Validation and Clinical Utility of the Patient Health Questionnaire-9 and Center for Epidemiologic Studies Depression Scale as Depression Screening Tools in Collegiate Student-Athletes

Lindsey Keenan, PhD, LAT, ATC*; Yvette Ingram, PhD, LAT, ATC†; Branden Green, MS, LAT, ATC‡; Rachel Daltry, PsyD*; Sebastian Harenberg, PhD§

*West Chester University, PA; †Lock Haven University, PA; ‡Ursinus College, Collegeville, PA; §St Francis Xavier University, Antigonish, Nova Scotia, Canada

Context: The National Athletic Trainers' Association recommends including mental health screening measures as part of the preparticipation examination for all student-athletes (SAs). Despite this recommendation, most mental health screening tools have not been validated in the SA population.

Objective: To validate and examine the clinical utility of 2 depression screening tools in the collegiate SA population.

Design: Cross-sectional mixed-methods study.

Setting: Two Northeastern United States university athletics programs.

Patients or Other Participants: A total of 881 (men = 426, 48.4%; women = 455, 51.6%; mean age = 19.7 ± 1.4 years) National Collegiate Athletic Association Division II collegiate SAs completed the Patient Health Questionnaire-9 (PHQ-9) and Center for Epidemiologic Studies Depression Scale (CES-D); 290 SAs participated in a Mini-International Neuropsychiatric Interview.

Main Outcome Measure(s): Depression symptoms were measured using 2 self-report depression screening tools, the PHQ-9 and CES-D, during the fall preparticipation examination. The SAs were selected using a random stratified sampling technique to participate in a Mini-International Neuropsychiatric Interview as the reference standard comparison for the receiver operating characteristic analysis.

Results: A cutoff score of 6 on the PHQ-9 corresponded to 78% sensitivity, 75% specificity, 17.3% positive predictive value, 98.1% negative predictive value (NPV), 3.2 positive likelihood ratio (+LR), and 0.3 negative likelihood ratio (−LR). A cutoff score of 15 on the CES-D corresponded to 83% sensitivity, 78% specificity, 19.7% positive predictive value, 98.6% NPV, 3.7 +LR, and 0.22 −LR.

Conclusions: This was the first study to validate depression screening tools in the collegiate SA population. The results suggest cutoff scores on the PHQ-9 and CES-D in SA may need to be lower than those recommended for the general population and provide strong evidence for use as screeners to rule out depression. Referral and confirmatory testing should be implemented to confirm the presence of depression for SAs scoring at or above the cutoff thresholds. Given its brevity, inclusion of a suicidality or self-harm question and evidence of −LR and NPV strength, the PHQ-9 is a practical and effective screener for the SA population.

Key Words: mental health, psychology, preparticipation examination

---

Key Points

- Athletic trainers have the unique opportunity to incorporate mental health screening measures into clinical practice during the preparticipation examination and throughout clinical practice.
- A cutoff of 6 on the Patient Health Questionnaire-9 or 15 on the Center for Epidemiologic Studies Depression Scale provides strong evidence for clinical utility in ruling out depression in the collegiate student-athlete population; clinicians should carefully select measures and cutoff scores based on the available evidence and resources.
- Clinicians may consider lowering the cutoff scores for both the Patient Health Questionnaire-9 and Center for Epidemiologic Studies Depression Scale in the student-athlete population from those previously reported for the general population.

D epression is the second most common mental health concern in college students[1] and is a leading cause of disability across the globe.[2] Despite the benefits of regular physical activity, collegiate student-athletes (SAs) may be at the same or higher risk for mental health conditions, including depression, as the general population.[3,4] The literature examining the prevalence of depression in SAs is limited, with reports ranging from 10% to 26% across various self-report depression symptom measures.[3–9] Involvement in collegiate-level athletics may cause excessive stress due to a

number of factors. These include time constraints, increased performance expectations, and other required academic- and athletic-related commitments such as study hall, traveling for competition, and missing classes for games.[4] Unsurprisingly, the stressors on SAs have been linked to a higher prevalence of mental health disorder symptoms (eg, stress, anxiety, depression).[9] In addition, with the first collegiate sport season, as well as the high risk of injury in collegiate sports, these factors may cause more emotional and psychological distress than in nonathlete college students.[10–12] Predictors of depression in SAs have been described in collegiate SAs and are a notable concern for their health and well-being.[13]

To ensure the health, safety, and well-being of SAs, athletic trainers (ATs) and other athletic health care professionals need to identify possible mental health concerns in their patients.[14] In 2013, the National Athletic Trainers' Association (NATA) along with the National Collegiate Athletic Association (NCAA) and numerous other associations related to sports medicine and mental health formed an interassociation group to address psychological concerns in collegiate SAs.[14] The group developed a consensus statement with several recommendations for recognizing and referring collegiate SAs with mental health conditions. In 2016, the NCAA also published a best-practices document aimed at understanding and supporting SAs' mental wellness.[15] Both publications recommended screening SAs for mental health concerns in the preparticipation examination (PPE) and provided examples of screening tools used in the health care industry. When implementing mental health screening into clinical practice, careful consideration is needed in choosing appropriate measures, interpreting scores, and applying them clinically for patient benefit.

Across the United States, ATs are beginning to adopt mental health screening tools into clinical practice and PPEs to meet the current mental health screening recommendations of the NATA and NCAA.[16] These recommendations include adopting mental health screening into PPEs with appropriate referral to a mental health professional for any SAs with elevated mental health symptoms based on screening results,[14,15] yet Kroshus[16] noted significant variability in mental health screening practices across NCAA divisions; fewer than one-third of US sports medicine departments surveyed implemented depression screening in 2016. Additionally, whether validated measures were used by these institutions is unclear, as this was not examined.[16]

Despite the numerous open-source mental health screeners widely available for clinical use for the general population, only the Generalized Anxiety Disorder 7-item scale (GAD-7) has been validated in collegiate SAs with identified sensitivity and specificity, and a lower cutoff score was identified.[17] Regarding depression, Long et al[18] recently conducted a factorial validation study on the adolescent version of the Patient Health Questionnaire-9 (PHQ-9) in high school SAs but did not provide clinically relevant cutoff scores. The PHQ-9[19] and the Center for Epidemiologic Studies Depression Scale (CES-D)[20] are 2 widely available depression screening measures used in the general population. These measures are free to access and easy to administer and score.[19,20] The factorial validity of the PHQ-9 was sufficient for use in the collegiate student population,[21] but neither the PHQ-9 nor the CES-D has been validated in the SA population. With minimal information available on these tools in SA populations, their clinical utility is limited; it is possible ATs who implement depression screening may be relying on cutoff scores for nonathlete populations to trigger a referral to a mental health care professional. To our knowledge, no depression screening tool has been validated in a collegiate SA population.

Therefore, the purpose of our study was to validate depression screening tools in the collegiate SA population. The specific aims of this study were to determine (1) whether the PHQ-9 and the CES-D depression screening tools are valid and reliable measures to detect clinically relevant depression symptoms in collegiate SAs and (2) the suggested cutoff scores for these depression screening tools in collegiate SAs based on sensitivity and specificity. We hypothesized that both the PHQ-9 and the CES-D would be valid measures in discriminating between collegiate SAs who did and those who did not meet the diagnostic criteria for clinical depression. We also hypothesized that the optimal cutoff scores based on sensitivity and specificity for both screening tools would be lower than for the general population.

METHODS

Study Design

We used a cross-sectional mixed-methods design to validate 2 depression screening tools in collegiate SAs. Methods consisted of administration of 2 quantitative measures of depression symptoms, the PHQ-9 and the CES-D, as well as a qualitative measure of depression, the Mini-International Neuropsychiatric Interview (MINI), in a collegiate SA population. Independent variables were sex, sport, and meeting the criteria for depression using the MINI. Dependent variables were screening tool total scores as measured by the PHQ-9 and CES-D.

Participants

All 943 SAs from 2 NCAA Division II universities in the northeast United States were invited to participate in this study during the in-person fall PPE on campus. Exclusion criteria were any SA under the age of 18 years or not a member of an institution's NCAA varsity athletic team. The study was reviewed and approved by both university institutional review boards.

Instrumentation

Patient Health Questionnaire-9. The PHQ-9 is a brief 9-item, self-report depression tool that is a reliable and valid measure used as a diagnostic screening for major depressive disorder (MDD) in the general population.[19] A PHQ-9 score ≥ 10 has 88% sensitivity and 88% specificity when using a structured neuropsychiatric interview as the criterion standard for major depression.[19,22] Total scores on the PHQ-9 are grouped into depression symptom levels, as described by Kroenke et al.[19] A score of 0 represents *no symptoms*; 1 through 4, *minimal symptoms*; 5 through 9, *mild symptoms*; 10 through 14, *moderate symptoms*; 15 through 19, *moderately severe symptoms*; and 20 through 27, *severe symptoms*. For clinical follow-up purposes in this study, a red flag for clinically relevant depression symptoms was determined by a score of ≥ 10 based on validation in the general population.[19,22] Any SA who scored in the red-flag range met privately with the AT and was offered a referral to a mental health provider. If any SA indicated he or she was having "thoughts that you would be better off dead or of hurting

yourself in some way" on question 9 of the PHQ-9, they were automatically red flagged regardless of the total score and offered a referral.

Center for Epidemiologic Studies Depression Scale. The CES-D is a 20-item, self-report depression scale that was developed as part of a National Institute of Mental Health study to measure depression and has been validated in the general and college student populations.[20,23] A meta-analysis of the validation of the CES-D in the general population indicated that a cutoff of ≥16 had a sensitivity of 87% and specificity of 70%.[24] The authors also reported 83% sensitivity and 78% specificity with a cutoff of ≥20 representing clinically relevant symptoms, suggesting that the latter might be a better tradeoff when interpreting scores.[24] The total scores of the CES-D in this study were grouped into depression symptom levels, as previously determined by Radloff.[20,23] A score of 0 represents *no symptoms*; 1 through 15, *mild symptoms*; 16 through 23, *moderate symptoms*; and 24 through 60, *severe symptoms*. Using previously identified cutoff scores for the young adult population,[23] a score of ≥16 was considered a red flag for clinically relevant depression symptoms in this study to provide a follow-up referral for the SA.

Mini-International Neuropsychiatric Interview. The MINI 6.0 is a shortened, structured diagnostic neuropsychiatric interview.[25] A structured clinical interview is considered the criterion standard for diagnosing mental health concerns based on the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5) criteria.[26] The MINI was developed by psychiatrists and clinicians for diagnosing the most common psychiatric disorders, including MDD, as defined by the DSM-5.[25,26] The MINI was designed as a shorter version of the traditional clinical psychiatric interview to enable accurate, structured clinical interviews in research and clinical settings.[25] It has been widely used in research as a reference standard for validation of mental health assessments and surveys.[22,24,25]

For this study, the MDD module of the MINI was used, which asks about depressive symptoms over the past 2 weeks, including feeling down, sad, or hopeless; loss of interest in things one used to enjoy; changes in eating and sleeping patterns; changes in locomotion; loss of energy; feelings of worthlessness or guilt; difficulty concentrating; and thoughts about death or suicide. For each *yes* response, a point is scored. With a score ≥5, including feeling down, sad, or hopeless and a loss of interest, the participant was coded as meeting the criteria for depression[25] and assigned to the positive MINI depression group for this study. Student-athletes who met the criteria for depression or indicated thoughts about death or suicide during the MINI were offered a referral to counseling services.

Procedures

The SAs completed the depression screening tools as part of their PPEs. At the PPE, they were given a consent form describing the research study and its purpose. The consent form stated that only the ATs and researchers would have access to their mental health screening results. Consenting participants completed electronic versions of the PHQ-9, CES-D, and a demographic questionnaire, administered by ATs at both participating institutions. The SAs provided their names on the screening tools as part of the PPE, which allowed the ATs to identify any SA with clinically relevant depression symptoms. If any SA's screening tool results

indicated clinically relevant depression symptoms based on cutoff scores for the general population (≥10 on the PHQ-9, ≥16 on the CES-D, or "thoughts that you would be better off dead or of hurting yourself in some way" on PHQ-9 question 9), the SA was offered a referral to the participating institution's on-campus counseling services center by his or her AT. All ATs at both institutions were educated and trained on mental health screening and referral protocols, as per the mental health management plan at each institution.

Once the screenings were completed, 300 SAs were selected using a 2-phase random stratified sampling technique to participate in the MINI. This technique was implemented to recruit participants with depression screening scores throughout the possible total score range of 0 to 27 on the PHQ-9 as well as to replicate the percentage of SAs represented per sport. The sampling method involved first grouping participants by sport to accurately represent the collegiate SA population at both institutions. Several SAs from each sport who represented the overall percentage of SAs at each institution were invited to participate in the MINI. The calculations for the stratified sampling technique were based on a goal of 300 MINI participants, 150 from each institution. For example, at university 1, football athletes represented 18% of all athletes at that institution; therefore, 18%, or 27 football athletes would be invited to participate in the MINI.

For the second phase of this sampling method, the SAs were ranked based on their total PHQ-9 score, and every third athlete from each score range was invited to participate in the MINI, until our target sample for that sport was achieved. This second phase (score ranking) was conducted to ensure representation across all possible scores on the PHQ-9 in order to determine a suggested cutoff score in this population. In some cases, we undersampled or oversampled from the sport to recruit the necessary athletes in each score range.

Six research assistants (RAs) who were graduate student clinicians enrolled in the master of counseling, master of social work, or master of school counseling program at either institution in this study were recruited to administer the MINIs. All RAs had prior education and training in clinical interviewing and completed 2 live, interactive training sessions on administration of the MINI, led by a licensed sport psychologist. The RAs were unaffiliated with the athletics and athletic training departments.

We used the MINI as the reference standard to determine whether the SAs met the criteria for clinical depression and the MINI results to determine the validity of the PHQ-9 and CES-D. All SAs who participated in the MINI received a $10 gift card to a local business as a participant incentive. The RAs administering the MINI were blinded to the PHQ-9 and CES-D depression screening results.

Data Analysis

The area under the curve (AUC) was used to indicate the maximal discrimination of the PHQ-9 and CES-D between patients with and those without depression symptoms based on meeting the criteria for a current MDD using the MINI. The cutoff values for the PHQ-9 and CES-D were determined using their respective sensitivity and specificity and the Youden index ([sensitivity + specificity] − 1), which identifies the cutoff score that optimally balances sensitivity and specificity. Positive and negative likelihood ratios (+LRs, −LRs), positive and negative predictive values (PPVs, NPVs),

Table 1. Participants by Sport (N = 881)

Sport	No. (%)
Football	170 (19.3)
Baseball	62 (7.0)
Women's soccer	60 (6.8)
Men's soccer	53 (6.0)
Women's lacrosse	53 (6.0)
Field hockey	51 (5.8)
Softball	43 (4.9)
Women's swimming and diving	39 (4.4)
Women's track and field	37 (4.2)
Women's volleyball	34 (3.9)
Women's rugby	34 (3.9)
Women's cross-country	29 (3.3)
Men's basketball	28 (3.2)
Men's track and field	27 (3.1)
Wrestling	26 (2.9)
Men's cross-country	25 (2.8)
Women's basketball	23 (2.6)
Men's swimming and diving	22 (2.5)
Women's gymnastics	21 (2.4)
Cheerleading	15 (1.7)
Women's tennis	9 (1.0)
Men's golf	7 (0.8)
Women's golf	7 (0.8)
Men's tennis	6 (0.7)

Table 2. Depression Symptom Levels on the Patient Health Questionnaire-9 and Center for Epidemiologic Studies Depression Scale (N = 881)

	No. (%)	
Depression Symptom Level	Patient Health Questionnaire-9	Center for Epidemiologic Studies Depression Scale
None	412 (46.8)	59 (6.7)
Minimal	313 (35.5)	NA
Mild	120 (13.6)	720 (81.7)
Moderate	25 (2.8)	60 (6.8)
Moderately severe	10 (1.1)	NA
Severe	1 (0.1)	42 (4.8)

Abbreviation: NA, not applicable.

and 95% CIs associated with the estimates of sensitivity and specificity for the Youden index cutoff scores were computed to assess clinical utility. Concurrent validity was assessed using a Pearson correlation between total PHQ-9 and total CES-D scores. Cronbach α was calculated to identify the internal consistency of the PHQ-9 and CES-D in the SA population. General descriptive statistics were used to examine the demographics and main outcome measure scores. Significance was accepted when $P < .05$ using SPSS (version 25.0; IBM Corp).

RESULTS

Participants

A total of 943 SAs were recruited during PPEs; 881 (93.4% response rate) completed the PHQ-9 and CES-D. Participants were nearly evenly split between men (n = 426, 48.4%) and women (n = 455, 51.6%), with 516 (58.6%) SAs from university 1 and 365 (41.4%) from university 2. The SAs were 19.7 ± 1.4 years old, and 4.9% (n = 43) identified their ethnicity as Hispanic. Most individuals identified as White (n = 693, 78.7%), followed by Black or African American (n = 126, 14.3%), 2+ races or mixed race (n = 49, 5.6%), and Asian (n = 10, 1.1%). Most SAs in the study engaged in football (n = 171, 19.4%), followed by baseball (n = 62, 7%) and women's soccer (n = 60, 6.8%; Table 1). Participants were nearly evenly split between underclassmen (freshmen and sophomores, n = 474, 53.8%) and upperclassmen (juniors, seniors, fifth-year seniors, graduate students, n = 407, 46.3%). In the medical histories collected as part of the study during the fall preseason PPE, 157 (17.8%) SAs reported a family history of depression, while 68 (7.7%) indicated being previously diagnosed with depression. Additionally, 24 (2.7%) SAs noted they were currently in therapy for depression treatment, and 26 (3%) stated they currently took medication for depression.

The PHQ-9 Results

Internal consistency of the PHQ-9 was good (Cronbach α = .84). A total of 36 (4%) SAs were red flagged on the PHQ-9 for clinically relevant depression symptoms (moderate to severe), using the general population cutoff score of ≥10. Depression symptom level results based on cut scores for the general population are found in Table 2. The mean PHQ-9 score of all 881 participants was 2.14 ± 3.23. Total scores ranged from 0 to 23.

The CES-D Results

The internal consistency of the CES-D was acceptable (Cronbach α = .79). Most SAs (n = 720, 81.7%) reported mild depression symptoms (Table 2). Using the cutoff score for the general population of ≥16, 102 (11.6%) SAs were red flagged on the CES-D for clinically relevant depression symptoms. The mean CES-D score of all 881 participants was 8.33 ± 7.17. Total scores ranged from 0 to 46.

Concurrent Validity of the PHQ-9 and CES-D

Of the 105 SAs who were red flagged during the PPE screening, 2.9% (n = 3) were flagged solely on the PHQ-9, 65.7% (n = 69) solely on the CES-D, and 31.4% (n = 33) on both measures, using the previously identified cutoff scores from the general population.[22] The Pearson correlation between the PHQ-9 and CES-D mean scores revealed a significant positive relationship ($r = 0.76$, n = 881, $P < .001$), confirming good concurrent validity of the PHQ-9 and CES-D.

The MINI Results

Within 2 weeks of completing the PHQ-9 and CES-D, a total of 290 (96.7%) SAs were administered a MINI, which took on average 11.9 ± 5.9 minutes to complete. This sample relatively matched the total participation in representation of men (n = 125, 43.1%), women (n = 165, 56.9%), and age (mean = 19.8 ± 1.3 years). Race was also nearly identical to the overall representation of White (n = 231, 79.7%), Black or African American (n = 41, 14.1%), 2+ or mixed race (n = 16, 5.5%), and Asian (n = 2, 0.7%) SAs. As in the overall sample, SAs were nearly evenly split between underclassmen (n = 135, 46.5%) and upperclassmen (n = 155, 53.5%) as well as participation between institutions, with 148 (51%) SAs from university 1 and 142 (49%) from university 2. We were also able to closely replicate representation by sport across the sample (Figure 1). The random stratified sampling
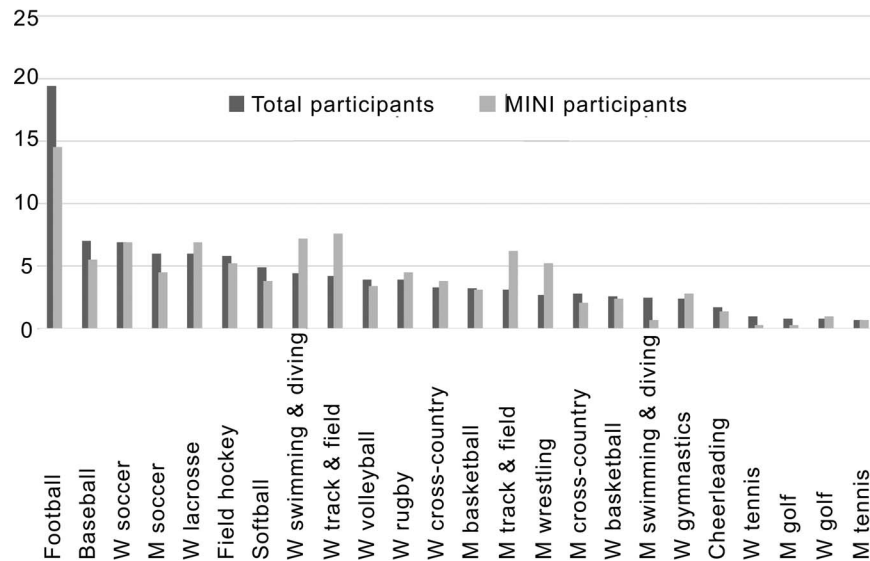
Figure 1. Mini-International Neuropsychiatric Interview (MINI) completers and total participants (%) by sport (N = 290). Abbreviations: M, men's; W, women's.

technique was successful in that SAs' PHQ-9 scores were represented across the total score range (Figure 2).

A total of 18 of the 290 SAs (6.2%) met the criteria for a current episode of MDD on the MINI and were grouped into the positive MINI depression group. The receiver operating characteristic (ROC) curves for the PHQ-9 and CES-D are shown in Figures 3 and 4, respectively. Based on the ROC analysis, the AUC for the PHQ-9 was 0.81 (95% CI = 0.71, 0.92) and for the CES-D was 0.84 (95% CI = 0.74, 0.95).

Clinical Utility Results

Sensitivity and specificity were calculated using the AUC, and the Youden J was computed to determine the optimal cut-off score between sensitivity and specificity for both the PHQ-9 (Table 3) and CES-D (Table 4). Based on the Youden index, the optimal cutoff score for clinically relevant depression symptoms was a total score of ≥ 6 on the PHQ-9 and ≥ 15 on the CES-D in SAs. A cutoff score of 6 on the PHQ-9 corresponded to a sensitivity of 78% (95% CI = 56%,

93%), specificity of 75% (95% CI = 70%, 80%), PPV of 17.3% (95% CI = 10.1%, 26.5%), NPV of 98.1% (95% CI = 95.6%, 99.4%), +LR of 3.2 (95% CI = 2.3, 4.4) and −LR of 0.3 (95% CI = 0.1, 0.7). A cutoff score of 15 on the CES-D corresponded to a sensitivity of 83% (95% CI = 62.3%, 95.6%) and specificity of 78% (95% CI = 72.4%, 82.3%), PPV of 19.7% (95% CI = 11.9%, 29.6%), NPV of 98.6% (95% CI = 96.4%, 99.6%), +LR of 3.7 (95% CI = 2.7, 5.0), and −LR of 0.22 (95% CI = 0.1, 0.7).

DISCUSSION

Our purpose was to validate depression screening tools in the collegiate SA population. We accepted our hypotheses
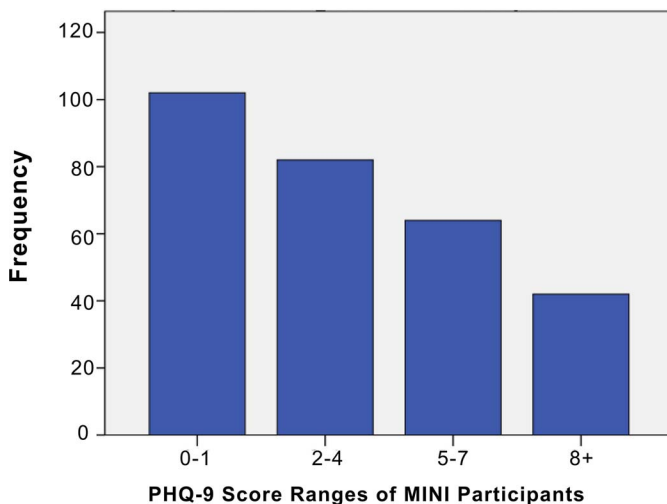

Figure 2. Patient Health Questionnaire-9 (PHQ-9) score ranges of Mini-International Neuropsychiatric Interview (MINI) participants.
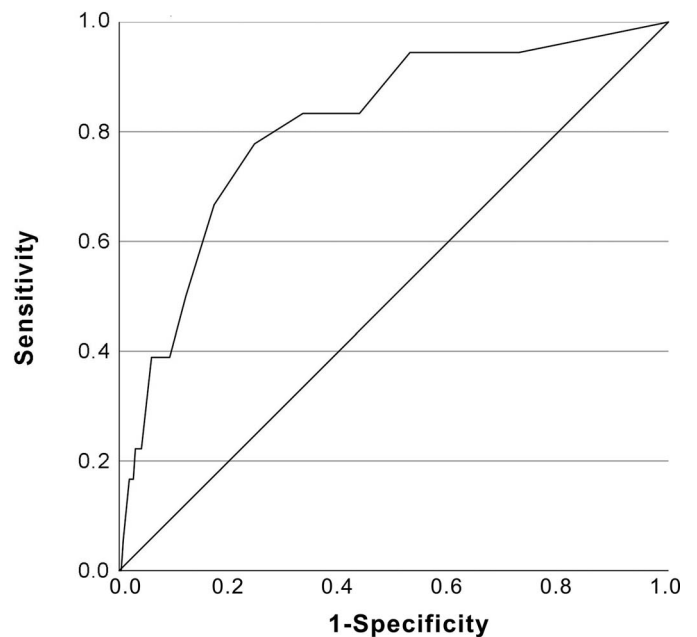

Figure 3. Receiver operating characteristic curve for the Patient Health Questionnaire-9 versus Mini-International Neuropsychiatric Interview.
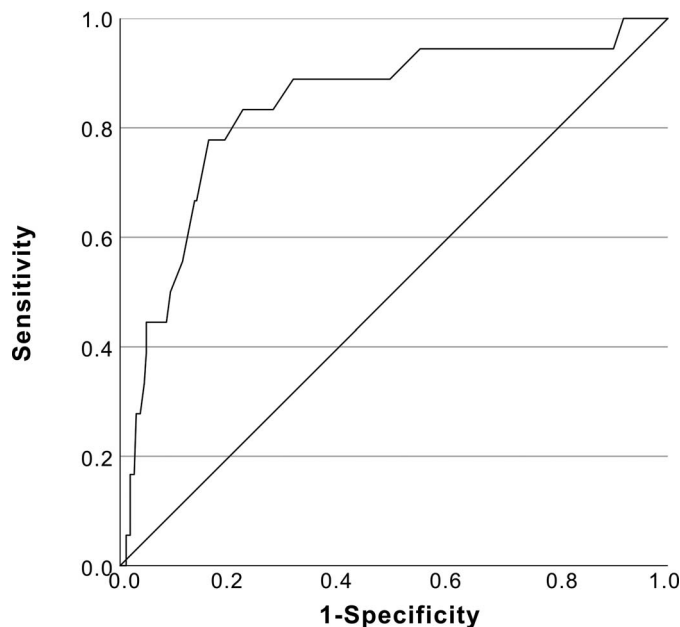
Figure 4. Receiver operating characteristic curve for the Center for Epidemiologic Studies Depression Scale versus Mini-International Neuropsychiatric Interview.

that both the PHQ-9 and CES-D would be valid measures in discriminating between collegiate SAs with and those without clinical levels of depression. The AUC measures accuracy and determines the ability of a test to distinguish between 2 groups, diseased and nondiseased or healthy participants and describes the fundamental validity of a test.[27] The AUC indices for the PHQ-9 and CES-D were 0.81 and 0.84, respectively. An AUC index of 0.5 is considered chance discrimination in correctly identifying groups; an AUC equal to 1 occurs when the test perfectly discriminates between groups.[27] Therefore, an AUC = 0.8 would be considered acceptable validity in our sample of SAs for both the

Table 3. Sensitivity, Specificity, and Youden J Results of the Patient Health Questionnaire-9

| Score | % | | Youden J Result |
	Sensitivity	Specificity	
0	100	0	0
1	94.4	27.2	0.216
2	94.4	37.1	0.315
3	94.4	47.1	0.415
4	83.3	56.2	0.395
5	83.3	66.5	0.498
6	77.8	75.4	0.532[a]
7	66.7	82.7	0.494
8	50	87.9	0.379
9	38.9	90.8	0.297
10	38.9	94.1	0.33
11	22.2	96	0.182
12	22.2	97.1	0.193
13	16.7	97.4	0.141
14	16.7	97.8	0.145
15	16.7	98.2	0.149
16	5.6	99.3	0.049
17	0	99.6	−0.004
18	0	1	0

[a] Highest Youden J value.

Table 4. Sensitivity, Specificity, and Youden J Results of the Center for Epidemiologic Studies Depression Scale

| Score | % | | Youden J Result |
	Sensitivity	Specificity	
0	100	0	0
1	100	6.6	0.066
2	100	8.1	0.081
3	94.4	9.9	0.043
4	94.4	18	0.124
5	94.4	23.2	0.176
6	94.4	31.2	0.256
7	94.4	37.1	0.315
8	94.4	41.5	0.359
9	94.4	45.2	0.396
10	88.9	50.7	0.396
11	88.9	55.9	0.448
12	88.9	59.9	0.488
13	88.9	68.4	0.573
14	83.3	72.1	0.554
15	83.3	77.6	0.609[a]
16	77.8	80.9	0.587
17	77.8	83.8	0.616
18	66.7	86	0.527
19	66.7	86.4	0.531
20	55.6	88.6	0.442
21	50.0	90.8	0.408
22	44.4	91.5	0.359
23	44.4	92.6	0.37
24	44.4	93	0.374
25	44.4	95.2	0.396
26	38.9	95.2	0.341
27	33.3	95.6	0.289
28	27.8	96.3	0.241
29	27.8	97.1	0.249
30	16.7	97.4	0.141
31	16.7	98.2	0.149
32	11.1	98.2	0.093
33	5.6	98.2	0.038
34	5.6	98.5	0.041
35	5.6	98.9	0.045
36	0	98.9	−0.011

[a] Highest Youden J.

PHQ-9 and CES-D. This is not surprising, given that both measures have been well documented as valid in other populations.[22,24]

Second, we accepted our hypotheses that the cutoff scores based on sensitivity and specificity for both the PHQ-9 and CES-D in SAs would be lower values than reported in the general population. The generally accepted cutoff score for the PHQ-9 for clinically relevant depression symptoms (moderate to severe symptoms) is 10.[22] At the cut point of 10, the sensitivity was 38.9%, and specificity was 94.1% in our sample. Comparatively, we found that a lower cutoff score of ≥6 on the PHQ-9 in SAs in our study optimized sensitivity and specificity at 77.8% and 75.4%, respectively. Researchers[22] have suggested a lower cutoff score on the PHQ-9 and recommended identifying specific cutoff scores for various settings. In contrast, the generally accepted cutoff score for clinically relevant depression symptoms on the CES-D is 16, but a score of ≥20 is often considered acceptable.[24] Similarly, we found an optimal cutoff of 15 in SAs using the Youden J with a sensitivity and specificity of 83.3% and 77.6%, respectively, while a ≥20 cutoff would yield a sensitivity of 55.6% and specificity of 88.6%. Authors[24,28,29]

of a few studies have observed lower cutoff scores on the CES-D to be the appropriate balance between sensitivity and specificity in other populations.

It is important to note that a cutoff score, as indicated by the AUC and appropriate index (eg, Youden J), should be determined by each clinician based on considered use of the measure and the relevant literature. For example, in epidemiologic research, the optimal balance between sensitivity and specificity (Youden index) is ideal; however, in the clinical setting, higher sensitivity, obtained by lowering the cutoff score, may be ideal when attempting to minimize missed cases.[27] More importantly, an effective screening process or instrument for clinical utility is typically useful in ruling out a health condition and relies on further testing to confirm a diagnosis, as seen in the debate around cardiovascular screening in the SA population.[30] The estimates of clinical utility in our study provide strong evidence that the PHQ-9 and CES-D are effective in ruling out depression; a high NPV corresponded with the 6+ cutoff score on the PHQ-9 and a sensitivity of 77.8%. Clinicians may choose to use this or another cutoff score that optimizes sensitivity to increase the probability that a SA without depression scores below this cutoff threshold. Conversely, minimizing the false-positive rate (ie, increasing specificity) may be considered in settings with limited resources, particularly for ATs with high patient-to-provider ratios who cannot perform numerous follow-up referrals with a high potential for false-positives; however, increasing specificity comes at the cost of more false-negatives. Furthermore, false-positive results may also have a negative effect on the patient, as found with other health concerns, such as cancer screening potentially leading to unnecessary medical interventions and patient stress.[31] For this reason, an optimal balance between sensitivity and specificity as identified in this study may be ideal for many ATs.

Mental health screening conducted by ATs should serve as an initial symptom screener. Despite the multidisciplinary approach to mental health suggested by the NCAA and NATA, it is necessary to emphasize that ATs cannot diagnose mental health conditions, and even when the PHQ-9 or CES-D is administered by a mental health professional, our results indicate neither should be implemented as a diagnostic tool in the SA population but rather as a screener to rule out depression. Given the LRs for the PHQ-9 and CES-D and the AT's acknowledged scope of practice, the AT must provide guidance for an SA during a referral after a positive depression screening. Positive screenings using the PHQ-9 or CES-D suggest a patient may be experiencing depression, but a confirmatory evaluation by a licensed mental health provider is necessary to determine the next steps in potential diagnosis and treatment. Additionally, as a cutoff score is used to identify SAs with symptoms that may reflect depression, many SAs may be struggling with depressive symptoms below this threshold and may still benefit from a referral to a mental health professional. Determining an appropriate balance between sensitivity and specificity on these depression screeners for various clinical scenarios is recommended.

The PHQ-9 and CES-D appear to be valid screening tools for depression symptoms in SAs. However, each tool has specific factors that may drive the use of one over another for clinical utility. For example, the PHQ-9 is a very brief measure, with only 9 symptom questions and 1 difficulty follow-up question, whereas the CES-D has 20 questions. When considering the use of a depression screening tool in a PPE and throughout clinical practice, the AT and athletic health care team may choose the PHQ-9, given the breadth of questions already burdening SAs regarding medical history and other screenings or patient-reported outcomes as well as evidence of NPV strength, as demonstrated in this study.

Furthermore, though we did not examine the factorial validity of the CES-D, it has been called into question in the literature,[32,33] and some researchers[24] have suggested it should not be used as a standalone measure of depression symptoms. With this in mind, the factor structure of the CES-D relevant to the collegiate SA population warrants further investigation. For example, in 1 study,[34] researchers cautioned against the use of the effort item ("everything I did was an effort") on the CES-D, particularly among Black men, for whom putting in effort is viewed as a positive trait. This item exhibited poor correlations and factor loading concerns for both positive and negative affect, as the item loaded simultaneously for both factors.[34] The authors pointed to theoretical considerations of effort in the community of Black men, among whom a high-effort active coping style, *John Henryism*, is adopted to overcome the demands of their environment.[35] Similar lines can be drawn for the SA population in a high-demand, high-stress environment, particularly for Black and African-American SAs. We noted that more than 25% of the SAs indicated "everything I did was an effort" at least a moderate amount to most of the time over the past week. It is possible the SAs varied in perceptions as to whether this was a positively or negatively worded item. Further evaluation is needed to explore the factorial structure of the CES-D across a diverse SA population.

Another important component to think about in choosing a screening tool is the inclusion of a question concerning suicidal ideation and self-harm. Before it is administered, this question should be carefully considered by the AT, team physician, and athletic health care team's mental health provider. The PHQ-9 includes a question about self-harm and thoughts of death and therefore should only be administered when a timely follow-up to that question is possible. The CES-D does not include a question on suicidal ideation or self-harm and therefore may not accurately replicate the total experience of depression; however, it can subsequently be administered without the need for immediate follow up. Despite this, the athletic health care team should contemplate inquiring about suicidal ideation and self-harm during depression screening referrals if using an instrument that does not include these items, as a depression screening may be one of the only opportunities an individual has to report suicidal thoughts or self-harm. This may be especially true in populations such as collegiate athletes, for whom the discussion of mental health is often inhibited by stigma and societal pressures.[36]

Student-athletes experienced several barriers to reporting distress and seeking help for their mental health, including stigma.[36] It is possible the stigma of mental health in sports and underreporting played a role in the lower cutoff scores on the depression screens we identified.[36–38] Researchers have characterized the likelihood of underreporting psychological distress by SAs and their reluctance to report struggles with mental health concerns.[39] More recently, the authors[40] of a meta-analysis of anonymous versus nonanonymous depression screening in SAs revealed that nonanonymous screening

methods only detected approximately half of the SAs struggling with depression. These phenomena have been well documented regarding concussion injury.[41] Similar to concussion, mental health concerns may be stigmatized as invisible health conditions that cannot be easily recognized by peers, coaches, or family.[42–44] Symptom underreporting may be due to pressure athletes experience to not be perceived as weak by their coaches or teammates or embarrassment for suffering from a mental health condition based on the negative mental health stigma in general society.[42]

Despite this, notable improvements in the conversation around mental health by highly regarded sports organizations such as the NATA[14] and the NCAA[15] as well as in the media may result in further acceptance and help to decrease the stigma of mental health in the sports field. Athletic trainers are often the first medical professional SAs may contact and interact with on a regular basis regarding their health.[14] Thus, ATs and sports medicine physicians in the collegiate and high school settings are in an advantageous position to normalize mental health as an aspect of well-being through screening and to identify and refer SAs who may be struggling with their mental health.[14,16] Athletic trainers should follow consensus recommendations to develop referral protocols for positive findings during mental health screenings in SAs.[14,15] The NPVs we found provide strong evidence that an athlete scoring below the reported cutoff scores (6 on the PHQ-9 and 15 on the CES-D) is unlikely to be experiencing depression. However, with positive screening results, referral is needed to verify the presence of depression for SAs scoring at or above cutoff thresholds. Implementing validated mental health screeners such as the PHQ-9 into the PPE and throughout athletic training services is warranted, and future investigators should examine the effect of screening on the mental health of SAs.

Limitations

This study had several limitations. First, the small number of SAs in the positive MINI group may have contributed to individual bias and decreased the strength of the ROC analysis. Further evaluation and adding to the pool of total participants would strengthen the outcomes. As previously mentioned, the lower cutoff scores we observed may be related to SAs possibly underreporting depression symptoms. It is important to point out the methods used when administering the PHQ-9 and the MINI. Student-athletes were asked to indicate their name on the screening as a part of the PPE for the AT to provide appropriate follow-up referrals. Identifiers were also necessary to match PHQ-9 and CES-D results with MINI outcomes. As the methods and clinical application of these tools were not anonymous, the SAs may have felt pressure to underreport or not report symptoms due to fear of identification.[40] Given the potential for underreporting on the PPE screenings, it is possible participants also underreported symptoms during the MINI. Future researchers may incorporate an oral confirmation of confidentiality of the MINI in addition to the consent form to potentially reduce the perceived risk of depression disclosure by SAs. Furthermore, we examined the validity of 2 commonly used depression measures, but the suggestion of lower cutoff scores in the SA population may not be applicable to other measures of depression or mental health screeners. Future authors should seek to validate the PHQ-9, CES-D, and other measures of mental health in the athlete population beyond the demographics in this study, especially as mental health screening becomes customary best practice in athletic training services for SAs.[14,16]

Clinical Significance

The clinical significance of our work is highlighted in the validation and determination of suggested cutoff scores for 2 depression measures in collegiate SAs. In an ideal world, the criterion standard of a psychiatric clinical diagnostic interview would be used to screen individuals for mental health concerns, but this is not reasonable given the limited time and resources in nearly every athletic training setting. Selecting and administering a valid depression screener is an ideal option when implementing mental health screening in the SA population. As recommended, mental health screening should only be implemented when a confirmed and ideally written mental health referral protocol is in place.[14,16] Furthermore, ATs should consider and weigh the balance between missed cases and false-negatives when determining a cutoff score for their specific setting and resources. Finally, as a cutoff score is used to identify SAs with symptoms that may indicate clinical depression, a referral to a licensed mental health provider is necessary for confirmatory evaluation before discussion of treatment. Student-athletes may also be struggling with depressive symptoms below this threshold or may not report symptoms on a screening but may still benefit from a referral to a mental health professional. Athletic trainers should use clinical judgment and consider offering a mental health referral for evaluation if any concerns about depression arise through observation, screening, or both.

CONCLUSIONS

Athletic trainers can incorporate depression screening into clinical practice through the PPE and throughout clinical practice. To our knowledge, this was the first study to validate any self-report depression measure and identify suggested cutoff scores in a collegiate SA population. Further research is needed to confirm our findings, but we recommend the use of lower cutoff scores for the SA population than recommended in the general population. Athletic health care teams should consider a valid depression screening tool such as the PHQ-9 and evaluate suggested cutoff scores based on the available evidence and available resources in their setting.

REFERENCES

1. Eisenberg D, Hunt J, Speer N. Mental health in American colleges and universities: variation across student subgroups and across campuses. *J Nerv Ment Dis*. 2013;201(1):60–67. doi:10.1097/NMD.0b013e31827ab077
2. Depressive disorder (depression). World Health Organization. Published March 31, 2023. Accessed July 20, 2023. http://www.who.int/mediacentre/factsheets/fs369/en/

3. Proctor SL, Boan-Lenzo C. Prevalence of depressive symptoms in male intercollegiate student-athletes and nonathletes. *J Clin Sport Psychol*. 2010;4(3):204–220. doi:10.1123/jcsp.4.3.204

4. Storch EA, Storch JB, Killiany EM, Roberti JW. Self-reported psychopathology in athletes: a comparison of intercollegiate student-athletes and non-athletes. *J Sport Behav*. 2005;28(1):86–98.

5. Hammond T, Gialloreto C, Kubas H, Davis H IV. The prevalence of failure-based depression among elite athletes. *Clin J Sport Med*. 2013;23(4):273–277. doi:10.1097/JSM.0b013e318287b870

6. McGuire LC, Ingram YM, Sachs ML, Tierney RT. Temporal changes in depression symptoms in male and female collegiate student-athletes. *J Clin Sport Psychol*. 2017;11(4):337–351. doi:10.1123/JCSP.2016-0035

7. Nixdorf I, Frank R, Hautzinger M, Beckmann J. Prevalence of depressive symptoms and correlating variables among German elite athletes. *J Clin Sport Psychol*. 2013;7(4):313–326. doi:10.1123/jcsp.7.4.313

8. Wolanin A, Hong E, Marks D, Panchoo K, Gross M. Prevalence of clinically elevated depressive symptoms in college athletes and differences by gender and sport. *Br J Sports Med*. 2016;50(3):167–171. doi:10.1136/bjsports-2015-095756

9. Yang J, Peek-Asa C, Corlette JD, Cheng G, Foster DT, Albright J. Prevalence of and risk factors associated with symptoms of depression in competitive collegiate student athletes. *Clin J Sport Med*. 2007;17(6):481–487. doi:10.1097/JSM.0b013e31815aed6b

10. Brewer BW, Linder DE, Phelps CM. Situational correlates of emotional adjustment to athletic injury. *Clin J Sport Med*. 1995;5(4):241–245. doi:10.1097/00042752-199510000-00006

11. Smith AM. Psychological impact of injuries in athletes. *Sports Med*. 1996;22(6):391–405. doi:10.2165/00007256-199622060-00006

12. Udry E, Gould D, Bridges D, Beck L. Down but not out: athlete responses to season-ending injuries. *J Sport Exerc Psychol*. 1997;19(3):229–248. doi:10.1123/JSEP.19.3.229

13. Wilson S, Harenberg S, Stilwell T, Vosloo J, Keenan L. Prevalence and predictors of depressive symptoms in NCAA Division III collegiate athletes. *J Athl Dev Exp*. 2022;4(1). doi:10.25035/jade.04.01.05

14. Neal TL, Diamond AB, Goldman S, et al. Inter-association recommendations for developing a plan to recognize and refer student-athletes with psychological concerns at the collegiate level: an executive summary of a consensus statement. *J Athl Train*. 2013;48(5):716–720. doi:10.4085/1062-6050-48.4.13

15. Mental health best practices. National Collegiate Athletic Association. Accessed July 20, 2023. https://www.ncaa.org/sports/2016/5/2/mental-health-best-practices.aspx

16. Kroshus E. Variability in institutional screening practices related to collegiate student-athlete mental health. *J Athl Train*. 2016;51(5):389–397. doi:10.4085/1062-6050-51.5.07

17. Tran AGTT. Using the GAD-7 and GAD-2 generalized anxiety disorder screeners with student-athletes: empirical and clinical perspectives. *Sport Psychol*. 2020;34(4):300–309. doi:10.1123/tsp.2020-0028

18. Long A, DeFreese JD, Bickett A, Price D. Factorial validity and invariance of an adolescent depression symptom screening tool. *J Athl Train*. 2022;57(6):592–598. doi:10.4085/1062-6050-343-21

19. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x

20. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385–401. doi:10.1177/014662167700100306

21. Keum BT, Miller MJ, Inkelas KK. Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. *Psychol Assess*. 2018;30(8):1096–1106. doi:10.1037/pas0000550

22. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ*. 2012;184(3):E191–E196. doi:10.1503/cmaj.110829

23. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *J Youth Adolesc*. 1991;20(2):149–166. doi:10.1007/BF01537606

24. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. *PLoS One*. 2016;11(5):e0155431. doi:10.1371/journal.pone.0155431

25. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(suppl 20):22–33.

26. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th ed. American Psychiatric Association; 2013.

27. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Int Med*. 2013;4(2):627–635.

28. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychol Aging*. 1997;12(2):277–287. doi:10.1037//0882-7974.12.2.277

29. Papassotiropoulos A, Heun R. Screening for depression in the elderly: a study on misclassification by screening instruments and improvement of scale performance. *Prog Neuropsychopharmacol Biol Psychiatry*. 1999;23(3):431–446. doi:10.1016/s0278-5846(99)00007-x

30. Winkelmann ZK, Crossway AK. Optimal screening methods to detect cardiac disorders in athletes: an evidence-based review. *J Athl Train*. 2017;52(12):1168–1170. doi:10.4085/1062-6050-52.11.24

31. Tosteson ANA, Fryback DG, Hammond CS, et al. Consequences of false-positive screening mammograms. *JAMA Intern Med*. 2014;174(6):954–961. doi:10.1001/jamainternmed.2014.981

32. Carleton RN, Thibodeau MA, Teale MJ, et al. The Center for Epidemiologic Studies Depression Scale: a review with a theoretical and empirical examination of item content and factor structure. *PloS One*. 2013;8(3):e58067. doi:10.1371/journal.pone.0058067

33. Harenberg S, Marshall-Prain N, Dorsch KD, Riemer HA. Factorial validity and gender invariance of the Center for Epidemiological Studies Depression in cardiac rehabilitation patients. *J Cardiopulm Rehabil Prev*. 2015;35(5):320–327. doi:10.1097/HCR.0000000000000127

34. Adams LB, Gottfredson N, Lightfoot AF, Corbie-Smith G, Golin C, Powell W. Factor analysis of the CES-D 12 among a community sample of black men. *Am J Mens Health*. 2019;13(2):1557988319834105. doi:10.1177/1557988319834105

35. James SA. John Henryism and the health of African-Americans. *Cult Med Psychiatry*. 1994;18(2):163–182. doi:10.1007/BF01379448

36. López RL, Levy JJ. Student athletes' perceived barriers to and preferences for seeking counseling. *J Coll Couns*. 2013;16(1):19–31. doi:10.1002/j.2161-1882.2013.00024.x

37. Watson JC. College student-athletes' attitudes toward help-seeking behavior and expectations of counseling services. *J Coll Stud Dev*. 2005;46(4):442–449. doi:10.1353/csd.2005.0044

38. Wahto RS, Swift JK, Whipple JL. The role of stigma and referral source in predicting college student-athletes' attitudes toward psychological help-seeking. *J Clin Sport Psychol*. 2016;10(2):85–98. doi:10.1123/JCSP.2015-0025

39. Mentink JW. *Major Depression in Collegiate Student-Athletes: Case Study Research*. Dissertation. Washington State University; 2001.

40. Harenberg S, Ouellet-Pizer C, Nieto M, et al. Anonymous vs. non-anonymous administration of depression scales in elite athletes: a meta-analysis. *Int Rev Sport Exerc Psychol*. 2022. doi:10.1080/1750984X.2022.2095658

41. Kroshus E, Kubzansky LD, Goldman RE, Austin SB. Norms, athletic identity, and concussion symptom under-reporting among male collegiate ice hockey players: a prospective cohort study. *Ann Behav Med*. 2015;49(1):95–103. doi:10.1007/s12160-014-9636-5

42. Corrigan PW, Watson AC, Barr L. The self-stigma of mental illness: implications for self-esteem and self-efficacy. *J Soc Clin Psychol*. 2006;25(8):875–884. doi:10.1521/jscp.2006.25.8.875

43. Jones AL, Butryn TM, Furst DM, Semerjian TZ. A phenomenological examination of depression in female collegiate athletes. *Athl Insight*. 2013;5(1):1–19.

44. Kamm RL. Interviewing principles for the psychiatrically aware sports medicine physician. *Clin Sports Med*. 2005;24(4):745–769, vii. doi:10.1016/j.csm.2005.06.002

Address correspondence to Lindsey Keenan, PhD, ATC, West Chester University, 700 S High Street, West Chester, PA 19383. Address email to LKeenan@wcupa.edu.