and Bu, F. (2021). Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. *Lancet Psychiat.* 8, 141–149. doi: 10.1016/S2215-0366(20)30482-X
- Franke, G. H., Jaeger, S., Glaesmer, H., Barkmann, C., Petrowski, K., and Braehler, E. (2017). Psychometric analysis of the brief symptom inventory 18 (BSI-18) in a representative German sample. *BMC Med. Res. Methodol.* 17:14. doi: 10.1186/s12874-016-0283-3
- García-Campayo, J., Zamorano, E., Ruiz, M. A., Pardo, A., Perez-Paramo, M., Lopez-Gomez, V., et al. (2010). Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. *Health Qual. Life Outc.* 8:8. doi: 10.1186/1477-7525-8-8
- Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. A., Murray, A. M., et al. (2014). Assessing somatic symptom burden: a psychometric comparison of the patient health questionnaire 15 (PHQ-15) and the somatic symptom scale-8 (SSS-8). *J. Psychosom. Res.* 78, 352–355. doi: 10.1016/j.jpsychores.2014.11.006
- Gottschick, C., Diexer, S., Massag, J., Klee, B., Broda, A., Purschke, O., et al. (2023). Mental health in Germany in the first weeks of the Russo-Ukrainian war. *BJPsych Open* 9:e66. doi: 10.1192/bjo.2023.21
- Hinz, A., Esser, P., Friedrich, M., Glaesmer, H., Mehnert-Theuerkauf, A., Schroeter, M. L., et al. (2023). Changes in anxiety in the general population over a six-year period. *PLoS ONE* 18:e0291206. doi: 10.1371/journal.pone.0291206
- Hinz, A., Klein, A. M., Brähler, E., Glaesmer, H., Luck, T., Riedel-Heller, S. G., et al. (2017). Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *J. Affect. Disord.* 210, 338–344. doi: 10.1016/j.jad.2016.12.012
- Hyland, P., Shevlin, M., Murphy, J., McBride, O., Fox, R., Bondjers, K., et al. (2021). A longitudinal assessment of depression and anxiety in the Republic of Ireland before and during the COVID-19 pandemic. *Psychiatry Res.* 300:113905. doi: 10.1016/j.psychres.2021.113905
- Johansson, R., Carlbring, P., Heedman, A. A., Paxling, B., and Andersson, G. (2013). Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. *PeerJ* 1:e98. doi: 10.7717/peerj.98
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2021). “Tools: Useful tools for structural equation modeling,” in *R Package Version 0.5–5*.
- José, H., Oliveira, C., Costa, E., Matos, F., Pacheco, E., Nave, F., et al. (2023). Anxiety and depression in the initial stage of the COVID-19 outbreak in a Portuguese sample: exploratory study. *Healthcare* 11:659. doi: 10.3390/healthcare11050659
- Kliem, S., Lohmann, A., Klatt, T., Mößle, T., Rehbein, F., Hinz, A., et al. (2017a). Brief assessment of subjective health complaints: Development, validation and population norms of a brief form of the Giessen Subjective Complaints List (GSCL-8). *J. Psychosom. Res.* 95, 33–43. doi: 10.1016/j.jpsychores.2017.02.003
- Kliem, S., Lohmann, A., Mößle, T., and Brähler, E. (2017b). German beck scale for suicide ideation (BSS): psychometric properties from a representative population survey. *BMC Psychiatry* 17:389. doi: 10.1186/s12888-017-1559-9
- Kliem, S., Lohmann, A., Mößle, T., and Brähler, E. (2018). Psychometric properties and measurement invariance of the beck hopelessness scale (BHS): results from a german representative population sample. *BMC Psychiatry* 18:110. doi: 10.1186/s12888-018-1646-6
- Kliem, S., Mößle, T., Rehbein, F., Hellmann, D. F., Zenger, M., and Brähler, E. (2015). A brief form of the perceived social support questionnaire (f-SozU) was developed, validated, and standardized. *J. Clin. Epidemiol.* 68, 551–562. doi: 10.1016/j.jclinepi.2014.11.003
- Kliem, S., Sachser, C., Lohmann, A., Baier, D., Brähler, E., Gündel, H., et al. (2024). Psychometric evaluation and community norms of the PHQ-9, based on a representative German sample. *Front. Psychiatry* 15:1483782. doi: 10.3389/fpsyg.2024.1483782
- Kocalevent, R.-D., Hinz, A., and Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *Gen. Hosp. Psychiatry* 35, 551–555. doi: 10.1016/j.genhosppsych.2013.04.006
- Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.0160090606.x
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., et al. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Med. Care* 46, 266–274. doi: 10.1097/MLR.0b013e318160d093
- Martin, A., Rief, W., Klalberg, A., and Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen. Hosp. Psychiatry* 28, 71–77. doi: 10.1016/j.genhosppsych.2005.07.003
- Matsuyama, S., Otsubo, T., Nomoto, K., Higa, S., and Takashio, O. (2024). Prevalence of generalized anxiety disorder in japan: a general population survey. *Neuropsychiat. Dis. Treatm.* 20, 1355–1366. doi: 10.2147/NDT.S456272
- McBride, O., Murphy, J., Shevlin, M., Gibson-Miller, J., Hartman, T. K., Hyland, P., et al. (2021). Monitoring the psychological, social, and economic impact of the COVID-19 pandemic in the population: Context, design and conduct of the longitudinal COVID-19 psychological research consortium (c19prc) study. *Int. J. Methods Psychiatr. Res.* 30:1. doi: 10.1002/mpr.1861

- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychol. Methods* 23, 412–433. doi: 10.1037/met0000144
- Monteiro, R. P., Nascimento, B. S., Monteiro, T. M. C., Da Silva, P. D. G., and Ferreira, A. J. C. (2022). Psychometric evidence of the 7-item generalized anxiety disorder questionnaire in Brazil. *Int. J. Ment. Health Addict.* 20, 1023–1034. doi: 10.1007/s11469-020-00423-9
- Parmentier, H., Garc-a-Campayo, J., and Prieto, R. (2013). Comprehensive review of generalized anxiety disorder in primary care in europe. *Curr. Med. Res. Opin.* 29, 355–367. doi: 10.1185/03007995.2013.770731
- Plummer, F., Manea, L., Trepel, D., and McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *Gen. Hosp. Psychiatry* 39:24–31. doi: 10.1016/j.genhosppsych.2015.11.005
- Pyra, N. (2023). “Scam: shape constrained additive,” in *R Package Version 1.2–13*.
- Remes, O., Brayne, C., van der Linde, R., and Lafortune, L. (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain Behav.* 6:e00497. doi: 10.1002/brb3.497
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Ruscio, A. M., Hallion, L. S., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., et al. (2017). Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. *JAMA Psychiatry* 74:465. doi: 10.1001/jamapsychiatry.2017.0056
- Shevlin, M., Butter, S., McBride, O., Murphy, J., Gibson-Miller, J., Hartman, T. K., et al. (2022). Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic. *BMC Psychiatry* 22:154. doi: 10.1186/s12888-022-03787-5
- Shevlin, M., McBride, O., Murphy, J., Miller, J. G., Hartman, T. K., Levita, L., et al. (2020). Anxiety, depression, traumatic stress and COVID-19-related anxiety in the UK general population during the COVID-19 pandemic. *BJPsych Open* 6:e125. doi: 10.1192/bjo.2020.109
- Skoda, E.-M., Spura, A., De Bock, F., Schweda, A., Dörrie, N., Fink, M., et al. (2021). Vernderung der psychischen Belastung in der COVID-19-Pandemie in Deutschland: ngste, individuelles Verhalten und die Relevanz von Information sowie Vertrauen in Behrden. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 64, 322–333. doi: 10.1007/s00103-021-03278-0
- Somers, J. M., Goldner, E. M., Waraich, P., and Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: A systematic review of the literature. *Can. J. Psychiat.* 51, 100–113. doi: 10.1177/070674370605100206
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166:1092. doi: 10.1001/archinte.166.10.1092
- Streit, F., Zillich, L., Frank, J., Kleineidam, L., Wagner, M., Baune, B. T., et al. (2022). Lifetime and current depression in the German National Cohort (NAKO). *World J. Biol. Psychiat.* 24, 865–880. doi: 10.1080/15622975.2021.2014152
- Terlizzi, E. P., and Zablotzky, B. (2024). “Symptoms of anxiety and depression among adults: United states, 2019 and 2022,” in *Technical Report 213*. Hyattsville, MD: National Center for Health Statistics.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Boston: Reading/Addison-Wesley.
- van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* 45, 1–67. doi: 10.18637/jss.v045.i03
- Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *J. Stat. Softw.* 48, 1–27. doi: 10.18637/jss.v048.i05
- William Revelle (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.
- Wu, H., and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika* 81, 1014–1045. doi: 10.1007/s11336-016-9506-0