General Hospital Psychiatry 60 (2019) 50–56

Contents lists available at ScienceDirect

General Hospital Psychiatry

journal homepage: www.elsevier.com/locate/genhospsych

Research paper

Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis

Zhicheng Luo[a,1], Yuqian Li[b,1], Yitan Hou[c], Haiqing Zhang[a], Xiaotian Liu[a], Xinling Qian[a], Jingjing Jiang[a], Yan Wang[a], Xue Liu[a], Xiaokang Dong[a], Dou Qiao[a], Fang Wang[d,*], Chongjian Wang[a,*]

[a] Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou, Henan, PR China
[b] Department of Clinical Pharmacology, School of Pharmaceutical Science, Zhengzhou University, Zhengzhou, Henan, PR China
[c] Department of Global Health, School of Health Sciences, Wuhan University, Wuhan, Hubei, PR China
[d] Department of Epidemiology, School of Public Health, Shanxi Medical University, Taiyuan, Shanxi, PR China

ARTICLE INFO

Keywords:
Generalized anxiety disorder
GAD-2
Chinese rural population
Meta-analysis

ABSTRACT

Objective: The two-item generalized anxiety disorder scale (GAD-2) has been investigated in different populations, but with limited evaluation in the Chinese population. This study aimed to validate the diagnostic accuracy of GAD-2 for identifying GAD through a validation study along with an updated meta-analysis.
Methods: We recruited 694 adults in 2015 from "the Henan Rural Cohort study" where the GAD-7 questionnaire was adopted as the gold standard diagnosis. Receiver operating characteristics (ROC) curve and Cronbach's α were determined. Furthermore, a meta-analysis was conducted to evaluate the effectiveness of the GAD-2.
Results: In this study, 37 patients [5.33%; 95% confidence interval (CI): 3.78–7.27%] met the criteria for current GAD. The GAD-2 showed a Cronbach's alpha of 0.806 and an area under the ROC curve (AUC) of 0.954 (95% CI: 0.936–0.968). At a cutoff of 3, GAD-2 had highest Youden's index of 0.845, with a sensitivity of 0.865 and a specificity of 0.980. A total of 13 studies were included in the meta-analysis. The pooled sensitivity, specificity and diagnostic odds ratio (DOR) were 0.80, 0.82 and 17.81, respectively, at the optimal cutoff of 3.
Conclusion: GAD-2 has acceptable properties for identifying GAD at a cutoff of 3 in the Chinese rural population.

1. Introduction

Anxiety disorders are the most common mental disorders globally [1], with a twelve-month prevalence of 2.4% to 18.2% in different regions [2]. Studies have demonstrated that anxiety disorders are the sixth leading cause of disability worldwide [3] and accounted for 15% of the disability-adjusted life years in 2010 [4]. Generalized Anxiety Disorder (GAD), characterized by chronic, excessive and uncontrollable worry or fearfulness about events [5], is one of the most common of all anxiety disorders [6]. GAD tends to coexist with other mental disorders, most often major depressive disorder [7], resulting in severe and debilitating consequences. In addition, there is evidence that GAD is associated with reduced quality of life [8], impaired occupational function [9] and increased risk of cardiovascular diseases [10]. At the population level early screening for GAD is important for targeted treatment and prevention [11].

However, only a minority of patients (36%) with anxiety are recognized through primary care [12]. One reason for an under-diagnosis of GAD might be that anxiety scale tools are time-consuming and expensive [13]. Consequently, brief patient-completed scales are needed for use in busy clinical settings or as part of comprehensive health questionnaires [14]. The two-item generalized anxiety disorder scale

Abbreviations: GAD-2, the two-item generalized anxiety disorder scale; GAD, generalized anxiety disorder; ROC, receiver operating characteristic; CI, confidence interval; AUC, area under the receiver operating characteristics curve; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition; PRISMA, the Preferred Reporting Items for Systematic Reviews and Meta-Analysis; CIDI, Composite International Diagnostic Interview; SCID, Structured Clinical Interview for Diagnostic Statistical Manual; QUADAS-2, Quality Assessment Tool for Diagnostic Accuracy Studies-2; SD, standard deviation; PPV, positive predictive values; NPV, negative predictive values; PLR, positive likelihood ratios; NLR, negative likelihood ratios; DOR, diagnostic odds ratios; sROC, summary ROC

* Corresponding authors at: Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, 100 Kexue Avenue, Zhengzhou 450001, Henan, PR China.
E-mail addresses: wfang@sxmu.edu.cn (F. Wang), tjwcj2005@126.com (C. Wang).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.genhosppsych.2019.07.008
Received 19 February 2019; Received in revised form 10 July 2019; Accepted 11 July 2019

(GAD-2) was initially developed by Kroenke, Spitzer and their colleagues in 2007 and is representative of the "ultra-short screening tools" (1–4 items, < 2 min to complete) [15]. Over the past decade GAD-2 has been widely applied to identify GAD in primary care settings [16,17] and in the general population [18–20], with reasonably good psychometric properties in different populations. However, to date no study has been conducted to investigate the validity of the GAD-2 scale in a Chinese rural population.

The aim of the present study was to test the reliability and validity of the GAD-2 scale in a Chinese rural population, and secondly, to systematically review the accuracy of the GAD-2 questionnaire with comparison to a validation study.

2. Methods

2.1. Setting and participants

The current study was conducted as part of "the Henan Rural Cohort study". The large population-based project was carried out in Yuzhou, Suiping, Xinxiang, Tongxu, and Yima counties of Henan province in China from July 2015 to October 2017, and was registered in the Chinese Clinical Trial Register (Registration number: ChiCTR-OOC-15006699) before the onset of patient enrollment. The target population was permanent residents aged 18–79 years who lived in the five rural areas. Using a multistage, stratified cluster sampling method for sample selection, the Henan Rural Cohort study eventually included 39,259 participants aged 18–79 years old with a response rate of 93.7%.

For this study the "gold standard" test for generalized anxiety (GAD-7) was administered only to participants in Suiping county to reduce the burden on all participants while still providing a sufficient sample for this analysis. This study took place between July and August of 2015. A standardized questionnaire was completed by participants in all Henan counties to collect detailed information on demographic characteristics, personal and family history of diseases and medication, lifestyle factors and GAD-2 information. Demographic characteristics included age, gender, education level (primary school or below, junior high school, senior high school or above), marital status (married/cohabiting, widowed/single/divorced), and *per capita* monthly income (< 500, 500~, and ≥ 1000 Renminbi (RMB)). Lifestyle factors included current cigarette smoking, current alcohol consumption, and physical activity (low, moderate, high). Personal history of chronic diseases focused on three main chronic diseases: coronary heart disease, type 2 diabetes mellitus and hypertension. All participants who completed the initial survey were invited to respond to a more detailed questionnaire asking about any recent psychological issues. From this group 694 participants were included in the current study, yielding sufficient responses to proceed with the study (the minimum required sample size for statistical analysis was 623).

This study was approved by the Zhengzhou University Life Science Ethics Committee (Code: [2015] MEC (S128)), and informed consent was obtained from all participants. The study was conducted according to the 1975 Declaration of Helsinki.

2.2. GAD-7 and GAD-2

The Chinese version of GAD-7 is a seven-item self-reported screening instrument based on the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV) criteria for the rapid detection of GAD [21]. It assesses the frequency of self-reported anxiety symptoms (not at all = 0, several days = 1, more than half the days = 2, and nearly every day = 3) over the prior two weeks. Total scores for GAD-7 range from 0 to 21 and higher scores indicate more severe anxiety symptoms. At a cutoff of 10, GAD-7 shows greatest accuracy for identifying GAD, with a sensitivity of 86.2% and specificity of 95.5% [21]. Therefore, this cutoff was adopted to identify GAD in the present study.

GAD-2 is an ultra-short screening tool that includes the first two items of GAD-7 (feeling nervous, anxious, or on edge and not being able to stop or control worrying) [15]. Total scores for GAD-2 range from 0 to 6. A cutoff of 3 or greater is recommended in the general population to screen for GAD [18].

2.3. Sample size calculation

The sensitivity of the GAD-2 instrument was considered to calculate the estimated sample size needed for statistical reliability. According to the results of Kroenke et al. [15], a sensitivity of 85% is expected for GAD-2. Referring to the calculation method described by Flahault et al. [22], at least 33 diagnoses are required to ensure that the minimum lower limit of the 95% confidence interval (CI) would not be < 60%. Assuming a prevalence of GAD in the Chinese population of 5.3% [23], a total of 623 persons were needed for this study.

2.4. Statistical analysis

Continuous variables are expressed as mean ± standard deviation (SD) and categorical variables are presented as numbers and corresponding proportions. Statistical comparisons between persons with and without GAD utilized the Student's *t*-test, the Mann-Whitney *U* test or a chi-squared test, as appropriate. Cronbach's α was calculated to determine the internal consistency of the GAD-2 [26]. Discriminant validity was tested by comparing the differences in GAD-2 total scores between the GAD group and the non-GAD group. With regard to consistency between GAD-2 and GAD-7, McNemar's test was used to compare the proportions of GAD diagnosis by the two GAD scales and a Pearson product-moment correlation coefficient was calculated to determine the correlation between the total scores of GAD-2 and GAD-7. Sensitivity, specificity, positive and negative predictive values (PPV, NPV) and likelihood ratios (PLR, NLR) for various cutoffs were analyzed using a receiver operating characteristics (ROC) curve. The area under the curve (AUC) was used to measure the overall accuracy of the GAD-2 tool. The best cutoff was determined by balancing sensitivity and specificity in order to maximize the Youden's index (sensitivity + specificity – 1) [27]. All statistical analyses were performed using the SAS 9.1 software package (SAS Institute, USA) with two-tailed tests where $P < 0.05$ was considered statistically significant.

2.5. Meta-analysis

A meta-analysis was carried out based on the guideline of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [24]. The PRISMA checklist is included in Supplementary Table 1. The previous literature on screening for GAD by GAD-2 was searched in PubMed, Web of Science, Wanfang and China National Knowledge Infrastructure (all years to December 31, 2018) library databases using the terms "GAD-2" and "two-item generalized anxiety disorder" without any qualifier. References from the retrieved articles were also collected as additional candidates to ensure that the search was comprehensive.

All studies that reported the accuracy of GAD-2 for screening GAD were selected. The included studies must provide sufficient information to generate a 2 × 2 table. Only studies that defined GAD according to standard diagnostic instruments were included; such instruments included but were not limited to the Composite International Diagnostic Interview (CIDI) or the Structured Clinical Interview for Diagnostic Statistical Manual (SCID).

Two reviewers evaluated the included studies and extracted the data independently. The Quality Assessment Tool for Diagnostic Accuracy Studies-2 (QUADAS-2) was used as a quality assessment criterion [25]. The tool assessed risk of bias and applicability of the individual studies to the review question in four domains: Patient Selection, Index Test, Reference Standard and Flow and Timing. Accuracy

Table 1
Characteristics of subjects (N = 694).

Variables	Total (n = 694)	Non-GAD (n = 657)	GAD[a] (n = 37)	P[b]
Age (years), mean ± SD	47.33 ± 11.860	47.35 ± 11.938	47.03 ± 10.508	0.872
Gender, n (%)				0.001
Male	304 (43.8)	298 (45.4)	6 (16.2)	
Female	390 (56.2)	359 (54.6)	31 (83.8)	
Marital status, n (%)				0.748
Married/cohabiting	643 (92.7)	609 (92.7)	34 (91.9)	
Widowed/single/divorced	51 (7.3)	48 (7.3)	3 (8.1)	
Education level, n (%)				0.196
Primary school or below	88 (12.7)	80 (12.2)	8 (21.6)	
Junior high school	381 (54.9)	361 (54.9)	20 (54.1)	
Senior high school or above	225 (32.4)	216 (32.9)	9 (24.3)	
Per capita monthly income, n (%)				0.939
<500	209 (30.1)	197 (29.9)	12 (32.5)	
500~	221 (31.8)	210 (32.0)	11 (29.7)	
≥1000	264 (38.0)	250 (38.1)	14 (37.8)	
Current smoking[c], n (%)				0.026
No	159 (22.9)	156 (23.7)	3 (8.1)	
Yes	535 (77.1)	501 (76.3)	34 (91.9)	
Current alcohol use[d], n (%)				0.205
No	141 (20.3)	137 (20.9)	4 (10.8)	
Yes	553 (79.7)	520 (79.1)	33 (89.2)	
Physical activity[e], n (%)				0.004
Low	176 (25.4)	167 (25.4)	9 (24.3)	
Moderate	344 (49.6)	318 (48.4)	26 (70.3)	
High	174 (25.1)	172 (26.2)	2 (5.4)	
Chronic diseases history[f], n (%)				0.319
No	374 (53.9)	357 (54.3)	17 (45.9)	
Yes	320 (46.1)	300 (45.7)	20 (54.1)	
BMI (kg/m^2), mean ± SD	24.29 ± 3.381	24.29 ± 3.392	24.36 ± 3.229	0.898
GAD-2 score, mean ± SD	0.62 ± 1.207	0.41 ± 0.790	4.19 ± 1.681	<0.001

Abbreviations: SD, standard deviation; BMI, body mass index; GAD-2, the two-item generalized anxiety disorder scale.
[a] Diagnosis from GAD-7.
[b] T-test or chi-square test was performed for the comparison of variables.
[c] A person who smoked more than one cigarette per day in the past six months was defined as current smoker.
[d] A person who consumed alcoholic drinks twelve or more times in the past year was defined as current alcohol user.
[e] The classification of physical activity was based on the international physical activity questionnaire (IPAQ 2001).
[f] The information of chronic diseases history was collected from self-report of subjects.

data was recorded to implement contingency tables of the reported cutoffs. Divergences between the two investigators were resolved by a discussion or a third assessor.

Stata Software Package, V11.2 (Stata Corp, College Station, Texas, USA) was used to execute the meta-analysis of the available literature by random-effects bivariate meta-analysis methods [28]. Thus, pooled estimates and corresponding 95% CI of sensitivity, specificity, PLR, NLR and diagnostic odds ratios (DOR) were produced for each cutoff. Heterogeneity was assessed using the I^2 statistic based on DOR [29]. Sufficient studies were found (n > 10 for a cutoff of 3) to conduct a meta-regression to further explore sources of heterogeneity in the data [30]. In addition, publication bias was explored by funnel plots [31].

3. Results

3.1. Participant characteristics

A total of 694 participants were enrolled in the present study and 37 patients [5.33%; 95% confidence interval (CI): 3.78–7.27%] met the criteria for current GAD according to the GAD-7 questionnaire. Demographic characteristics of the study population are presented in Table 1. The study population consisted of 304 males (43.8%) and 390 females (56.2%). The mean age of the participants was 47.3 ± 11.9 years (ranging from 18 to 77). Compared to participants without GAD, those diagnosed with GAD were more likely to be female, be current smokers and have a low-level of exercise (all P < 0.05).

3.2. Reliability and comparison analysis

Comparative psychometric properties of GAD-2 as a case-finding tool for GAD were tested. The results show that the GAD-2 tool has robust internal consistency with a Cronbach's alpha value of 0.806. The corrected correlations between the total scores of the GAD-2 and each item were 0.916 and 0.914 (both P < 0.001), respectively. The correlation between the two items of the GAD-2 was 0.675 (P < 0.001).

A comparison between the GAD-2 and GAD-7 tools was conducted. Using the recommended GAD-7 cutoff of 10 or greater, we divided the population into a GAD group and a non-GAD group. Application of the GAD-2 tool showed that the total scores of the GAD-2 assessment differed significantly between the two groups (P < 0.001), indicating good discriminant validity for the GAD-2 questionnaire. Regarding consistency between GAD-2 and GAD-7, the proportion of GAD diagnoses with the GAD-2 tool (6.48%) was not statistically different from the number obtained with the GAD-7 tool (5.33%) (P = 0.096). In addition, there was also a robust correlation between total scores of the GAD-2 and GAD-7 (r = 0.894, P < 0.001).

3.3. Operating characteristics of the GAD-2

Table 2 summarizes the operating characteristics of the GAD-2 tool at various cutoffs. Unsurprisingly, the sensitivity declined and the specificity increased in a continuous fashion as the cutoff increased. At a cutoff of 3 the GAD-2 tool yielded the highest Youden's index of 0.845, with a sensitivity of 0.865 and a specificity of 0.980 respectively.

Table 2
Operating characteristics of the GAD-2 at various cutoffs.

Cutoffs	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	PLR	NLR
≧1	94.6 (81.8–99.3)	74.3 (70.8–77.6)	17.2 (15.2–19.4)	99.6 (98.6–99.9)	3.68 (3.2–4.3)	0.07 (0.02–0.3)
≧2	91.9 (78.1–98.3)	86.9 (84.1–89.4)	28.3 (24.1–33.0)	99.5 (98.5–99.8)	7.02 (5.6–8.7)	0.09 (0.03–0.3)
≧3	86.5 (71.2–95.5)	98.0 (96.6–98.9)	71.1 (58.6–81.1)	99.2 (98.3–99.7)	43.71 (25.1–76.0)	0.14 (0.06–0.3)
≧4	73.0 (55.9–86.2)	99.5 (98.7–99.9)	90.0 (74.1–96.6)	98.5 (97.5–99.1)	159.81 (50.8–502.7)	0.27 (0.2–0.5)
≧5	43.2 (27.1–60.5)	99.9 (99.2–100.0)	94.1 (68.6–99.2)	96.9 (95.9–97.6)	294.11 (38.7–2084.6)	0.59 (0.5–0.8)

PPV: positive predictive value; NPV: negative predictive value; PLR: positive likelihood ratio; NLR: negative likelihood value: AUC: area under the ROC curve.

GAD-2 was accurate in 71.1% (PPV) of participants with a positive screen and 99.2% (NPV) of participants that did not screen positive.

An ROC curve of the GAD-2 questionnaire was used to determine the sensitivity and specificity at various cutoffs (data not shown). In addition, it suggested that the AUC was 0.954 (95% CI: 0.936–0.968).

3.4. Meta-analysis

A total of 13 studies [15–20,32–38] met the inclusion and exclusion criteria (see Supplementary Fig. 1). Three of the included studies validated GAD-2 as a screening test in the general population [18–20]; five in primary care services [15–17,32,33] and five in samples from the hospital [34–38]. The mean age of the included participants ranged from 26.8 to 62.7 years. Within the 13 included studies, the prevalence of GAD diagnosed by an accepted standard test ranged from 4.0% to 77.0%. The details are shown in Table 3.

The methodological quality of the included studies was assessed using the QUADAS-2 tool. The results of the quality assessment are shown in Supplementary Table 2. The main risk of bias identified for these studies included not selecting patients consecutively or randomly, not avoiding inappropriate exclusions, not reporting multiple cutoffs for pre-specified thresholds, and a lack of information on blinding of assessments.

Studies reported cutoffs ranging from ≥1 to ≥6 for this measure. Sufficient studies were found at cutoffs 2–5 ($n \geq 4$) to conduct a meta-analysis. Table 4 shows a summary of the results of the meta-analysis. A cutoff of 3 had the highest sensitivity and specificity balance, with an area under the summary ROC (sROC) curve of 0.88 (95% CI: 0.85–0.91); however, the heterogeneity among studies was high (I^2 = 62.8%). To reduce the heterogeneity to an acceptable level, a sensitivity analysis was performed. After removing two outlying studies [17,36] from the analysis, the heterogeneity decreased markedly from 62.8% to 37.0% (see Supplementary Table 3). The removal of the outliers slightly improved the pooled specificity from 0.82 (95% CI: 0.72–0.90) to 0.83 (95% CI: 0.71–0.91) and decreased the pooled sensitivity from 0.80 (95% CI: 0.67–0.89) to 0.78 (95% CI: 0.61–0.89). The sROC curves at a cutoff of 3 are displayed in Supplementary Fig. 2, and indicate no significant change in the AUC (0.88 *vs.* 0.88) after adjusting for outliers.

In this meta-analysis the UK group displayed a lower mean age and the Korean group showed a higher proportion of women [35,36], compared with other included studies. Therefore, another sensitivity analysis excluding these two studies was conducted. After removing these two studies from the analysis, the heterogeneity decreased from 62.8% to 57.2%, and the pooled sensitivity and specificity improved from 0.80 (95% CI: 0.67–0.89) to 0.83 (95% CI: 0.69–0.92) and 0.82 (95% CI: 0.72–0.90) to 0.83 (95% CI: 0.70–0.91), respectively (see Supplementary Table 4). Simultaneously, the AUC of the sROC curves at a cutoff of 3 also increased from 0.88 to 0.90 (see Supplementary Fig. 3). According to previous studies, the diagnostic performance of GAD-2 varies in different settings. Therefore, a subgroup analysis by settings (general population and primary care *vs.* hospital settings) was conducted to estimate accuracy. DOR in hospital settings (DOR =15.16; 95% CI: 5.98–38.41) was lower than that in the general population and primary care (DOR = 19.66; 95% CI: 10.75–35.95). However, heterogeneity in the two subgroups remained high (general population and primary care I^2 = 63.8%; hospital settings I^2 = 66.0%). Supplementary Table 5 shows a detailed comparative summary of diagnostic properties of GAD-2 at a cutoff of 3 between the two subgroups.

A meta-regression was performed given the heterogeneity between studies. The possible sources of heterogeneity included mean age, proportion of women, consecutive or random sample, and an interval of two weeks or less. There was no evidence for publication bias according to Deeks funnel plot results (see Supplementary Fig. 4).

4. Discussion

4.1. Main findings

To the best of our knowledge, this is the first study investigating the validity of the ultra-short GAD-2 as a screening instrument for GAD in a Chinese rural population. The results show that the GAD-2 screening tool had robust internal consistency as indicated by a Cronbach's alpha value (α = 0.806) above the conventional threshold of 0.70 [39]. Youden's index for GAD-2 reached a maximum (0.845) at a cutoff of 3, with a sensitivity of 0.865 and specificity of 0.980. The area under the ROC curve (AUC) was 0.954 (95% CI: 0.936–0.968), which indicated that the GAD-2 questionnaire was extremely useful based on conventional guidelines where AUC values ≥0.90 indicate high accuracy [40].

Meta-analysis also revealed that the pooled sensitivity and specificity values were acceptable at a cutoff of 3 [sensitivity: 0.80 (95% CI: 0.67–0.89), specificity: 0.82 (95% CI: 0.72–0.90)]. At that cutoff GAD-2 had reasonably good accuracy, with an area under the sROC curve of 0.88 (95% CI: 0.85–0.91). The stratified meta-analysis by setting showed higher diagnostic performance in the general population and primary care compared with hospital settings (sensitivity: 0.83 *vs.* 0.80, specificity: 0.84 *vs.* 0.83, DOR: 19.66 *vs.* 15.16). These results, however, should be interpreted with caution due to the high levels of heterogeneity.

4.2. Comparison with existing literature

Previous studies have validated GAD-2 as a screening instrument for GAD in different populations. The original validation study for GAD-2 was conducted in primary care patients and reported an AUC of 0.908, a sensitivity of 86%, and a specificity of 83% at a cutoff of 3 [15]. In an Australian population-based study, the sensitivity was 57.6% and the specificity was 86.3% with a cutoff of 3 or more [18]. A German study conducted in elderly people aged 58 to 82 revealed that the sensitivity was 67% and the specificity was 90% at the optimal cutoff of 2 [19]. In a Dutch population-based study, a web-based GAD-2 showed a sensitivity of 83% and a specificity of 61% with the best cutoff at 4 [20]. Overall, the sensitivity and specificity in the current study were higher than those obtained in the above studies at the respective best cutoffs. Language differences may contribute to the differing best cutoff values in these studies [41].

Additionally, a meta-analysis that included 13 studies was conducted to systematically review the accuracy of GAD-2. Plummer et al. reported a diagnostic meta-analysis of GAD-2 in 2015 and demonstrated its acceptable properties for identifying GAD [42]. Our meta-analysis is similar to Plummer's in some aspects such as the search

Table 3
Characteristics of included studies.

Authors	Country	Age (mean and range)	Gender (% female)	Setting	Sample size	Number of patients meeting diagnostic criteria for GAD (%)	Index test: best cutoff	Index test: cutoffs reported	Language	Reference test	Publish year
Kroenke et al. [15]	USA	47.1 (18–87)	69	Primary care	965	73 (7.6)	3	GAD-2: ≥2–3	English	SCID	2007
Kujanpaa et al. [16]	Finland	62.7 (range not reported)	69.3	Primary care	150	6 (4.0)	3	GAD-2: ≥2–4	Finnish	MINI	2014
Garcia-Campayo et al. [17]	Spain	47.59 (19–85)	72.6	Primary care	212	106 (50)	3	GAD-2: ≥1–6	Spanish	MINI	2012
Christensen et al. [18]	Australia	Mean not reported (18–65)	63.5	General population	326	33 (10.2)	3	GAD-2: ≥3	English	MINI	2011
Wild et al. [19]	Germany	Mean not reported (58–82)	55	General population	438	27 (6.2)	2	GAD-2: ≥1–6	German	SCID	2014
Donker et al. [20]	The Netherlands	43 (18–80)	57	General population	157	30 (19)	4	GAD-2: ≥2–6	Dutch	CIDI	2011
Heyningen et al. [32]	South Africa	26.8 (18–48)	100	Primary care	376	86 (22.9)	2	GAD-2: ≥2	English, isiXhosa, Afrikaans	MINI	2018
Cano-Vindel et al. [33]	Spain	44.0 (range not reported)	70.2	Primary care	178	137 (77.0)	3	GAD-2: ≥1–6	Spanish	CIDI	2018
Micoulaud-Franchi et al. [34]	France	39.38 (range not reported)	63.4	Clinical neurophysiology department	145	49 (33.8)	2	GAD-2: ≥2	French	MINI	2017
Seo et al. [35]	Korea	40.0 (16–65)	86.3	Outpatients headache clinic	146	32 (21.9)	2	GAD-2: ≥1–3	Korean	MINI	2015
Delgadillo et al. [36]	UK	35 (23–54)	23	Community drugs treatment service	103	31 (30)	3	GAD-2: ≥2–5	English	CIS-R	2012
Qian et al. [37]	China	Not reported	Not reported	Hospital outpatient	300	16 (5.3)	3	GAD-2: ≥3	Chinese	MINI	2011
Wang et al. [38]	China	60.8 (18–87)	46.3	Psycho-cardiological outpatient	201	74 (36.8)	3	GAD-2: ≥2–5	Chinese	CIDI	2014

Abbreviations: GAD, general anxiety disorder; CIDI, Composite International Diagnostic Interview; CIS-R, Clinical Interview Schedule Revised; SCID, Structured Clinical Interview for DSM Disorders; MINI, Mini International Neuropsychiatric Interview.

Table 4
GAD-2 for identifying GAD: heterogeneity and pooled estimates of sensitivity, specificity, PLR, NLR and DOR by cutoffs.

Cutoffs	Number of studies	Pooled sample size	Sensitivity (95% CI)	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)	Heterogeneity: I^2 (%)
2	11	3071	0.94 (0.84–0.98)	0.59 (0.43–0.73)	2.29 (1.75–3.01)	0.10 (0.05–0.23)	16.07 (8.89–29.02)	62.7
3	11	3176	0.80 (0.67–0.89)	0.82 (0.72–0.90)	4.25 (2.90–6.21)	0.24 (0.15–0.39)	17.81 (10.88–29.15)	62.8
4	7	1439	0.70 (0.57–0.81)	0.88 (0.76–0.94)	5.49 (2.86–10.51)	0.34 (0.24–0.49)	15.65 (7.40–33.07)	76.3
5	6	1289	0.43 (0.29–0.58)	0.94 (0.85–0.98)	6.20 (2.71–14.14)	0.61 (0.49–0.75)	10.31 (4.75–22.39)	66.8

PLR: positive likelihood ratio; NLR: negative likelihood value: DOR: area under the ROC curve.

strategy, the exclusion criteria, and the quality assessment criteria (QUADAS-2). However, our meta-analysis updated the literature search from March, week 2, 2014, to December 31, 2018 and extended the overall library databases by including two Chinese databases (CNKI and Wanfang). As a result, seven additional reports were included in the current meta-analysis along with the six reports of Plummer's study. With these additional studies we were able to carry out a more comprehensive analysis including a sensitivity analysis, subgroup analyses and meta-regression. Both the earlier and the current meta-analyses show similar results with no differences in the sensitivity (0.80 *vs.* 0.80), specificity (0.81 *vs.* 0.82) and DOR (17.52 *vs.* 17.81).

A comparison between the validation study and the updated meta-analysis was also performed. Although it could be argued that the meta-analysis should only include studies in China or rural China for best comparability, there are insufficient Chinese studies ($n < 4$) to support this limited meta-analysis. The result of our larger meta-analysis suggests that the accuracy of GAD-2 for identifying GAD is best at a cutoff of 3 as found in the validation study. In addition, the sensitivity in the validation study was similar to that in the meta-analysis while the specificity was significantly higher, which indicated a lower false positive rate in the validation study.

4.3. Limitations

There are several limitations to be noted in the validation study. Firstly, it might be inappropriate that a Chinese version of the GAD-7 scale was used as the standard for GAD diagnosis. Although GAD-7 is not currently recognized as a standard diagnostic instrument, a recent meta-analysis revealed that GAD-7 had acceptable psychometric properties for identifying GAD [42]. In addition, GAD-7 has been used as a reference test in the general population [43] and in patient groups [44]. The close correspondence between GAD-2 and GAD-7 suggests that GAD-2 would perform comparably against other criteria like the CIDI, although future work is required to confirm this hypothesis.

Another limitation is possible selection bias due to the default selection of literate participants who were able to read the questionnaire. The included samples were mostly well educated and different outcomes might arise if the samples were less educated. Whether GAD-2 might perform better in less educated participants requires further research.

For the meta-analysis in this study, the most notable limitation was the bias from the selective reporting of cutoffs. It is possible that researchers tend to report more information on the best cutoff rather than those cutoffs that had low sensitivity and specificity. However, a diagnostic meta-analysis is based on a different number of studies for each cutoff, each of which included a different population and had different methodologic characteristics.

4.4. Implications and recommendations

As an ultra-brief screening tool, GAD-2 could be useful and valuable in certain circumstances. For instance, when busy primary care [16,17] or large population studies [18,19] are undertaken, GAD-2 might be quite suitable for saving time. It is likely that this scale will have more applications in the future.

This validation study suggests that GAD-2 has reasonable accuracy at cutoffs of 2 or 3 in a Chinese rural population. To be specific, GAD-2 displayed a higher sensitivity at a cutoff of 2 (91.9% *vs.* 86.5%) and a higher specificity at a cutoff of 3 (98.0% *vs.* 86.9%). Nevertheless, determining optional cutoffs should not merely consider indexes such as sensitivity and specificity, but consider the purposes of utilization. Patients with GAD would suffer further pain and a greater risk of suicide [45] if they are not identified in a timely manner. Given this consideration, clinicians might expect to choose a cutoff of 2, which appeared to have a higher sensitivity value. In contrast, a cutoff of 3 might be applied if decreasing the number of false positive patients was important, thereby avoiding unnecessary treatment risks.

Given the large population and regional differences in China, further multi-site studies are encouraged to maximize sample size and promote the generalizability of results. In addition, it is recommended that researchers report diagnostic accuracy for all cutoffs, not just best cutoffs, to ensure the integrity of data.

5. Conclusion

GAD-2 was found to be a valid screening instrument for GAD in a Chinese rural population, with cutoffs of 2 or 3 depending on the purpose of the study. Its brevity and ease of use indicate that it is a valuable tool for the timely identification of GAD.

Ethics approval

Ethics approval was obtained from the “Zhengzhou University Life Science Ethics Committee”, and written informed consent was obtained for all participants. Ethic approval code: [2015] MEC (S128).

Funding

This research was supported by the Foundation of National Key Research and Development Program of China (Grant NO: 2016YFC0900803), National Natural Science Foundation of China (Grant NO: 81573243, 81602925), Henan Natural Science Foundation of China (Grant NO: 182300410293), Science and Technology Foundation for Innovation Talent of Henan Province (Grant NO: 164100510021), Science and Technology Innovation Talents Support Plan of Henan Province Colleges and Universities (Grant NO: 14HASTIT035), High-level Personnel Special Support Project of Zhengzhou University (Grant NO: ZDGD13001). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank all of the participants, coordinators, and administrators for their support and help during the research. In addition, the authors would like to thank Dr. Tanko Abdulai for his critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genhosppsych.2019.07.008.

References

[1] Remes O, Brayne C, van der Linde R, Lafortune L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. Brain Behav 2016;6:e00497.
[2] Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004;291:2581–90.
[3] Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. The global burden of anxiety disorders in 2010. Psychol Med 2014;44:2363–74.
[4] Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013;382:1575–86.
[5] Glass RM. Diagnostic and Statistical Manual of Mental Disorders (DSM) AMA Manual of Style [Internet Oxford University Press; 2009.
[6] Somers JM, Goldner EM, Waraich P, Hsu L. Prevalence and incidence studies of anxiety disorders: a systematic review of the literature. Can J Psychiatry 2006;51:100–13.
[7] Ruscio AM, Hallion LS, Lim CCW, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. JAMA Psychiat 2017;74:465–75.
[8] Barrera TL, Norton PJ. Quality of life impairment in generalized anxiety disorder, social phobia, and panic disorder. J Anxiety Disord 2009;23:1086–90.
[9] Waghorn G, Chant D, White P, Whiteford H. Disability, employment and work performance among people with ICD-10 anxiety disorders. Aust N Z J Psychiatry 2005;39:55–66.
[10] Celano CM, Daunis DJ, Lokko HN, Campbell KA, Huffman JC. Anxiety disorders and cardiovascular disease. Curr Psychiatry Rep 2016;18:101.
[11] Osuch EA, Vingilis E, Summerhurst C, Forster CI, Ross EE, Wrath AJ. Process evaluation of an early-intervention program for mood and anxiety disorders among older adolescents and young adults. Psychiatr Serv 2015;66:1113–7.
[12] Kessler D, Lloyd K, Lewis G, Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. BMJ 1999;318:436–9.
[13] Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. Arch Gen Psychiatry 2001;58:787–94.
[14] Swinson RP. The GAD-7 scale was accurate for diagnosing generalised anxiety disorder. Evid Based Med 2006;11:184.
[15] Kroenke K, Spitzer RL, Williams JB, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146:317–25.
[16] Kujanpää T, Ylisaukko-Oja T, Jokelainen J, Hirsikangas S, Kanste O, Kyngäs H, et al. Prevalence of anxiety disorders among Finnish primary care high utilizers and validation of Finnish translation of GAD-7 and GAD-2 screening tools. Scand J Prim Health Care 2014;32:78–83.
[17] García-Campayo J, Zamorano E, Ruiz MA, Pérez-Páramo M, López-Gómez V, Rejas J. The assessment of generalized anxiety disorder: psychometric validation of the Spanish version of the self-administered GAD-2 scale in daily medical practice. Health Qual Life Outcomes 2012;10:114.
[18] Christensen H, Batterham PJ, Grant JB, Griffiths KM, Mackinnon AJ. A population study comparing screening performance of prototypes for depression and anxiety with standard scales. BMC Med Res Methodol 2011;11.
[19] Wild B, Eckl A, Herzog W, Niehoff D, Lechner S, Maatouk I, et al. Assessing generalized anxiety disorder in elderly people using the GAD-7 and GAD-2 scales: results of a validation study. Am J Geriatr Psychiatry 2014;22:1029–38.
[20] Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res 2011;188:58–64.
[21] He X, Li C, Qian J, Cui H, Wu W. Reliability and validity of a generalized anxiety disorder scale in general hospital outpatients. Shanghai Archives of Psychiatry. 22. 2010. p. 200–3.
[22] Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol 2005;58:859–62.
[23] Yu W, Singh SS, Calhoun S, Zhang H, Zhao X, Yang F. Generalized anxiety disorder in urban China: prevalence, awareness, and disease burden. J Affect Disord 2018;234:89–96.
[24] Moher D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264.
[25] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
[26] Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. J Clin Epidemiol 1991;44:381–90.
[27] YOUDEN WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.
[28] Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–90.
[29] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
[30] Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559–73.
[31] Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002;31:88–95.
[32] van Heyningen T, Honikman S, Tomlinson M, Field S, Myer L. Comparison of mental health screening tools for detecting antenatal depression and anxiety disorders in South African women. PLoS One 2018;13:e0193697.
[33] Cano-Vindel A, Muñoz-Navarro R, Medrano LA, Ruiz-Rodríguez P, González-Blanch C, Gómez-Castillo MD, et al. A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional disorders in primary care. J Affect Disord 2018;234:247–55.
[34] Micoulaud-Franchi JA, Bartolomei F, McGonigal A. Ultra-short screening instruments for major depressive episode and generalized anxiety disorder in epilepsy: the NDDIE-2 and the GAD-SI. J Affect Disord 2017;210:237–40.
[35] Seo JG, Park SP. Validation of the Generalized Anxiety Disorder-7 (GAD-7) and GAD-2 in patients with migraine. J Headache Pain 2015;16.
[36] Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, et al. Brief case finding tools for anxiety disorders: validation of GAD-7 and GAD-2 in addictions treatment. Drug Alcohol Depend 2012;125:37–42.
[37] Qian J, Bian C, Cui H, Wu W, Li C. Comparative study of reliability and validity between several screening scales for anxiety. J Intern Med Concepts Pract 2011;6:176–9.
[38] Wang L, Lu K, Wang C, Sheng L, Hu D, Ding R. Reliability and validity of GAD-2 and GAD-7 for anxiety screening in cardiovascular disease clinic. Sichuan Ment Health 2014;27:198–201.
[39] Hundleby JD, Nunnally J. Psychometric theory. Am Educ Res J 1968;5:431.
[40] Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–93.
[41] Zis P, Gatzonis S. Estimating the diagnostic value of the Neurological Disorders Depression Inventory for Epilepsy in different languages. Epilepsia 2014;55:941.
[42] Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24–31.
[43] Silva MT, Caicedo Roa M, Martins SS, da Silva ATC, Galvao TF. Generalized anxiety disorder and associated factors in adults in the Amazon, Brazil: a population-based study. J Affect Disord 2018;236:180–6.
[44] Hughes AJ, Dunn KM, Chaffee T, Bhattarai JJ, Beier M. Diagnostic and clinical utility of the GAD-2 for screening anxiety symptoms in individuals with multiple sclerosis. Arch Phys Med Rehabil 2018;99:2045–9.
[45] Gilmour H. Threshold and subthreshold Generalized Anxiety Disorder (GAD) and suicide ideation. Health Rep 2016;27:13–21.