RESEARCH

Poor performance of PHQ-9 and GAD-7 in screening clinical depression and anxiety among a large sample of Chinese children and adolescents

Sifan Wang[1,2†], Jin Lu[3,4,5†], Guiqing Zheng[1], Yandie He[1], Shuqing Liu[1], Yi Xiang[1], Xinyi Liu[1], Xiang Wang[1] and Yuanyuan Xiao[1*]

Abstract

Background The Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder scale (GAD-7) are widely used symptom-based instruments for screening depression and anxiety. However, their validation in Chinese children and adolescents remains insufficient. We aim to investigate the performance and optimal cut-offs of PHQ-9 and GAD-7 in Chinese children and adolescents in screening clinical depression and anxiety, and to discuss the influencing factors of the cut-offs.

Methods The study subjects were chosen from 3 sites of the Mental Health Survey for Children and Adolescents in Yunnan (MHSCAY), a total of 2,237 participants who had been screened positive by self-administered questionnaire and further diagnosed by using the gold standard were included in the final analysis. The Receiver Operating Characteristic (ROC) curves were used to determine the discriminative ability of the two instruments, measured by using the area under curve (AUC). The optimal cut-offs of the two instruments were determined by the maximum Youden's index. A series of stratified analyses were performed to discover the best cut-offs for children and adolescents of different characteristics. Logistic regression models were adopted to evaluate the influence of self-harm (SH) on identified optimal cut-offs.

Results We found the performance was generally poor for both the PHQ-9 and GAD-7 in screening clinical depression and anxiety in our study sample, with the AUCs ranged only between 0.622–0.712. When using the two instruments for diagnosis purpose, 11 was the optimal cut-off for both clinical depression and anxiety (AUC for PHQ-9: 0.664, AUC for GAD-7: 0.669). For study subjects of different age, gender, race, and left-behind status, discordant cut-offs were identified. SH also showed conspicuous influence on the optimal cut-offs of PHQ-9 and GAD-7, and the combination of SH information can increase screening accuracy of PHQ-9 in some subgroups.

Conclusions Both the PHQ-9 and GAD-7 showed poor performance in screening clinical depression and anxiety in our study sample. This crucial finding suggests that, despite the wide use of the two scales, they might be

---

†Sifan Wang and Jin Lu contributed equally as joint first authors.

*Correspondence:
Yuanyuan Xiao
33225647@qq.com
Full list of author information is available at the end of the article

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 2 of 11

fundamentally inadequate for depression and anxiety screening in Chinese children and adolescents. Other screening tools of higher accuracy should be developed and used in this age group.

Keywords Children, Adolescents, PHQ-9, GAD-7, Optimal cut-off, Depression, Anxiety

Introduction

Depression and anxiety are the most common mental diseases and the major causes of disease burden in young people [1]. A previously published meta-analysis revealed that the global prevalence rates of depressive and anxiety disorders were 2.6% and 6.5% in children and adolescents [2]. In Chinese children and adolescents, the estimated prevalence rates were 3.0% and 4.7% [3]. Early-onset depressive and anxiety disorders are related to various negative outcomes, such as low academic performance, poor interpersonal relationships, drug dependence, and other mental illnesses [4, 5]. More importantly, their detrimental influence could extend into adulthood. For instance, prospective studies found that depression and anxiety disorders in childhood and adolescence were associated with more than five-fold risk of adult suicide attempt [6, 7].

Screening children and adolescents of depression or anxiety is critical for timely diagnosis and treatment. Because of conciseness and ideal performance, the Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder scale (GAD-7) are the most widely used screening tools for depression and anxiety. Both instruments measure two-week symptoms, and firstly developed for the adult populations, with corroborated validity and reliability [8, 9]. Although the use of PHQ-9 and GAD-7 in children and adolescents has also been validated subsequently [10, 11], some newly emerged evidence suggested less optimistic discriminative accuracy in screening clinical depression and anxiety in this age group [12, 13]. Moreover, the performance of the two instruments in Chinese youth populations has not been adequately discussed.

Besides, controversies still exist in cut-offs for the use of the two scales in children and adolescents. For instance, a score of 10 has been normally adopted for screening depression in adults by using PHQ-9 [14], although a previous study supported the use of this same cut-off in Chinese children and adolescents [15], another research indicated that the cut-off of 15 reached the best joint accuracy [16]. For GAD-7, the optimal cut-off for youths is also inconclusive: a cut-off of 9 had been validated in outpatients [17], another study found 7 could be preferable [18]. These discordant findings introduce great confusion when applying these two screening instruments in Chinese youngsters. Another important thing to be noticed is that the distribution of depressive and anxiety symptoms varies in some characteristics, such as age, gender, race, self-harm (SH) behaviors [19, 20]. Therefore, it is highly possible that specific cut-offs should be applied for these two instruments for different subgroups of children and adolescents. However, this assumption has not been effectively corroborated.

In this study, by using a large sample of Chinese children and adolescents from an ongoing population-based survey project, we aim to evaluate the performance of PHQ-9 and GAD-7 in screening clinical depression and anxiety, investigate the optimal cut-offs of the two instruments, and discuss the influence of important characteristics on the optimal cut-offs. The following major hypotheses will be tested: (1) Whether the discriminative accuracy of PHQ-9 and GAD-7 is satisfactory; (2) What are the optimal cut-offs for PHQ-9 and GAD-7 in screening clinical depression and anxiety among children and adolescents; (3) Whether important characteristics impose significant influence on the optimal cut-offs of the two instruments.

Methods

Participants

The data used for analysis in this study were obtained from the Mental Health Survey for Children and Adolescents in Yunnan (MHSCAY). We used the combined database from 3 study sites (Zhenxiong, Yuxi, and Kaiyuan), which finished survey between May and October 2021. In Zhenxiong and Yuxi, a two-stage probability proportionate to sample size (PPS) sampling method was used to choose study subjects: in the first stage, a certain number of primary, secondary, and senior high schools were randomly selected from all schools within the study site; in the second stage, 3–4 classes were randomly selected from each chosen school based on pre-calculated required sample size, and all students in the chosen classes who satisfied the inclusion criteria were eligible participants. In Kaiyuan, we conducted a census for all left-behind children (LBC), defined as children under 18 years old, with one parent or both parents migrated to other places for more than 6 months in the past year. Other detailed inclusion and exclusion criteria can be referred to in our previous publication [21]. The study protocol has been reviewed and approved by the Ethics Committee of Kunming Medical University (No. KMMU2020MEC047). Prior to the survey, we obtained

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 3 of 11

written informed consents from the participants' parents or legal guardians.

Procedure

The survey at all study sites was carried out in two phases. In phase one, all participants were required to finish a self-administered questionnaire. In phase two, positive subjects screened out by using relevant symptom-based instruments which incorporated into the questionnaire were interviewed by pediatric psychiatrists deployed at site using the Kiddie-SADS-Present and Lifetime Version (K-SADS-PL) [22], a semi-structured diagnostic interview used to assess current and past psychotic episodes in children and adolescents following the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition) (DSM-5).

Positive subjects to be further diagnosed in phase two were: 1) PHQ-9 score ≥ 5 or/and; 2) GAD-7 score ≥ 5 or/and; 3) Repeated or severe self-harm (SH) behaviors occurred or/and; 4) Severe suicidal ideation, or ever made suicide plan, or ever tried suicide or/and; 5) Positive in cross-symptom measurement of Kiddie-SADS-Present and Lifetime Version (K-SADS-PL) Level 1. To protect the privacy of the interviewees, interviews were conducted face-to-face in a separate room. For subjects who were identified high suicidal risk or serious mental symptoms, timely psychological counseling will be provided at site, and the guardians will be formally notified.

Measurements

Depressive symptoms

In the self-administered questionnaire, depressive symptoms were measured by using the PHQ-9. It has nine items that correspond to the nine diagnostic criteria for depression (interest in doing things, mood fluctuations, sleep quality, vitality, appetite, self-evaluation, concentration on things, speed of movement, thoughts of suicide) [23]. Each item is scored from 0 to 3 points, indicating "not at all" (0 point), "a few days" (1 point), "more than half of the days" (2 points), and "almost every day" (3 points). The total score of PHQ-9 ranges from 0 to 27 points, with a higher score indicates severer depressive symptoms.

Anxiety symptoms

Two-week anxiety symptoms were measured by using the GAD-7, a seven-item scale which evaluates nervousness, anxiety, uncontrollable worry, excessive worry, inability to relax, inability to sit still, irritability and foreboding [24]. Each item has four choices in frequency of anxiety symptoms, from "not at all" (0 point) to "almost every day" (3 points), with a combined score of 0–21. A higher combined score reflects more serious anxiety symptoms.

SH behaviors

SH behaviors were measured using the Modified Adolescent Self-Harm Scale (MASHS) developed exclusively for Chinese adolescents [25]. Likert-style responses were used to assess lifetime frequency (never, once, two to four times, five times, more than five times) and severity (non-observable, slight, moderate, severe, critical) of the 18 most common SH behaviors. Repeated SH was defined as committed SH "two to four times" and above, severe SH was defined as the severity of SH ever reached "moderate" and above.

Statistical analysis

Data were analyzed by using the R software (Version 4.2.2, The R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics were used to present general characteristics of the study subjects. Sensitivity (Se), specificity (Sp), and the Receiver Operating Characteristic (ROC) curves were analyzed by using the "pROC" package, discriminative ability of the instrument was measured by the area under curve (AUC). The maximum Youden's index (sensitivity + specificity − 1) was used to determine the optimal cut-offs for GAD-7 and PHQ-9, which can help balance the cut-offs between sensitivity and specificity, leading to optimal model performance. The value of Youden's index ranges from −1 to 1, with a larger Youden's index indicating better authenticity of the screening tool [26]. A series of stratified analyses were performed to determine the best cut-offs for GAD-7 and PHQ-9 considering different characteristics of children and adolescents, like sex, age, race, left-behind status, specific diagnosis of depression or anxiety, and SH. Moreover, the influence of SH on the diagnostic accuracy of depression or anxiety based on the identified optimal cut-offs was evaluated by comparing the pseudo R^2 statistic between univariate and multivariate logistic regression models. The significance level for all analyses was set as $p < 0.05$ (two-tailed analysis).

Results

General characteristics

A total of 15,368 study subjects completed phase one self-administered questionnaire survey, 2,245 were screened positive and entered phase two diagnostic interview, 8 subjects were excluded because of incomplete information, all the rest 2,237 participants who had been diagnosed by the gold standard (K-SADS-PL) were included in the final analysis. Table 1 showed the general characteristics of analyzed participants. The mean of age was 13.73 years (standard deviation, S.D.: 1.99); most of the subjects were teenagers (74.43%) aged 13–17; girls took a larger proportion of 63.75%; LBC and non-left-behind

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 4 of 11

Table 1 General characteristics of study subjects (*N* = 2237)

Characteristics	*N* (%)	Mean (SD)
Sex		
Boys	811 (36.35)	
Girls	1426 (63.75)	
Age		13.73 (1.922)
10–12 years	572 (25.57)	
13–17 years	1665 (74.43)	
Race		
Han majority	1870 (83.59)	
Minorities	367 (16.41)	
Left-behind children		
Yes	1105 (49.40)	
No	1132 (50.60)	
Self-harm behaviors		
Yes	1820 (81.36)	
Repetitive self-harm	1470 (65.71)	
Moderate and above self-harm	913 (40.81)	
No	417 (18.64)	
Depression disorders	574 (25.66)	
Major depression disorder	364 (16.27)	
Other depression disorders	210 (9.39)	
Anxiety disorders	180 (8.05)	
Generalized anxiety disorder	101 (4.51)	
Other anxiety disorders	79 (3.53)	

children (NLBC) were similar in proportion (49.40% versus 50.60%); more than 80% of the participants reported SH behaviors, and most of them committed repetitive SH (80.74%); 574 (25.66%) and 180 (8.05%) subjects were diagnosed with depressive disorder and anxiety disorder.

Performance and optimal cut-offs of PHQ-9 and GAD-7

We found that the overall discriminative ability for both PHQ-9 and GDA-7 was poor for screening clinical depression and anxiety in our study sample: AUC for PHQ-9 was 0.664 (Se = 0.551; Sp = 0.695), AUC for GAD-7 was 0.669 (Se = 0.683; Sp = 0.600). For children and adolescents of different sex, race, and left-behind status, the AUCs for PHQ-9 and GAD-7 were also largely below acceptable level, only ranged between 0.622–0.712. For the analytical sample in general, the optimal cut-off for both PHQ-9 and GAD-7 was 11 (Fig. 1).

For children and adolescents of different characteristics, as shown in panel C-H in Fig. 1, Table 2, and Fig. 2, we found that gender, left-behind status, race, and age showed conspicuous influence on the optimal cut-offs of PHQ-9 and GAD-7. Sex showed particularly strong influence: for PHQ-9, the optimal cut-offs in screening clinical depression for boys and girls were 11 (AUC: 0.675) and 15 (AUC: 0.648); for GAD-7, the optimal cut-offs

in screening clinical anxiety for boys and girls were 13 (AUC: 0.712) and 11 (AUC: 0.639). Race also exerted strong impact on PHQ-9, for Han majority, the optimal cut-off was 11 (AUC: 0.672), whereas for minorities, it was 16 (AUC: 0.627).

The influence of SH on optimal cut-offs of PHQ-9 and GAD-7

We divided study subjects into two subgroups by the presence of SH behaviors: for those who reported SH behaviors, the optimal cut-offs for PHQ-9 and GAD-7 were 14 (AUC: 0.654) and 11 (AUC: 0.662), higher than 11 (AUC: 0.653) and 9 (AUC: 0.678) for those without SH behaviors. We also estimated the influence of SH repetition and severity on the optimal cut-offs for PHQ-9 and GAD-7 in self-harmed study subjects exclusively, analytical results revealed that compared with study subjects with one-off SH behaviors, those who reported repeated SH were observed lower optimal cut-offs in both instruments. For severity of SH, study subjects who reported severer SH were found higher optimal cut-off of PHQ-9, and lower cut-off of GAD-7 (Fig. 3, Table 3).

Combination of SH on screening accuracy of PHQ-9 and GAD-7

We further explored the combination of SH on screening accuracy of PHQ-9 and GAD-7 based on pre-determined optimal cut-offs, and the results were summarized in Table 4. The results showed that the combination of SH could significantly improve screening accuracy for PHQ-9, but not for GAD-7. In general, the inclusion of SH could increase the screening accuracy of PHQ-9 by 23.5% in the overall study sample. For children and adolescents of different characteristics, the combination of SH was related to the most prominent increase in screening accuracy of PHQ-9 for minorities (79.3%), adolescents (31.5%), and girls (28.8%).

Discussion

In the current study, we discussed the performance and optimal cut-offs of PHQ-9 and GAD-7 for screening clinical depression and anxiety in a large sample of Chinese children and adolescents. First and foremost, we found the overall performance of the two scales was poor, as the diagnostic accuracy was generally under 70%. Second, as expected, the optimal cut-offs of the two scales were different for Chinese children and adolescents, compared with teenagers of other origins. Moreover, for children and adolescents of different characteristics, the optimal cut-offs varied considerably. Besides, for PHQ-9, the combination of SH information could increase its screening accuracy for clinical depression among youngsters. These important findings could provide guidance in
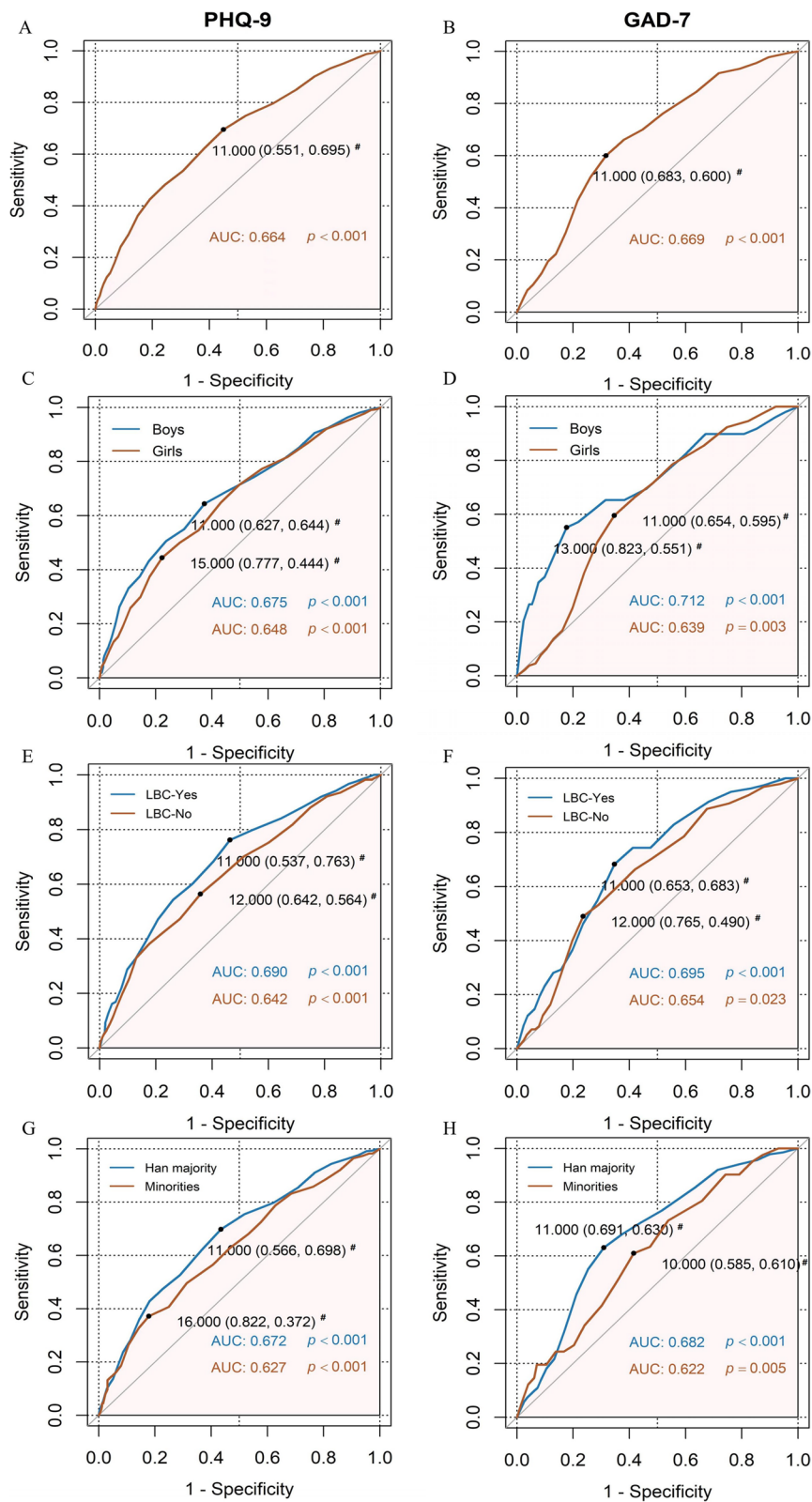
Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 5 of 11

Fig. 1 ROC curves with AUC values for PHQ-9 and GAD-7 in Chinese children and adolescents. **A** and **B**: all children and adolescents; **C** and **D**: by sex; **E** and **F**: by left-behind status; **G** and **H**: by race. [#] Optimal cut-off (sensitivity, specificity)

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 6 of 11

Table 2 The optimal cut-offs of PHQ-9 and GAD-7 for children and adolescents by age

Age (years)	PHQ-9				GAD-7			
	Cut-off	AUC (95%CI)	Se, Sp	p	Cut-off	AUC (95%CI)	Se, Sp	p
10	14	0.545 (0.422–0.668)	0.296, 0.817	0.144	NA	NA	NA	NA
11	12	0.677 (0.581–0.773)	0.615, 0.655	< 0.001	NA	NA	NA	NA
12	13	0.664 (0.581–0.746)	0.566, 0.698	< 0.001	12	0.680 (0.549–0.810)	0.631, 0.690	0.004
13	11	0.693 (0.636–0.751)	0.755, 0.567	< 0.001	10	0.654 (0.568–0.740)	0.697, 0.622	0.002
14	12	0.671 (0.620–0.723)	0.664, 0.587	< 0.001	11	0.617 (0.540–0.694)	0.520, 0.676	0.003
15	16	0.691 (0.628–0.756)	0.436, 0.872	< 0.001	12	0.666 (0.564–0.769)	0.548, 0.764	0.001
16	15	0.609 (0.515–0.701)	0.454, 0.813	0.007	8	0.695 (0.592–0.798)	0.818, 0.486	0.001
17	11	0.619 (0.527–0.710)	0.640, 0.570	0.006	10	0.842 (0.774–0.910)	0.933, 0.695	< 0.001
Children (10–12)	13	0.642 (0.585–0.698)	0.479, 0.732	< 0.001	12	0.670 (0.566–0.774)	0.586, 0.716	0.371
Adolescents (13–17)	11	0.669 (0.639–0.698)	0.714, 0.536	< 0.001	11	0.671 (0.629–0.712)	0.596, 0.695	< 0.001

better use of symptom-based screening tools for detecting clinical depression and anxiety among Chinese children and adolescents.

Because of the brief nature and easy scoring method, PHQ-9 and GAD-7 have become two of the most popular symptom-based screening tools for depression and anxiety. However, our findings raised serious concerns for their valid use in Chinese children and adolescents, as the AUCs were generally under 0.7 for the combined sample and across all demographic subgroups, suggesting high chance of both misdiagnosis and missed diagnosis. The poor performance of the two questionnaires probably can be partly ascribed to the choosing method of study subjects, as we only included children and adolescents who were screened out in phase one self-administered survey, for only this part of study subjects were all diagnosed by the gold standard. Based on the screening criteria that we described above, it is possible that the false positive rate in our study subjects was higher than in the general population, which resulted in a lower specificity. Nevertheless, the sensitivity of the two scales was also below the acceptable level, only 0.551 for the PHQ-9, and 0.683 for the GAD-7. All the above results suggest that symptom-based questionnaires of better screening performance should better be used in Chinese youths. A newly published study revealed ideal performance of the Patient Health Questionnaire for Adolescents (PHQ-A), a modified version of PHQ-9, in 248 Chinese adolescents with clinical depression [27]. Another study consisted of 699 English children and adolescents disclosed a high accuracy for the Spence Children's Anxiety Scale (Parent version) (SCAS-P) [28] in detecting anxiety disorders. Owing to small and unrepresentative samples, the performance of these promising tools in screening depression and anxiety in Chinese children and adolescents should be further validated.

Although most published studies favored the uniform cut-off of 10 for screening clinical depression and anxiety in children and adolescents when using PHQ-9 and GAD-7 [29–31], our study found that 11 was the optimal cut-off, and this same optimal cut-off was also identified for youths in the US and Chile [11, 32, 33]. However, heterogeneity exists among published studies of Chinese children and adolescents. Take PHQ-9 for example, a small sample size study found the cut-off of 10 was the optimal [15], whereas another study originated from Taiwan province favored the cut-off of 15 [14]. Our findings support the use of population-specific optimal cut-offs for PHQ-9 and GAD-7 in children and adolescents. Besides, considering the scarcity of currently available high-quality evidence, together with discordant results, the optimal cut-offs for the two instruments in Chinese children adolescents should be further investigated.

Another important finding of our study would be that some characteristics of study subjects posed significant influence on the optimal cut-offs of the two scales. Some scholars emphasized the influence of population characteristics on cut-off selection for psychometric tools [34, 35]. In this study, we found the optimal cut-off of PHQ-9 was significantly higher in girls (15) than in boys (11), probably because girls present greater tendency in reporting affective symptoms (e.g., sadness, hopelessness) at similar depression levels than boys [36, 37]. However, for the GAD-7, the optimal cut-off was higher in boys (13) than in girls (11), which could be supported by the evidence that a greater proportion of females are characterized by elevated levels of constructs of neuroticism and negative affectivity [38], therefore anxious females were more likely to be diagnosed with an anxiety disorder [39, 40]. Stark interracial difference existed for the optimal cut-off of PHQ-9, compared with Han majority, a much higher cut-off has been revealed for minorities.
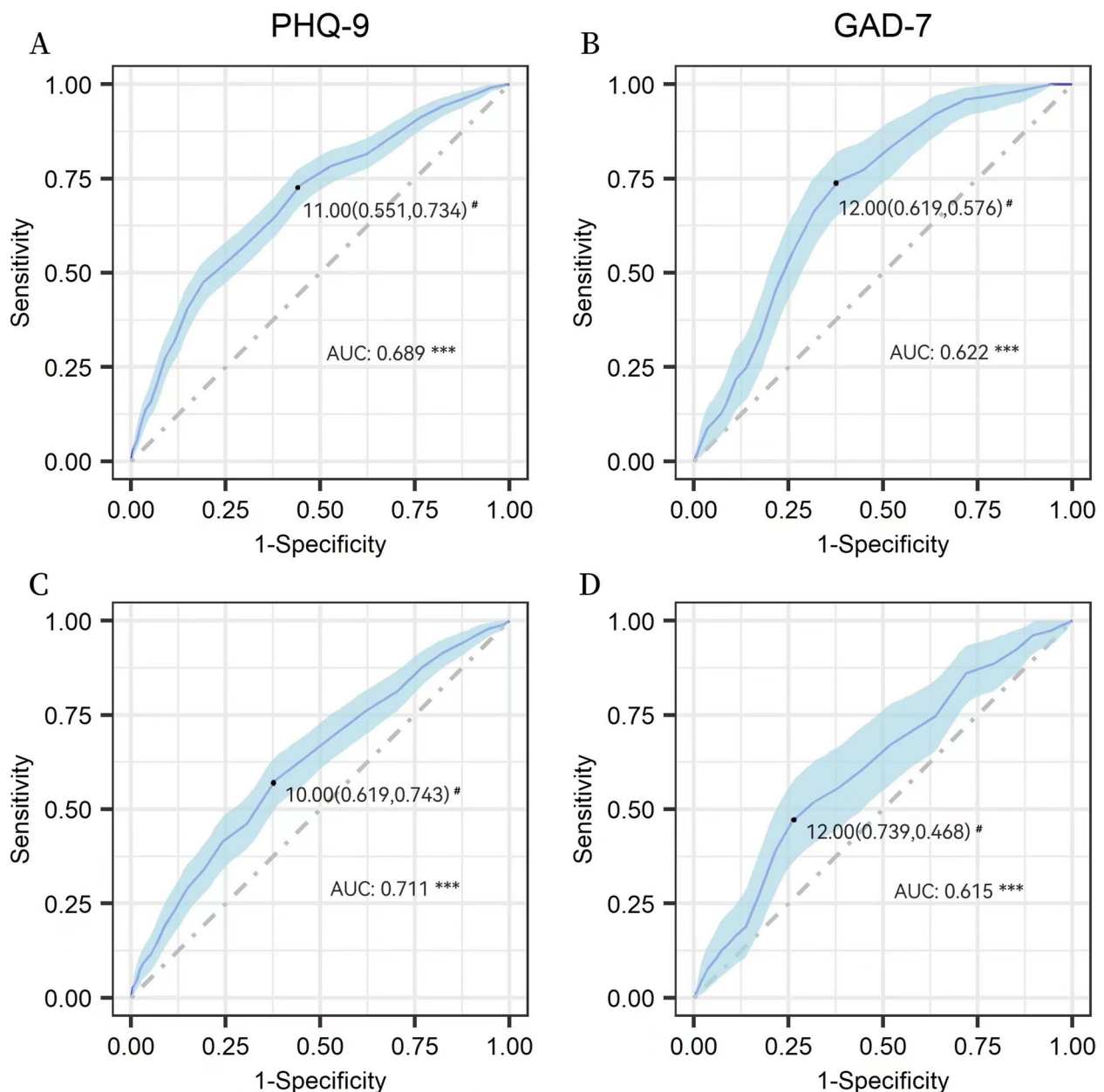
Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 7 of 11

Fig. 2 ROC curves with AUC values of PHQ-9 and GAD-7 by different anxiety and depression diagnoses. **A**: major depression disorder, MDD; **B**: generalized anxiety disorder, GAD; **C**: other types of depression; **D**: other types of anxiety. $^{***}p < 0.001$, $^{\#}$ Optimal cut-off (sensitivity, specificity)

Cultural biases have been disclosed for GAD-7 among different ethnic groups [41], it seems that the same situation may also apply for PHQ-9. All these discrepancies highlight the necessity of applying characteristic-specific cut-offs for PHQ-9 and GAD-7 in Chinese children and adolescents.

The analytical results suggest that SH behaviors significantly influenced the optimal cut-offs of PHQ-9 and GAD-7, particularly for PHQ-9. Children and adolescents who reported any SH behaviors were observed a

higher PHQ-9 cut-off, indicating SH may be related to lower risk of clinical depression for youths who present depressive symptoms. Perhaps because SH is an effective way for adolescents to relieve pressure when facing external stimuli that may bring psychological distress [42]. Among self-harmed subjects, a lower cut-off of PHQ-9 has been revealed for those who committed repetitive SH, suggesting increased clinical depression risk in this group. This finding could be well supported by longitudinal evidence that for adolescents, repeated SH was
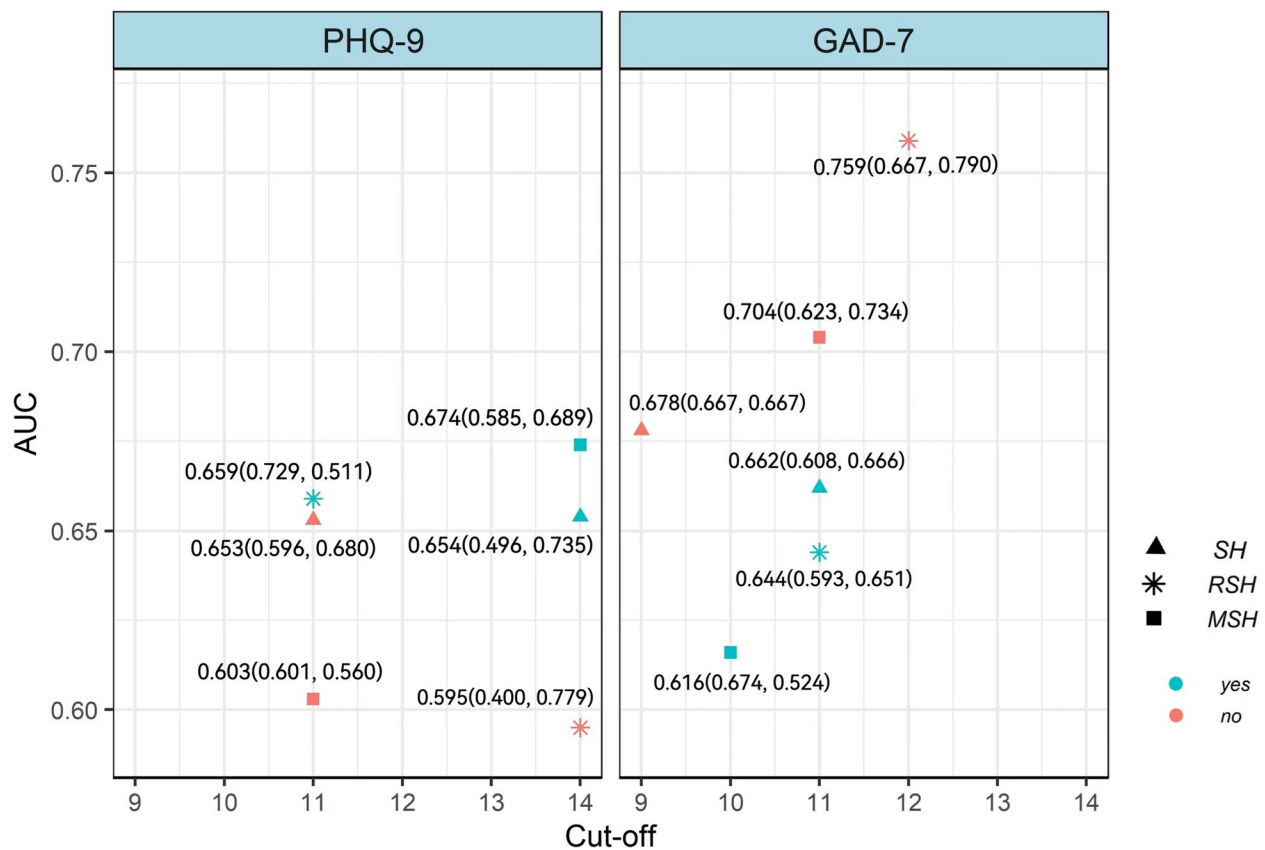
Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 8 of 11

Fig. 3 The optimal cut-offs and AUC values of PHQ-9 and GAD-7 in children and adolescents by SH. SH: self-harm behaviors; RSH: repetitive self-harm behaviors; MSH: moderate and above self-harm behaviors

Table 3 Optimal cut-offs for PHQ-9 and GAD-7 in children and adolescents with different levels if SH

	General subjects	Without SH	SH	RSH	MSH
PHQ-9	11	11	14	11	14
GAD-7	11	9	11	11	10

SH self-harm behaviors, *RSH* repetitive self-harm behaviors, *MSH* moderate and above self-harm behaviors

related to 4.65 folds of depressive disorders risk [43]. For severity of SH, a higher cut-off of PHQ-9 has been identified for those with severer SH, which seems to be contradictory to published evidence that more serious SH could be associated with higher risk of mental disorders [44]. Nevertheless, insignificant association between SH severity and psychological consequences has also been reported [45].

Except for its noticeable influence on the optimal cut-offs of the two scales, we also found that the combination use of SH information can significantly improve screening accuracy of PHQ-9: for all children and adolescents,

the accuracy improved by nearly one-fourth (23.5%). However, for GAD-7, the combination of SH information only slightly improved screening accuracy by 1.75%. This stark difference could probably be attributed to much more frequent SH among depressed individuals, as a previous study found that compared with study subjects free of psychiatric diagnoses, the odds ratio of SH was 4.6 times for those who were diagnosed with depressive disorders, whereas only 1.5 times for those with anxiety disorders [46]. SH can help distinguish individuals at a higher risk of clinical depression from children and adolescents with depressive symptoms, possibly because individuals diagnosed with depressive disorders had a significantly higher risk of SH than those with subclinical depression (76% vs. 44%) [47, 48]. The mechanism behind this may be that SH is an effective way for depressed individuals to relieve their misery and negative feelings [49], and that manifestations of depression, such as sadness, feelings of worthlessness, guilt, and other negative emotions, were all prominent triggers of SH [37].

We also found that the elevation of accuracy of PHQ-9 combining with SH can be even higher for minorities (79.3%), adolescents (31.5%), and girls (28.8%). It may be

Wang *et al. BMC Psychiatry*　　(2025) 25:301

Page 9 of 11

Table 4 Univariate and multivariate logistic regression models fitting results for the influence of SH on screening accuracy of PHQ-9 and GAD-7

	Univariate		Multivariate		
	b (*p*)	Pseudo R^2	*b* (*p*)	*b*-SH (*p*)	Pseudo R^2
All subjects					
PHQ-9	1.027 (0.001)	0.068	0.962 (0.001)	0.738 (0.001)	0.084
GAD-7	1.173 (0.001)	0.057	1.163 (0.001)	0.170 (0.001)	0.058
By sex					
Boys					
PHQ-9	1.110 (0.001)	0.073	1.049 (0.001)	0.493 (0.041)	0.081
GAD-7	1.741 (0.001)	0.105	1.780 (0.001)	0.338 (0.331)	0.108
Girls					
PHQ-9	1.028 (0.001)	0.066	0.984 (0.001)	0.843 (0.001)	0.085
GAD-7	1.022 (0.001)	0.046	1.014 (0.001)	0.395 (0.181)	0.048
By age					
Children					
PHQ-9	0.922 (0.001)	0.050	0.894 (0.001)	0.337 (0.248)	0.054
GAD-7	1.272 (0.001)	0.057	1.290 (0.001)	0.227 (0.635)	0.058
Adolescents					
PHQ-9	1.062 (0.001)	0.073	0.992 (0.001)	0.892 (0.001)	0.096
GAD-7	1.212 (0.001)	0.064	1.200 (0.001)	0.271 (0.275)	0.065
By LBC					
LBC-yes					
PHQ-9	1.314 (0.001)	0.102	1.244 (0.001)	0.809 (0.001)	0.118
GAD-7	1.399 (0.001)	0.076	1.390 (0.001)	0.135 (0.690)	0.077
LBC-no					
PHQ-9	0.841 (0.001)	0.048	0.790 (0.001)	0.704 (0.001)	0.065
GAD-7	1.138 (0.001)	0.053	0.191 (0.001)	0.213 (0.508)	0.054
By ethnicity					
Han majority					
PHQ-9	1.101 (0.001)	0.076	1.044 (0.001)	0.641 (0.001)	0.088
GAD-7	1.339 (0.001)	0.071	1.330 (0.001)	0.172 (0.489)	0.066
Minorities					
PHQ-9	1.001 (0.001)	0.058	0.916 (0.001)	1.146 (0.004)	0.094
GAD-7	0.788 (0.020)	0.030	0.782 (0.022)	0.068 (0.886)	0.030

related to differences in the strength of the depression-SH association in different subgroups. Girls with depression are more likely to report SH than boys, which may be related to the earlier onset of adolescence in females. Hormonal changes and shifting social roles may put girls at higher risk for SH [50]. Additionally, prevalence of SH peaks in adolescence [51], with more than half of depressed adolescents experiencing SH [52], which explains the higher elevation of SH for clinical depression screening accuracy in adolescents than in children. However, whether SH behaviors are more likely to occur in ethnic minorities with depression is still unclear. Under any situation, the above findings suggest, when using PHQ-9 to screen for clinical depression in children and adolescents, an extra simple question for previous SH behaviors should be considered, especially for key subpopulations, to gain a higher level of screening accuracy. Although incorporating SH can improve PHQ-9 screening accuracy, self-reported SH may be underestimated due to shame and stigma [53], making detection challenging. Therefore, how to accurately and safely collect SH information should be further investigated.

Our study tested the performance and provided the optimal cut-offs of PHQ-9 and GAD-7 in screening clinical depression and anxiety among Chinese children and adolescents. Particularly, we discussed the influence of some important characteristics on the optimal cut-offs and screening accuracy of the two instruments. The major results of our study can provide important guidance for better use of symptom-based scales in screening clinical depression and anxiety in Chinese youngsters. Even though, some limitations should be noticed. At first, all study subjects were chosen from one province in China, so the extrapolation should be cautiously made. Secondly, due to the purpose of our original study, we only included children and adolescents aged 10–17, for children less than 10 years old, the discriminative accuracy and optimal cut-offs of PHQ-9 and GAD-7 should be further discussed. Thirdly, although the overall sample size in the current study was large, some subgroups, such as ethnic minorities, had relatively small sample size, which definitely affected the statistical power. Furthermore, considering the cross-sectional nature of our study, future longitudinal design is warranted to further validate our major findings.

Conclusion

By using population-based representative large sample, we effectively discussed the performance and optimal cut-offs of PHQ-9 and GAD-7 for screening clinical depression and anxiety among Chinese children and adolescents. Based on analytical results, we found the overall discriminative accuracy of the two scales is poor, which makes them unreliable

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 10 of 11

screening tools for clinical depression and anxiety in this population. If still choose to use the two scales in this age group, then characteristic-specific optimal cut-offs should be applied whenever it is possible. Besides, for PHQ-9, SH information should be simultaneously incorporated to improve its performance. Other symptom-based screening tools of better performance should be devised and validated in Chinese youths.

Declarations

Ethics approval and consent to participate
The study protocol has been reviewed and approved by the Ethics Committee of Kunming Medical University. All study procedures were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]NHC Key Laboratory of Drug Addiction Medicine, Division of Epidemiology and Health Statistics, School of Public Health, Kunming Medical University, 1168 West Chunrong Road, Kunming, Yunnan 650500, China. [2]Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China. [3]Psychiatry Department, The First Affiliated Hospital, Kunming Medical University, Kunming, Yunnan, China. [4]Mental Health Institute of Yunnan, the First Affiliated Hospital, Kunming Medical University, Kunming, Yunnan, China. [5]Yunnan Clinical Research Center for Mental Health, Kunming, Yunnan, China.

References

1. Patel V. Why adolescent depression is a global health priority and what we should do about it. J Adolesc Health. 2013;52(5):511–2. https://doi.org/10.1016/j.jadohealth.2013.03.003.
2. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015;56(3):345–65. https://doi.org/10.1111/jcpp.12381.
3. Li F, Cui Y, Li Y, Guo L, Ke X, Liu J, et al. Prevalence of mental disorders in school children and adolescents in China: diagnostic data from detailed clinical assessments of 17,524 individuals. J Child Psychol Psychiatry. 2022;63(1):34–46. https://doi.org/10.1111/jcpp.13445.
4. Clayborne ZM, Varin M, Colman I. Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. J Am Acad Child Adolesc Psychiatry. 2019;58(1):72–9. https://doi.org/10.1016/j.jaac.2018.07.896.
5. Woodward LJ, Fergusson DM. Life course outcomes of young people with anxiety disorders in adolescence. J Am Acad Child Adolesc Psychiatry. 2001;40(9):1086–93. https://doi.org/10.1097/00004583-200109000-00018.
6. Stapinski LA, Bowes L, Wolke D, Pearson RM, Mahedy L, Button KS, et al. Peer victimization during adolescence and risk for anxiety disorders in adulthood: a prospective cohort study. Depress Anxiety. 2014;31(7):574–82. https://doi.org/10.1002/da.22270.
7. Weissman MM, Wolk S, Goldstein RB, Moreau D, Adams P, Greenwald S, et al. Depressed adolescents grown up. JAMA. 1999;281(18):1707–13. https://doi.org/10.1001/jama.281.18.1707.
8. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
9. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;22;166(10):1092–7. https://doi.org/10.1001/archinte.166.10.1092.
10. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, Herzberg PY. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care. 2008 Mar;46(3):266–74. https://doi.org/10.1097/MLR.0b013e318160d093.
11. Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, et al. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. Pediatrics. 2010;126(6):1117–23. https://doi.org/10.1542/peds.2010-0852.
12. Pilunthanakul T, Goh TJ, Fung DSS, Sultana R, Allen JC, Sung M. Validity of the patient health questionnaire 9-item in autistic youths: a pilot study. BMC Psychiatry. 2021;21(1):564. https://doi.org/10.1186/s12888-021-03556-w.
13. Marlow M, Skeen S, Grieve CM, Carvajal-Velez LC, Åhs JW, Kohrt BA, et al. Detecting depression and anxiety among adolescents in South Africa: validity of the isiXhosa Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7. J Adolesc Health. 2023;72(1S):S52–60. https://doi.org/10.1016/j.jadohealth.2022.09.013.
14. Tsai FJ, Huang YH, Liu HC, Huang KY, Huang YH, Liu SI. Patient health questionnaire for school-based depression screening among Chinese adolescents. Pediatrics. 2014;133(2):e402–9. https://doi.org/10.1542/peds.2013-0204.
15. Hu X, Zhang Y, Liang W, Zhang H, Yang S. Reliability and validity of the patient health questionnaire-9 in Chinese adolescents. Mental Health in Sichuan. 2014;27:357–60. https://doi.org/10.3969/j.issn.1007-3256.2014.04.021.(InChinese).
16. He XY, Li CB, Qian J, Cui HS, Wu WY. Reliability and validity of a generalized anxiety disorder scale in general hospital outpatient. Shanghai Arch. Psychiatry 2010;22:200–203. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=JSYI201004003&DbName=CJFQ2010. (In Chinese).
17. Ip H, Suen YN, Hui CLM, Wong SMY, Chan SKW, Lee EHM, et al. Assessing anxiety among adolescents in Hong Kong: psychometric properties and validity of the Generalised Anxiety Disorder-7 (GAD-7) in an epidemiological community sample. BMC Psychiatry. 2022;22(1):703. https://doi.org/10.1186/s12888-022-04329-9.
18. Cartwright-Hatton S, McNicol K, Doubleday E. Anxiety in a neglected population: prevalence of anxiety disorders in pre-adolescent children. Clin Psychol Rev. 2006;26(7):817–33. https://doi.org/10.1016/j.cpr.2005.12.002.

Wang *et al. BMC Psychiatry*   (2025) 25:301

Page 11 of 11

19. Cheng J, Sun YH. Depression and anxiety among left-behind children in C hina: a systematic review. Child Care Health Dev. 2015;41(4):515–23. https://doi.org/10.1111/cch.12221.

20. Wang YY, Xiao L, Rao WW, Chai JX, Zhang SF, Ng CH, et al. The prevalence of depressive symptoms in "left-behind children" in China: a meta-analysis of comparative studies and epidemiological surveys. J Affect Disord. 2019;244:209–16. https://doi.org/10.1016/j.jad.2018.09.066.

21. Xiao Y, Ran H, Fang D, Che Y, Peng J, Sun H, et al. Childhood maltreatment and depressive disorders in Chinese children and adolescents: A population-based case-control study. Asian J Psychiatr. 2022;78: 103312. https://doi.org/10.1016/j.ajp.2022.103312.

22. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry. 1997;36(7):980–8. https://doi.org/10.1097/00004583-199707000-00021.

23. Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ. 2019;365: l1476. https://doi.org/10.1136/bmj.l1476. 23. Erratum in: BMJ. 2019 Apr 12;365:l1781.

24. 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. https://doi.org/10.1016/j.genhosppsych.2015.11.005.

25. Feng Y. The relation of adolescents' self-harm behaviors, individual emotion characteristics and family environment factors. Central China Normal University. 2008. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=2008116844.nh&DbName=CMFD2008 (In Chinese).

26. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5. https://doi.org/10.1002/1097-0142(1950)3:1%3c32::aid-cncr2820030106%3e3.0.co;2-3.

27. Chen YQ, Huang XJ, Yang F, Yang JJ, Zhong J, Yao KM, et al. A Chinese adaptation of the Patient Health Questionnaire for Adolescents (PHQ-A): factor structure and psychometric properties. BMC Psychiatry. 2024;24(1):331. https://doi.org/10.1186/s12888-024-05783-3.

28. Reardon T, Spence SH, Hesse J, Shakir A, Creswell C. Identifying children with anxiety disorders using brief versions of the Spence Children's Anxiety Scale for children, parents, and teachers. Psychol Assess. 2018;30(10):1342–55. https://doi.org/10.1037/pas0000570.

29. Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. Br J Gen Pract. 2007;57(541):650–2. https://bjgp.org/content/57/541/650.short.

30. Löwe B, Gräfe K, Zipfel S, Witte S, Loerch B, Herzog W. Diagnosing ICD-10 depressive episodes: superior criterion validity of the Patient Health Questionnaire. Psychother Psychosom. 2004;73(6):386–90. https://doi.org/10.1159/000080393.

31. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry. 2007;29(5):388–95. https://doi.org/10.1016/j.genhosppsych.2007.06.004.

32. Borghero F, Martínez V, Zitko P, Vöhringer PA, Cavada G, Rojas G. Screening depressive episodes in adolescents. Validation of the Patient Health Questionnaire-9 (PHQ-9). Rev Med Chil. 2018;146(4):479–486. https://doi.org/10.4067/s0034-98872018000400479.

33. Mossman SA, Luft MJ, Schroeder HK, Varney ST, Fleck DE, Barzman DH, Gilman R, DelBello MP, Strawn JR. The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: Signal detection and validation. Ann Clin Psychiatry. 2017;29(4):227–234A. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5765270/.

34. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. Compr Psychiatry. 2011;52(1):96–101. https://doi.org/10.1016/j.comppsych.2010.04.013.

35. Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, et al. Validity of the Patient Health Questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Res. 2013;210(2):653–61. https://doi.org/10.1016/j.psychres.2013.07.015.

36. Kendel F, Wirtz M, Dunkel A, Lehmkuhl E, Hetzer R, Regitz-Zagrosek V. Screening for depression: Rasch analysis of the dimensional structure of the PHQ-9 and the HADS-D. J Affect Disord. 2010;122(3):241–6. https://doi.org/10.1016/j.jad.2009.07.004.

37. Hack J, Martin G. Expressed Emotion, Shame, and Non-Suicidal Self-Injury. Int J Environ Res Public Health. 2018;15(5):890. https://doi.org/10.3390/ijerph15050890.

38. Lewinsohn PM, Gotlib IH, Lewinsohn M, Seeley JR, Allen NB. Gender differences in anxiety disorders and anxiety symptoms in adolescents. J Abnorm Psychol. 1998Feb;107(1):109–17. https://doi.org/10.1037//0021-843x.107.1.109.

39. Albano A M, Krain A. Anxiety and anxiety disorders in girls//Handbook of behavioral and emotional problems in girls. Boston, MA: Springer US, 2005: 79–116.

40. Howell HB, Brawman-Mintzer O, Monnier J, Yonkers KA. Generalized anxiety disorder in women. Psychiatr Clin North Am. 2001;24(1):165–78. https://doi.org/10.1016/s0193-953x(05)70212-4.

41. Parkerson HA, Thibodeau MA, Brandt CP, Zvolensky MJ, Asmundson GJ. Cultural-based biases of the GAD-7. J Anxiety Disord. 2015;31:38–42. https://doi.org/10.1016/j.janxdis.2015.01.005.

42. Carranza AB, Wallis CRD, Jonnson MR, Klonsky ED, Walsh Z. Nonsuicidal Self-Injury and Intimate Partner Violence: Directionality of Violence and Motives for Self-Injury. J Interpers Violence. 2022;37(3–4):1688–707. https://doi.org/10.1177/0886260520922372.

43. Wilkinson PO, Qiu T, Neufeld S, Jones PB, Goodyer IM. Sporadic and recurrent non-suicidal self-injury before age 14 and incident onset of psychiatric disorders by 17 years: prospective cohort study. Br J Psychiatry. 2018;212(4):222–6. https://doi.org/10.1192/bjp.2017.45.

44. Kiekens G, Hasking P, Claes L, Mortier P, Auerbach RP, Boyes M, et al. The DSM-5 nonsuicidal self-injury disorder among incoming college students: Prevalence and associations with 12-month mental disorders and suicidal thoughts and behaviors. Depress Anxiety. 2018;35(7):629–37. https://doi.org/10.1002/da.22754.

45. Andrewes HE, Hulbert C, Cotton SM, Betts J, Chanen AM. Relationships between the frequency and severity of non-suicidal self-injury and suicide attempts in youth with borderline personality disorder. Early Interv Psychiatry. 2019;13(2):194–201. https://doi.org/10.1111/eip.12461.

46. Chai Y, Luo H, Wong GHY, Tang JYM, Lam TC, Wong ICK, et al. Risk of self-harm after the diagnosis of psychiatric disorders in Hong Kong, 2000–10: a nested case-control study. Lancet Psychiatry. 2020Feb;7(2):135–47. https://doi.org/10.1016/S2215-0366(20)30004-3.

47. Zhang B, Zhang W, Sun L, Jiang C, Zhou Y, He K. Relationship between alexithymia, loneliness, resilience and non-suicidal self-injury in adolescents with depression: a multi-center study. BMC Psychiatry. 2023;23(1):445. https://doi.org/10.1186/s12888-023-04938-y.

48. Hu R, Peng LL, Du Y, Feng YW, Xie LS, Shi W, Jia P, Jiang LH, Zhao L. Reciprocal effect between non-suicidal self-injury and depressive symptoms in adolescence. Front Public Health. 2024;11:1243885. https://doi.org/10.3389/fpubh.2023.1243885.

49. Brown MZ, Comtois KA, Linehan MM. Reasons for suicide attempts and non-suicidal self-injury in women with borderline personality disorder. J Abnorm Psychol. 2002;111(1):198–202. https://doi.org/10.1037/0021-843X.111.1.198.

50. Serra M, Presicci A, Quaranta L, Caputo E, Achille M, Margari F, et al. Assessing Clinical Features of Adolescents Suffering from Depression Who Engage in Non-Suicidal Self-Injury. Children (Basel). 2022;9(2):201. https://doi.org/10.3390/children9020201.

51. Plener PL, Kaess M, Schmahl C, Pollak S, Fegert JM, Brown RC. Nonsuicidal Self-Injury in Adolescents. Dtsch Arztebl Int. 2018;115(3):23–30. https://doi.org/10.3238/arztebl.2018.0023.

52. Wu Y, Zhang Y, Wang C, Huang B. A meta-analysis on the lifetime and period prevalence of self-injury among adolescents with depression. Front Public Health. 2024;12:1434958. https://doi.org/10.3389/fpubh.2024.1434958.

53. Burke TA, Ammerman BA, Hamilton JL, Stange JP, Piccirillo M. Nonsuicidal self-injury scar concealment from the self and others. J Psychiatr Res. 2020;130:313–20. https://doi.org/10.1016/j.jpsychires.2020.07.040.

Publisher's Note