An Analysis of the Sport Mental Health Assessment Tool in Team USA Athletes

By: Travis Anderson, William M. Adams, Jessica D. Bartley, Angel L. Brutus, Amber T. Donaldson, Jonathan T. Finnof

Abstract:

Objective The Sport Mental Health Assessment Tool 1 (SMHAT-1) was introduced as a critical component to the athlete health evaluation. However, the effectiveness of the initial triage step questionnaire (Athlete Psychological Strain Questionnaire (APSQ)) has yet to be analysed within a National Olympic and Paralympic Committee delegation. This study evaluated the ability of the APSQ to identify athletes at risk for mental health concerns.

Methods Athletes completed the APSQ and all subsequent screening questionnaires of the SMHAT-1 as part of their Tokyo and Beijing Olympic and Paralympic Games health history screening. Each questionnaire was scored according to published guidelines, and the false-negative rate (FNR) for the APSQ identifying athletes that were positively screened on the subsequent questionnaires was computed.

Results 1066 athletes from 51 different Olympic and Paralympic and Summer and Winter sports completed the SMHAT-1. The FNRs for all athletes who were positively screened on a subsequent questionnaire with an APSQ score of <17 ranged from 4.8% to 66.7%. The global FNR for being positively screened on any questionnaire was 67.5%. Female, Paralympic and Winter athletes scored higher on one or more questionnaires compared with male, Olympic and Summer athletes, respectively (p<0.05).

Conclusion Due to the high FNR of the APSQ detecting a potential mental health concern, we recommend athletes complete the APSQ and all subsequent questionnaires of the SMHAT-1 rather than using only the APSQ as an initial screening test.

Keywords: Sport Mental Health Assessment Tool | Athlete Psychological Strain Questionnaire | Olympic athletes | SMHAT-1 | APSQ

Article:

Introduction

The Sport Mental Health Assessment Tool (SMHAT-1) was developed by the International Olympic Committee in 2021 to screen for mental health concerns in athletes.1 The SMHAT-1 consists of a multistep process; step 1 consists of a short questionnaire to triage the athletes; step 2 employs additional questionnaires to screen for specific mental health concerns; and step 3 consists of follow-up and formal evaluations by clinicians who then provide recommendations for additional services when appropriate. In order to first triage athletes, the 10-item Athlete Psychological Strain Questionnaire (APSQ)2 is used, and athletes scoring 17 or greater are assessed further by a series of validated questionnaires: the General Anxiety Disorder-73 (GAD-7, assesses for the presence of anxiety symptoms (reliability: internal consistency=0.92, test–retest reliability=0.83, and validity: sensitivity=89% and specificity=82%)1 3); the Patient Health Questionnaire-94 (PHQ-9, assesses for the presence of depression symptoms (reliability: internal consistency=0.83–0.89, and validity: sensitivity=88% and specificity=88%)1 4); the Athlete Sleep Screening Questionnaire5 (ASSQ, assesses for the presence of potential sleep disturbances (reliability: internal consistency=0.74, test–retest reliability=0.86, and validity: sensitivity=81% and specificity=93%)1 6); the Alcohol Use Disorders Identification Test Consumption7 (AUDIT-C, assesses for the presence of alcohol misuse (reliability: test–retest reliability=0.6–0.9 and validity: area under the curve (AUC) receiver operating characteristic (ROC)=0.85–0.94)1 8; a SMHAT1 adapted version of Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye Openings Adapted to Include Drugs9 (CAGE-AID, assesses for the presence of drug misuse (reliability≥0.9 and validity: sensitivity=79% and specificity=97%)1); and and Brief Eating Disorders in Athletes Questionnaire10 (BEDA-Q (reliability: internal consistency=0.81, and validity: sensitivity=82.1% and specificity=84.6%)1 10). Each validated questionnaire has its own threshold (and sex-specific thresholds, in the case of AUDIT-C), above which a clinical referral and assessment are recommended as the third and final step of the SMHAT-1. In the event of an APSQ score lower than 17, no further assessment is recommended.1

WHAT IS ALREADY KNOWN ON THIS TOPIC
• The Sport Mental Health Assessment Tool 1 (SMHAT-1) is used to screen athletes for potential mental health concerns and is constructed from validated questionnaires.
WHAT THIS STUDY ADDS
• This is the first study to test the effectiveness of the triage step of the SMHAT-1 and to report the false-negative rates (FNRs) for subsequent questionnaires completed after athlete triage. The overall FNR of the Athlete Psychological Strain Questionnaire (APSQ) was 67.5%.
HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY
• This study suggests that the APSQ should not be used as a stand-alone initial triage step in the use of the SMHAT-1 in athletes; rather, the combination of steps 1 and 2 of the SMHAT-1 as the initial triage may best serve our athlete populations. If only the APSQ is used, we believe adding a screening question regarding self-harm and suicide should be strongly considered.

The APSQ was validated against the Kessler Psychological Distress Survey (K10),2 which, in turn, was previously validated against clinical assessments of mental health disorders (eg, anxiety or mood disorders) as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.11 The score of ≥17 was aligned with the K10 'high' category, despite the 'very high' category showing a greater AUC of the ROC curve.12 The choice to use a lower

threshold is undoubtedly the correct decision, but this background illustrates that the APSQ was not initially created or validated for triaging athletes who may be at risk of anxiety, depression, sleep, disordered eating or eating disorders, or substance misuse or abuse.

Furthermore, the APSQ was validated against the K10 only (1) in athletes in Australia; (2) from three sports (Australian football, cricket and soccer), all of which were summer sports and only one of which was an Olympic sport; (3) with a disproportionately male sample (92.3% male); and (4) with athletes aged >18 years.[12] Despite these potential limitations, the APSQ is a short-form questionnaire with high internal consistency for male (Cronbach α=0.87) and female (Cronbach α=0.84) athletes,[12] an acceptable discrimination for a number of psychological conditions and developed specifically for use in athletes, and was therefore chosen as the triage tool for the SMHAT-1. Since the SMHAT-1 is deployed internationally, across Summer and Winter sports, and with some athletes younger than 18 years of age, it is critical to test and validate the APSQ in different athlete populations.

Given these limitations and the associated concerns, this study aimed to evaluate the APSQ false-negative rate (FNR) for identifying which athletes would have a positive screen on subsequent SMHAT-1 questionnaires. Further, we aimed to validate the APSQ's ability to correctly identify at-risk athletes across Olympic and Paralympic, Summer and Winter sports in elite male and female athletes (ie, examine the sensitivity and specificity of the triage step) and to evaluate alternative options for conducting the SMHAT-1.

Methods

Study design and setting

A retrospective analysis of SMHAT-1 responses of Team USA athletes, including those competing in the Tokyo 2020 Olympic and Paralympic Games and the Beijing 2022 Olympic and Paralympic Games, was conducted. SMHAT-1 responses collected from 1 January 2021 to 19 July 2022 were included in this analysis. Patients were not directly involved in the design, conduct, reporting or dissemination of this research, but our research priority remains to improve the care we provide to the athletes included and not included in this work, including those athletes at other National Olympic and Paralympic Committees.

Equity, diversity and inclusion statement

No specific efforts were made to recruit participants explicitly based on participant diversity. However, this study was conducted using Team USA athletes, a diverse population of elite athletes (sex: male=48.6% and female=51.4%; self-identified ethnic origin: Asian=2.44%, black=17.4%, white=62.8%, other=1.13%, two or more races=6.10%, declined to respond=9.94%; self-identified ethnicity: of Hispanic/Latino/Spanish origin=5.44%, not of Hispanic/Latino/Spanish origin=84.6%, declined to respond=9.94%; games: Olympic=90.9% and Paralympic=9.10%; season: Summer=66.5% and Winter=33.5%). The research team comprises a diverse, balanced group of expert clinicians and researchers (50% female). Because all athletes were required to complete the health history questionnaire before the games, efforts were made to ensure that they had access to the resources they required regardless of regional geographical differences, education or socioeconomic levels.

Data collection and measures

The SMHAT-1 was deployed via an online survey platform (Qualtrics, Provo, Utah, USA) and as part of a comprehensive athlete health history questionnaire for all Team USA athletes, including athletes who competed at the Tokyo 2020 and Beijing 2022 Olympic and Paralympic Games and athletes competing in associated national and international competitions. Unlike the SMHAT-1 implementation methods recommended by the International Olympic Committee,[1] athletes completed the APSQ and all subsequent questionnaires regardless of their score on the APSQ. This was an a priori decision made by the Sports Medicine team ahead of the Tokyo 2020 Games, given the novelty of the SMHAT-1's use among elite athletes.

SMHAT-1 data were exported, and each questionnaire was scored using published guidelines.[1] Readers are directed to Gouttebarge et al and the accompanying online supplemental material for additional details regarding the internal consistency, test–retest reliability, and specificity and sensitivity of the subsequent questionnaires. In this analysis, as shown in table 1, scores exceeding the thresholds were provided a clinical follow-up by a mental health provider as part of the Team USA Sports Medicine Program within 48 hours of SMHAT-1 completion. In addition, Question 9 of the Patient Health Questionnaire-9 (PHQ-9 Q9, assesses for the presence of suicidal thoughts) was scored and athletes that screened positive were contacted by a mental health provider within 20 min of submitting the health history questionnaire and were provided with appropriate care.

Statistical analysis

Descriptive statistics were computed and differences in each questionnaire between male and female (sex), Olympic and Paralympic (games), and Summer and Winter (season) athletes were tested by general linear models. Next, confusion matrices were constructed for all athletes, as well as by sex, games and season. The FNR (percentage of athletes scoring <17 on the APSQ but >threshold on one or more subsequent questionnaires) was computed for each confusion matrix. Next, logistic regression models were computed using a binary predictor variable ($17 > APSQ \geq 17$) and a binary outcome (specific threshold >subsequent questionnaire ≥specific threshold) for all athletes and then separately for each level of sex, games and season. While multiple observations for some athletes were present in the data, the degree of dependence did not warrant mixed-effects generalised linear models as the fit of these models was singular, suggesting little variance was accounted for by this strict lack of independence. We, therefore, opted to continue using simple logistic regressions. The AUC of the ROC curves was subsequently computed for each model. Finally, we computed separate logistic regression models using cut points from 10 to 50 and extracted the APSQ threshold that maximised the AUC of the ROC for each subsequent questionnaire. All analyses were completed using R Statistical Software,[13] and the alpha level was set at a p value of <0.05 for all inferential statistics. Additionally, we calculated post hoc power analyses for our logistic and linear regression models to provide additional context to the results presented here.

Table 1 Summary of Sport Mental Health Assessment Tool 1 questionnaires by sex, games and season

Questionnaire	Theoretical score range	Threshold	All athletes	Sex		Games		Season	
				Male	Female	Olympic	Paralympic	Summer	Winter
APSQ	10–50	≥17	14.1±5.3 (10–35)	13.3±4.7 (10–34)	14.9±5.2 (10–35)*	13.9±4.9 (10–34)	16.2±6.2 (10–35)	14.5±5.3 (10–35)	13.3±4.3 (10–30)
GAD-7	0–21	≥10	2.1±3.13 (0–18)	1.5±2.6 (0–18)	2.7±3.4 (0–17)*	2.0±2.9 (0–18)	3.3±3.9 (0–16)	2.2±3.2 (0–18)	1.8±2.8 (0–15)*
PHQ-9	0–27	≥10	1.7±2.7 (0–18)	1.2±2.2 (0–18)	2.1±3.0 (0–18)*	1.6±2.6 (0–18)	2.5±3.3 (0–15)	1.9±2.9 (0–18)	1.2±2.0 (0–12)*
PHQ-9 Q9	0–3	≥1	0.01±0.1 (0–1)	0.01±0.1 (0–1)	0.02±0.1 (0–1)	0.01±0.1 (0–1)	0.03±0.2 (0–1)	0.02±0.1 (0–1)	0±0 (0–0)*
ASSQ	0–17	≥8	4.9±2.8 (0–16)	4.8±2.8 (0–15)	5.1±2.8 (0–16)	4.9±2.8 (0–16)	5.4±2.5 (0–13)	5.1±2.9 (0–16)	4.6±2.6 (0 to 15)*
AUDIT-C	0–13	Male: ≥4 Female: ≥3	2.4±1.7 (0–11)	2.6±1.8 (0–11)	2.2±1.6 (0–8)*	2.4±1.7 (0–11)	2.7±1.7 (1–9)	2.2±1.7 (0–11)	2.8±1.7 (0–9)*
CAGE-AID	0–4	≥2	0.05±0.3 (0–4)	0.07±0.4 (0–4)	0.04±0.3 (0–4)	0.05±0.4 (0–4)	0.04±0.2 (0–1)	0.05±0.3 (0–4)	0.06±0.4 (0–4)
BEDA-Q	0–34	≥4	9.8±6.3 (0–31)	8.4±5.7 (0–24)	11.0±6.5 (0–31)*	9.7±6.3 (0–31)	10.5±6.3 (0–29)	10.4±6.3 (0–31)	8.6±6.0 (0–28)*

*Significantly different from male athletes, Olympic athletes and summer athletes for sex, games and season, respectively (p<0.05).

APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye openings Adapted to Include Drugs; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; PHQ-9 Q9, Question 9 of the Patient Health Questionnaire-9.

Results

One thousand sixty-six athletes completed the APSQ and all of the subsequent questionnaires. The sample consisted of 518 men (Summer Olympics, n=322; Winter Olympics, n=140; Summer Paralympics, n=6; and Winter Paralympics, n=50) and 548 women (Summer Olympics, n=355; Winter Olympics, n=152; Summer Paralympics, n=26; and Winter Paralympics, n=15) from 51 sports (Summer Olympics, n=28; Winter Olympics, n=9; Summer Paralympics, n=8; and Winter Paralympics, n=6). The greatest number of athletes were positively screened on the AUDIT-C and the least on the CAGE-AID (figure 1).
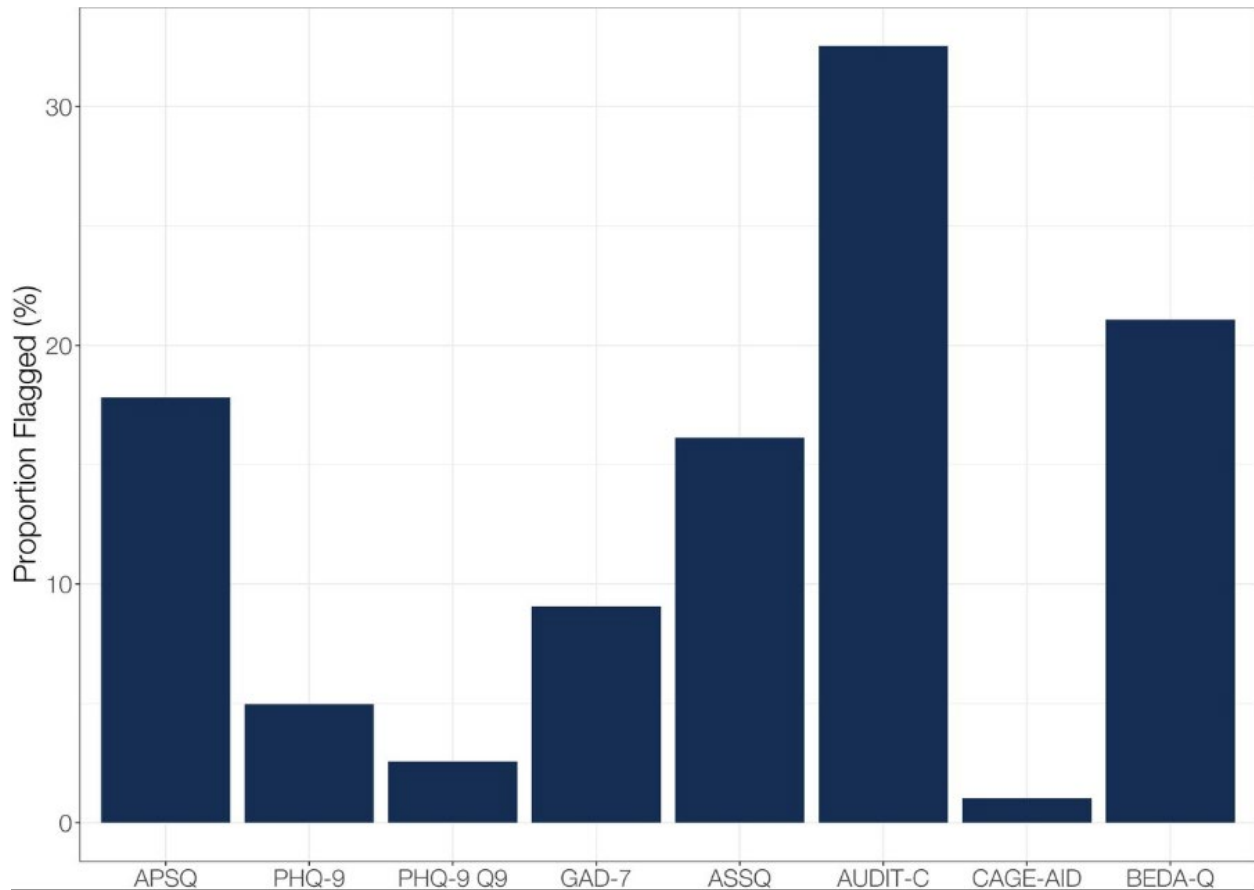

Figure 1 Proportion of athletes with positive screen on APSQ and all subsequent questionnaires. APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye Openings Adapted to Include Drugs; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9.

Comparison by sex, season and games

When considering all athletes, APSQ scores ranged from 10 to 35. Female athletes had greater APSQ ($p<0.001$), GAD-7 ($p<0.001$), PHQ-9 ($p<0.001$) and BEDA-Q ($p<0.001$) scores and lower AUDIT-C ($p<0.001$) scores compared with male athletes. Paralympic athletes did not statistically differ from Olympic athletes on any scale. Winter athletes did not statistically differ from Summer athletes in APSQ score ($p=0.432$) but scored significantly lower on the GAD-7 ($p<0.001$), PHQ-9

(p<0.001), PHQ-9 Q9 (p<0.001), ASSQ (p=0.008) and BEDA-Q (p<0.001), but significantly higher on AUDIT-C (p<0.001) compared with Summer athletes (table 1 and figure 2).
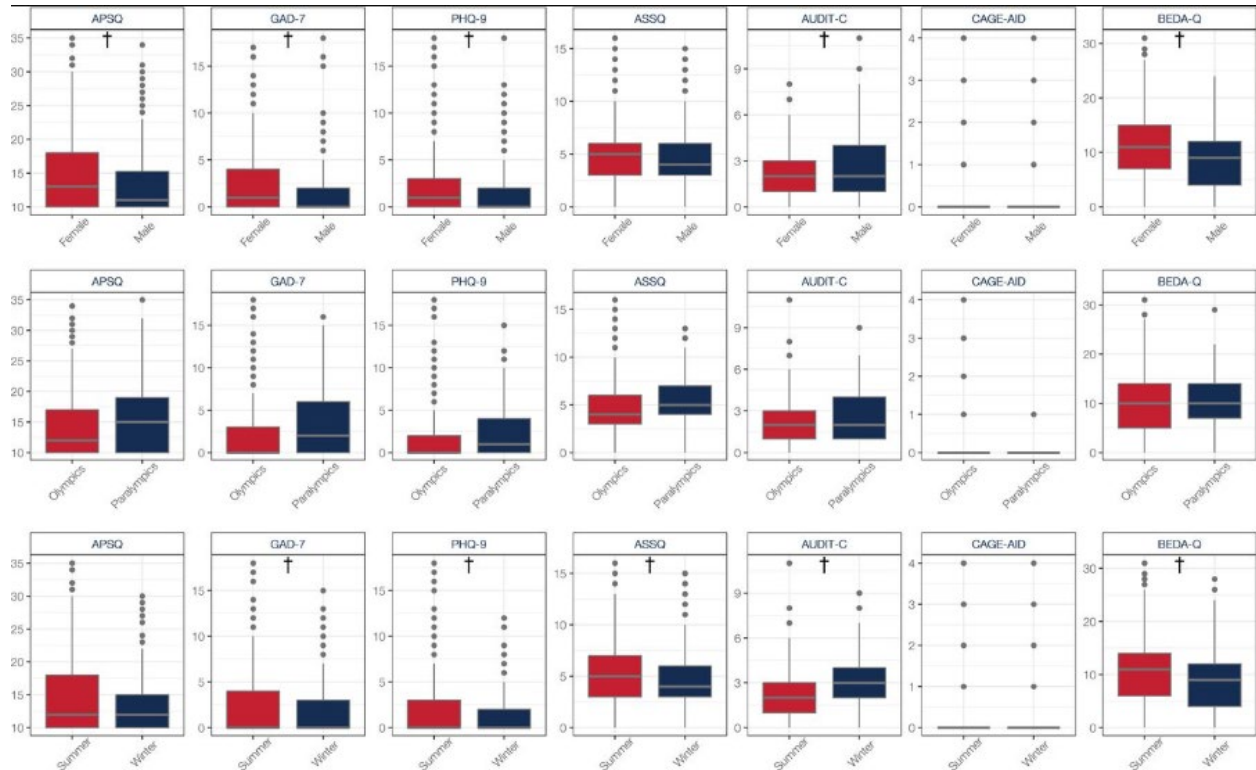

Figure 2 Questionnaire responses by sex, games, and season. APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling, and Eye-openings Adapted to Include Drugs; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9.

False-negative rates

The FNR for athletes who were positively screened on a subsequent questionnaire with an APSQ score of <17 was between 4.8% (PHQ-9) and 66.7% (BEDA-Q) (table 2). When considering athletes who did not positively screen on the APSQ but did so on any subsequent questionnaire, the global FNR of the APSQ was 67.4%. As shown in table 2, women had a lower FNR than men for all subsequent questionnaires, except for the PHQ-9, wherein men had an FNR of 0%. Similarly, Paralympic athletes had a lower FNR than Olympic athletes for all subsequent questionnaires and multiple instances of an FNR of 0%. Regarding PHQ-9 Q9 for all athletes, we observed an FNR of 6.7%.

Specificity and sensitivity

The greatest and lowest AUC of the ROC curve for any subcategory of any subsequent questionnaire using a threshold of APSQ score of ≥17 was 0.84 (PHQ-9 and GAD-7) and 0.34 (CAGE-AID) for All/Olympic athletes and Paralympic athletes, respectively (table 3). The average AUC of the ROC curve for the GAD-7 and PHQ-9 using a threshold of APSQ score of ≥17 was

Table 2 Confusion matrices and FNRs by sex, games and season

		Sex									Games						Season					
		All athletes			Male			Female			Olympic			Paralympic			Summer			Winter		
		APSQ score			APSQ score			APSQ score			APSQ score			APSQ score			APSQ score			APSQ score		
		<17	≥17	FNR (%)	<17	≥17	FNR (%)	<17	≥17	FNR (%)	<17	≥17	FNR (%)	<17	≥17	FNR (%)	<17	≥17	FNR (%)	<17	≥17	FNR (%)
GAD-7	<10	750	276	5	403	109	16.7	347	167	2.9	698	240	6.5	52	36	0	464	214	6.5	286	62	0
	≥10	2	38		1	5		1	33		2	29		0	9		2	29		0	9	
PHQ-9	<10	751	294	4.8	404	109	0	347	185	6.3	699	254	6.3	52	40	0	466	225	0	285	69	33.3
	≥10	1	20		0	5		1	15		1	15		0	5		0	18		1	2	
PHQ-9 Q9	<1	745	299	6.7	403	109	16.7	348	191	0	699	258	8.3	52	42	0	465	229	6.7	286	71	0
	≥1	1	14		1	5		0	9		1	11		0	3		1	14		0	0	
ASSQ	<10	685	209	39	364	72	48.8	321	137	30	638	176	40	47	33	29.4	418	163	25.9	267	46	43.2
	≥10	67	105		40	42		27	63		62	93		5	12		28	80		19	25	
AUDIT-C	<10	545	174	59.7	301	63	66.9	244	111	53.9	508	146	61	37	28	46.9	357	147	53.2	188	27	69
	≥10	207	140		103	51		104	89		192	123		15	17		109	96		98	44	
CAGE-AID	<10	147	308	45.5	400	111	57.1	347	197	25	695	263	45.5	52	45	0	463	241	60	284	67	33.3
	≥10	5	6		4	3		1	3		5	6		0	0		3	2		2	4	
BEDA-Q	<10	866	23	66.7	441	12	74.6	425	11	59.9	698	21	68.6	168	2	48.8	768	20	62.5	98	3	75.6
	≥10	582	291		300	102		282	189		541	248		41	43		371	223		211	68	

APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye Openings Adapted to Include Drugs; FNR, false-negative rate; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; PHQ-9 Q9, Question 9 of the Patient Health Questionnaire-9.

Table 3 Odds of being positively screened on a subsequent questionnaire if APSQ score of ≥17 by sex, games and season

| | Sex | | | | | | Games | | | | | | Season | | | | | |
| | All athletes | | Male | | Female | | Olympic | | Paralympic | | Summer | | Winter | |
	Odds	AUC	Odds	AUC	Odds	AUC	Odds	AUC	Odds	AUC	Odds	AUC	Odds	AUC
GAD-7	51.6*	0.84	18.5*	0.81	68.6*	0.82	42.2*	0.84	–	–	31.4*	0.81	–	–
PHQ-9	51.0*	0.84	–	–	28.1*	0.80	41.3*	0.84	–	–	–	–	8.3	0.74
PHQ-9 Q9	35.0*	0.82	18.5*	0.81	–	–	29.8*	0.82	–	–	28.4*	0.80	–	–
ASSQ	5.1*	0.69	5.3*	0.67	5.5*	0.70	5.4*	0.69	3.4*	0.65	4.3*	0.67	7.6*	0.71
AUDIT-C	2.1*	0.42	2.4*	0.42	1.9*	0.43	2.2*	0.42	1.5	0.55	2.1*	0.41	3.1*	0.41
CAGE-AID	2.9	0.63	2.7	0.34	5.3	0.69	3.2	0.64	–	–	1.3	0.47	8.5*	0.74
BEDA-Q	18.8*	0.65	12.5*	0.61	25.9*	0.69	15.2*	0.64	88.1*	0.75	23.1*	0.68	10.5*	0.61

Note: Odds are exponentiated log odds from generalised linear model (GLM). Odds and AUC are not reported for instances wherein a threshold of APSQ score of >17 created perfect prediction (see table 2).

*Significant predictor (p<0.05).

APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUC, area under the curve; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye Openings Adapted to Include Drugs; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; PHQ-9 Q9, Question 9 of the Patient Health Questionnaire-9.

0.82 and 0.81, respectively. An APSQ score of ≥17 threshold was a significant predictor of all subsequent questionnaires except CAGE-AID for all athletes (p=0.08), including for men (p=0.197) and women (p=0.151) when analysed separately. An APSQ score of ≥17 threshold was also not predictive of a positively screen score on CAGE-AID for Olympic athletes (p=0.058) or Summer athletes (p=0.787) when analysed separately, nor of positively screened scores on the AUDIT-C (p=0.352) or PHQ-9 (p=0.087) for Paralympic athletes and Winter athletes, respectively. All other models demonstrated a significant association between an APSQ score of ≥17 threshold and binary outcome on the subsequent questionnaire (there were insufficient data to test the association between the APSQ and having a positive screen on multiple subsequent questionnaires; see table 3).

Recalculating the APSQ threshold

The lowest optimised threshold for all athletes for the APSQ to predict a positively screened score for one of the subsequent questionnaires was determined to be ≥11 (table 4).

False-positive rates

Owing to the relatively high FNR observed, post hoc analyses were completed to evaluate the effect of removing the APSQ from the SMHAT-1 entirely and completing only the subsequent questionnaires. Specifically, we calculated the number of athletes who would exceed the APSQ score of ≥17 but would not exceed the threshold on any subsequent questionnaires (and thus, according to the current SMHAT-1 framework, be recommended for additional monitoring). As a result, only 7 of 1066 athletes (0.7%) were positively screened on the APSQ but no other subsequent questionnaire. However, because of the high number of athletes positively screened on the BEDA-Q, we also assessed how many athletes would be missed for additional monitoring if they took only the GAD-7, PHQ-9, ASSQ, AUDIT-C and CAGE-AID. In this case, 96 athletes would have been positively screened by the APSQ but no other subsequent questionnaire.

Post hoc power analyses

We calculated post hoc power analyses using our observed effect sizes; a full accounting of these analyses are available in online supplemental material 1. For our logistic regression models, we demonstrated power (1−β) of 0.41–1.00 for all athletes; 0.05–1.00 and 0.05–1.00 for male and female athletes, respectively; 0.46–1.00 and 0.05–1.00 for Olympic and Paralympic athletes, respectively; and 0.05–1.00 and 0.05–1.00 for Summer and Winter athletes, respectively. These post hoc power calculations include all models, regardless of whether a significant effect of sex, games or season was observed. With regard to our linear models comparing SMHAT-1 scores by sex, games and season (table 1), we found that only ASSQ by season (power=0.74) demonstrated a statistically significant difference when the post hoc power calculation suggested a power of <0.80.

Table 4 Optimised thresholds for APSQ in predicting a positively screen on a subsequent questionnaire by sex, games and season

	Sex						Games				Season			
	All athletes		Male		Female		Olympic		Paralympic		Summer		Winter	
	Threshold	AUC	Threshold	AUC	Threshold	AUC	Threshold	AUC	Threshold	AUC	Threshold	AUC	Threshold	AUC
GAD-7	51.6*	0.84	18.5*	0.81	68.6*	0.82	42.2*	0.84	–	–	31.4*	0.81	–	–
PHQ-9	51.0*	0.84	–	–	28.1*	0.80	41.3*	0.84	–	–	–	–	8.3	0.74
PHQ-9 Q9	35.0*	0.82	18.5*	0.81	–	–	29.8*	0.82	–	–	28.4*	0.80	–	–
ASSQ	5.1*	0.69	5.3*	0.67	5.5*	0.70	5.4*	0.69	3.4*	0.65	4.3*	0.67	7.6*	0.71
AUDIT-C	2.1*	0.42	2.4*	0.42	1.9*	0.43	2.2*	0.42	1.5	0.55	2.1*	0.41	3.1*	0.41
CAGE-AID	2.9	0.63	2.7	0.34	5.3	0.69	3.2	0.64	–	–	1.3	0.47	8.5*	0.74
BEDA-Q	18.8*	0.65	12.5*	0.61	25.9*	0.69	15.2*	0.64	88.1*	0.75	23.1*	0.68	10.5*	0.61

Note: AUC of ROC curve for BEDA-Q when modelling Olympic athletes only was equivalent for a threshold of ≥11 and ≥13. Models and therefore optimised thresholds could not be computed for some predictions due to the absence of positive cases or where models' best threshold performed worse than an AUC of ROC curve=0.5.

APSQ, Athlete Psychological Strain Questionnaire; ASSQ, Athlete Sleep Screening Questionnaire; AUC, area under the curve; AUDIT-C, Alcohol Use Disorders Identification Test Consumption; BEDA-Q, Brief Eating Disorders in Athletes Questionnaire; CAGE-AID, Cutting Down, Annoyance by Criticism, Guilty Feeling and Eye Openings Adapted to Include Drugs; GAD-7, General Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; PHQ-9 Q9, Question 9 of the Patient Health Questionnaire-9; ROC, receiver operating characteristic.

Discussion

The present study is among the first to analyse the efficacy of the APSQ in predicting outcomes of the subsequent questionnaires within the SMHAT-1, a novel screening tool that has been proposed to identify athletes potentially at risk of mental health concerns. In this analysis of 1066 Olympic and Paralympic athletes, 29.5% of athletes screened positively on the APSQ, slightly more than the 23.2% of athletes previously reported.[12] We further demonstrated that the current recommendations for an APSQ≥17 threshold results in an FNR on subsequent SMHAT-1 screening questionnaires between 4.8% and 66.7%, which tended to be higher than those previously published.[1] However, we also demonstrated that a threshold of ≥17 results in a reasonably high AUC of ROC curves (0.42–0.84) and a significant ability to predict a score that would have resulted in a positive screen on one of the subsequent questionnaires. Thus, for most subsequent questionnaires, an APSQ score of ≥17 results in empirically maximising both the sensitivity and specificity of predictions.

Additionally, we demonstrated clinically meaningful FNRs for the APSQ, greater than those previously reported, particularly for the PHQ-9 and GAD-7.[1] Recently, Mountjoy and colleagues evaluated Canadian collegiate athletes and demonstrated an alarmingly high presence of symptoms associated with anxiety (30%), depression (26%), sleep disturbance (39%), alcohol misuse (55%), drug use (10%) and disordered eating (83%)[14]; however, given the FNR in our research, the work of Mountjoy et al may in fact be underestimating the presence of potential mental health concerns and thus the number of athletes that require additional follow-up.

In our sample, female athletes tended to score higher on the APSQ, GAD-7 and PHQ-9 compared with male athletes, consistent with previous findings.[12] [14–17] Some preliminary data also suggest that Paralympic and Winter sport athletes experience a greater degree of psychological distress,[17] which supports our observation of greater scores on the APSQ and other subscales compared with the Olympic and Summer sport counterparts. However, the underlying reasons for these statistically significant differences and their practical clinical importance require further investigation. It has, however, been previously noted that Paralympic athletes likely face different stressors than Olympic athletes[18]; our data further confirm the recommendation from Leyland et al[17] that more qualitative research is needed to inform best mental health practices across all populations.

Acceptable FNRs

Most critically, a low FNR, especially in the event of just a single observation missed (as in the case of PHQ-9 Q9), would be ideal in most academic and theoretical contexts. However, it is crucial to recognise that the acceptable FNR is dependent and inversely proportional to the predicted outcome's importance and severity. As with other physical and mental health concerns, an FNR as close to 0 as possible is highly desirable, and the more critical and severe the predicted outcome, the more important this becomes. In the context of mental health, this would most obviously apply to PHQ-9 Q9, as we know that suicidal ideation can be as high as 17.4% in some populations,[19] and suicide contributes substantially to all-cause mortality in elite athletes.[20] Previous work has demonstrated that including suicide screening does not increase the likelihood of those ideations.[21] Conversely, incorrectly triaging an athlete at risk of self-harm and, thus, neglecting to perform a formal clinical evaluation and relevant implementation of proper treatment plans for the athlete could potentially have catastrophic consequences.

Clinical implications and recommended changes to the SMHAT-1

The SMHAT-1 is in its infancy, having been devised and deployed for use in athletics settings for less than 2 years (released in 2021). Therefore, it is expected that this tool will be modified as additional data are gathered to inform proper clinical decisions. Indeed, we believe the SMHAT-1 is a critical component of our health history screening for all Olympic and Paralympic athletes and will continue to champion its use. However, given the wide range of FNRs (4.8%–66.7%) demonstrated here, despite the otherwise acceptable performance of the APSQ score of ≥17 threshold in a purely quantitative sense, we suggest that the APSQ should not be used as a stand-alone initial triage step in the use of the SMHAT-1 in athletes; rather, the combination of steps 1 and 2 of the SMHAT-1 as an initial triage step prior to referral of athletes to a mental health provider may best serve our athlete populations. In this case, athletes could be positively screened on subsequent questionnaires—including, critically, the PHQ-9 Q9—for clinical assessment and care without needing to be positively screened on the APSQ. Additionally, those athletes who would positively screen on the APSQ but not on subsequent questionnaires (ie, in our sample, when not including the BEDA-Q, n=95) and require follow-up and additional monitoring would not be missed. Alternatively, it may be that the time requirements for athletes or the medical resources required for qualified professional clinicians to review and efficiently contact athletes would be too great. In this case, we recommend that at least the APSQ be adapted for specific use as a triage tool in the SMHAT-1 by including an 11th question on suicide risk (eg, the inclusion of PHQ-9 Q9). In this case, the question with the most significant consequence is included as part of the triage.

Limitations

Despite the large sample size and robust analysis and subanalyses of the SMHAT-1 in our current study, several limitations should be acknowledged. First, these data were collected after the start of and during the SARS-CoV-2 global pandemic, at which time the Tokyo 2020 Games were delayed, resulting in many athletes having to adjust to alternative training and competition schedules. The Beijing 2022 Games also introduced additional stressors for athletes that may not be present during a non-Games year, both related and unrelated to pandemic adjustments. These concerns likely influenced, to some degree, the prevalence of positive SMHAT-1 screens observed here. Moreover, these data, much like the original APSQ-validation studies, consist of athletes from a single country. It is acknowledged that this dataset is more diverse regarding sex, sport and seasons. However, additional data from other countries, especially countries with varying mental health educational systems, mental healthcare infrastructure and global sociocultural differences, should be analysed to confirm our results. Relatedly, only a small number of athletes were positively screened on subsequent questionnaires, especially on the PHQ-9 Q9. This small number of positive cases may produce biased estimates and unreliable (high or low) FNRs (see PHQ-9 Q9 in Winter athletes, table 2).

We also acknowledge that more specific analyses of Paralympic athletes are required, accounting for category and degree of impairments. Regrettably, we do not have the requisite sample size to conduct this analysis, but we strongly encourage this in future SMHAT-1 studies. Although we included all available athletes in these analyses, we also acknowledge that the sample size may be underpowered, particularly for some of our subgroup analyses. To investigate this further, we calculated post hoc power analyses and reported those values in online supplemental

material 1. However, we also recognise and acknowledge that post hoc power analyses have substantial statistical issues22–24 and thus cannot necessarily be relied on to confirm adequate sample size for all analyses and comparisons. For example, the statistically significant differences in ASSQ score by season had a post hoc power ($1-\beta$) of <0.80, and thus these differences should therefore be interpreted with caution, although the true effect size in the population may not be represented by our observed effect size, and thus these power analyses may be misleading. We therefore further encourage other National Olympic and Paralympic Committees to continue collaborating and aggregating data to better understand how the SMHAT-1 can be improved for the sake of athletes' mental health and well-being.

Conclusions

In conclusion, based on our findings, we recommend having athletes complete the APSQ and all subsequent questionnaires of the SMHAT rather than using the APSQ as a screening test. If only the APSQ is used as a screening test in time-limited or resource-limited settings, we recommend adding PHQ-9 Q9 to the APSQ to ensure those at risk to themselves are identified. These results are presented to enhance the effectiveness of the SMHAT-1, and we continue to advocate for the inclusion of the SMHAT-1 for mental health screening in elite athletes.

Data availability statement

No data are available.

Ethics statements
Patient consent for publication

Not applicable.

Ethics approval

This study obtained expedited approval by the institutional review board at the University of North Carolina at Greensboro (IRB-FY22-116).

Acknowledgments

References

Gouttebarge V, Bindra A, Blauwet C, et al. International Olympic Committee (IOC) sport mental health assessment tool 1 (SMHAT-1) and sport mental health recognition tool 1 (SMHRT-1): towards better support of athletes' mental health. Br J Sports Med 2021;55:30–7. doi:10.1136/bjsports-2020-102411

Rice SM, Parker AG, Mawren D, et al. Preliminary psychometric validation of a brief screening tool for athlete mental health among male elite athletes: the athlete psychological strain questionnaire. Int J Sport Exerc Psychol 2020;18:850–65. doi:10.1080/1612197X.2019.1611900

Spitzer RL, Kroenke K, Williams JBW, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092–7. doi:10.1001/archinte.166.10.1092

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13. doi:10.1046/j.1525-1497.2001.016009606.x

Samuels C, James L, Lawson D, et al. The athlete sleep screening questionnaire: a new tool for assessing and managing sleep in elite athletes. Br J Sports Med 2016;50:418–22. doi:10.1136/bjsports-2014-094332

Bender AM, Lawson D, Werthner P, et al. The clinical validation of the athlete sleep screening questionnaire: an instrument to identify athletes that need further sleep assessment. Sports Med Open 2018;4:23. doi:10.1186/s40798-018-0140-5

Bush K, Kivlahan DR, McDonell MB, et al. Ambulatory care quality improvement project (ACQUIP. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Arch Intern Med 1998;158:1789–95. doi:10.1001/archinte.158.16.1789

Bradley KA, DeBenedetti AF, Volk RJ, et al. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 2007;31:1208–17. doi:10.1111/j.1530-0277.2007.00403.x

Couwenbergh C, Van Der Gaag RJ, Koeter M, et al. Screening for substance abuse among adolescents validity of the CAGE-AID in youth mental health care. Subst Use Misuse 2009;44:823–34. doi:10.1080/10826080802484264

Martinsen M, Holme I, Pensgaard AM, et al. The development of the brief eating disorder in athletes questionnaire. Med Sci Sports Exerc 2014;46:1666–75. doi:10.1249/MSS.0000000000000276

Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32:959–76. doi:10.1017/s0033291702006074

Rice S, Olive L, Gouttebarge V, et al. Mental health screening: severity and cut-off point sensitivity of the athlete psychological strain questionnaire in male and female elite athletes. BMJ Open Sport Exerc Med 2020;6:e000712. doi:10.1136/bmjsem-2019-000712

R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria, 2013. Available: httpwww R-Proj Org

Mountjoy M, Edwards C, Cheung CP, et al. Implementation of the international olympic committee sport mental health assessment tool 1: screening for mental health symptoms in a Canadian multisport university program. Clin J Sport Med 2023;33:5–12. doi:10.1097/JSM.0000000000001077

Åkesdotter C, Kenttä G, Eloranta S, et al. The prevalence of mental health problems in elite athletes. J Sci Med Sport 2020;23:329–35. doi:10.1016/j.jsams.2019.10.022

Löwe B, Decker O, Müller S, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46:266–74. doi:10.1097/MLR.0b013e318160d093

Leyland S, Currie A, Board E, et al. A survey of the mental health of UK olympic and paralympic sport athletes. J Athlete Dev Exp 2022;4. doi:10.25035/jade.04.02.03

Swartz L, Hunt X, Bantjes J, et al. Mental health symptoms and disorders in paralympic athletes: a narrative review. Br J Sports Med 2019;53:737–40. doi:10.1136/bjsports-2019-100731

Timpka T, Spreco A, Dahlstrom O, et al. Suicidal thoughts (Ideation) among elite athletics (track and field) athletes: associations with sports participation, psychological resourcefulness and having been a victim of sexual and/or physical abuse. Br J Sports Med 2021;55:198–205. doi:10.1136/bjsports-2019-101386

Rao AL, Asif IM, Drezner JA, et al. Suicide in National Collegiate Athletic Association (NCAA) athletes: a 9-year analysis of the NCAA resolutions database. Sports Health 2015;7:452–7. doi:10.1177/1941738115587675

Crawford MJ, Thana L, Methuen C, et al. Impact of screening for risk of suicide: randomised controlled trial. Br J Psychiatry 2011;198:379–84. doi:10.1192/bjp.bp.110.083592

Lakens D. Sample size justification. Collabra 2022;8. doi:10.1525/collabra.33267

Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med 1994;121:200. doi:10.7326/0003-4819-121-3-199408010-00008

Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. Am Stat 2001;55:19–24. doi:10.1198/000313001300339897

Footnotes

Twitter @travaldinho

Contributors TA was responsible for data analysis, data visualisation and data interpretation and was the lead author. WMA, JDB, ALB, ATD and JTF were responsible for the study design, data collection and data interpretation. All authors were responsible for the critical review and editing of the manuscript. JDB and ALB provided domain expertise throughout the project. WMA is the guarantor.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Provenance and peer review Not commissioned; externally peer reviewed.