

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/287958571>

Clinician-friendly physical performance tests in athletes part 3: A systematic review of measurement properties and correlations to injury for tests in the upper extremity

Article in *British Journal of Sports Medicine* · December 2015

DOI: 10.1136/bjsports-2015-095198

CITATIONS

3

READS

1,086

4 authors, including:



Daniel Tarara

High Point University

10 PUBLICATIONS 72 CITATIONS

[SEE PROFILE](#)



Jeffrey Taylor

High Point University

78 PUBLICATIONS 329 CITATIONS

[SEE PROFILE](#)



Eric J Hegedus

High Point University

90 PUBLICATIONS 1,813 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Physical Performance Measures [View project](#)



Expanded Practice Physical Therapy Safety Event Rates: A Single-Center Observational Study [View project](#)

Clinician-friendly physical performance tests in athletes part 3: a systematic review of measurement properties and correlations to injury for tests in the upper extremity

Daniel T Tarara,¹ Lucas K Fogaca,² Jeffrey B Taylor,³ Eric J Hegedus³

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2015-095198>).

¹Department of Exercise Science, High Point University, School of Health Sciences, High Point, North Carolina, USA

²Department of Biology, High Point University, High Point, North Carolina, USA

³Department of Physical Therapy, High Point University, School of Health Sciences, High Point, North Carolina, USA

Correspondence to

Daniel T Tarara, Department of Exercise Science, High Point University, School of Health Sciences, 833 Montlieu Avenue, High Point 27268, NC, USA; dtarara@highpoint.edu

Accepted 25 November 2015

ABSTRACT

Objective In parts 1 and 2 of this systematic review, the methodological quality as well as the quality of the measurement properties of physical performance tests (PPTs) of the lower extremity in athletes was assessed. In this study, part 3, PPTs of the upper extremity in athletes are examined.

Methods Database and hand searches were conducted to identify primary literature addressing the use of upper extremity PPTs in athletes. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed and the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist was used to critique the methodological quality of each paper. The Terwee Scale was used to analyse the quality of the measurement properties of each test.

Results 11 articles that examined 6 PPTs were identified. The 6 PPTs were: closed kinetic chain upper extremity stability test (CKCUEST), seated shot put (2 hands), unilateral seated shot put, medicine ball throw, modified push-up test and 1-arm hop test. Best evidence synthesis provided moderate positive evidence for the CKCUEST and unilateral seated shot put. Limited positive evidence was available for the medicine ball throw and 1-arm hop test.

Conclusions There are a limited number of upper extremity PPTs used as part of musculoskeletal screening examinations, or as outcome measures in athletic populations. The CKCUEST and unilateral seated shot put are 2 promising PPTs based on moderate evidence. However, the utility of the PPTs in injured populations is unsubstantiated in literature and warrants further investigation.

Numerous healthcare organisations and sports bodies have released consensus statements regarding the importance of the preparticipation screening examination to identify athletes at risk of injury.^{1 2} Specifically, the IOC has stated that screening must be reliable, sensitive, specific, inexpensive, easy to perform and widely available.² Physical performance tests (PPTs) meet this definition and go one step further: they are portable and can be performed in many different environments and contexts.

Despite the obvious benefits of PPTs, little is known about whether these tests are appropriate tools for musculoskeletal injury screening, and what little is known is uninspiring.³ Two previous systematic reviews of PPTs in the lower extremity

found little evidence to support their use as either outcome measures or prognostic variables.^{4 5} Specifically, no test possessed the measurement properties outlined by the IOC to serve as part of a musculoskeletal injury screening examination. Further, no single PPT had sufficient measurement properties to warrant use in monitoring recovery and determining readiness for return to play. Therefore, a comprehensive research agenda, including a review of literature, on upper extremity PPTs is needed.

In comparison to PPTs that focus on the lower extremity, there are fewer upper extremity PPTs. This is despite the time lost from sport and long-term sequelae that can result from upper extremity injuries.⁶ To date, such metrics as the reliability, validity and responsiveness of upper extremity PPTs have yet to be comprehensively reported. Therefore, the purpose of this study was to systematically review the literature pertaining to PPTs of the upper extremity, and to examine the quality of the literature and measurement properties of established tests. Our hypothesis was that findings would be similar to those of the lower extremity: limited evidence to support the use of PPTs as screening examinations or outcome measures for musculoskeletal injury.

METHODS

We framed our research question using the Patient, Intervention, Comparison and Outcome (PICO) method. Our primary research question was: ‘Do upper extremity performance tests have any relationship to upper extremity injuries in athletes of any age?’ Secondary research questions were whether PPTs predicted injury, and whether PPTs could be used as an outcome measure in the clinic.

Literature search

The search was conducted in PubMed and translated with the assistance of a librarian for CINAHL and SPORTDiscus databases (see online supplementary appendix A). All articles were included up to 17 April 2015. The search strategy included terms pertaining to athletes and injury confined to the upper extremity, and these were combined with terminology relating to performance tests. Results were limited to English language and humans. In addition, a hand search was conducted based on the identification of review articles using the ‘Clinical Queries’ function in PubMed. Reference lists of these reviews were assessed for potentially

To cite: Tarara DT, Fogaca LK, Taylor JB, et al. *Br J Sports Med* Published Online First: [please include Day Month Year] doi:10.1136/bjsports-2015-095198

eligible articles, as were the reference lists of any article that we read in full. Finally, Google Scholar was searched using the names of known upper extremity PPTs.

Eligibility criteria

Two authors (DTT and LKF) reviewed the titles and abstracts of all studies identified during the literature search to generate a list of selected articles to read in full. Discrepancies were resolved via consensus discussion, and a third author (EJH) refereed if disagreements could not be resolved.

The inclusion criteria were:

- ▶ Analysed at least one upper extremity PPT that was accessible across settings (clinic, courtside, field side), and feasible (portable, affordable and easy to administer).
- ▶ (1) Studied an exclusive population of organised professional, collegiate or high school sports athletes (eg, baseball, volleyball, tennis, javelin); (2) used terms 'elite athlete', 'professional athlete', 'semi-professional athlete', 'club sport', 'intramural athlete', 'recreational athlete' or 'sport participant'; or (3) when 50% of study's participants had a physical activity level 5 or high based on the Tegner scale.⁷
- ▶ Original, primary research (reviews were excluded).

We operationally defined PPT as a subset of functional assessments⁸ used by sports medicine clinicians to discern aspects of athleticism (power, agility, endurance, flexibility), injury risk and return-to-play readiness.⁴ PPTs were accepted as upper extremity if they tested the region of the body from shoulder girdle, including the scapula, and supporting structures to the end of the fingers.

We excluded studies of PPTs that used equipment for three-dimensional motion capture, upper body ergometers, rowing ergometers, or other technology-dependent instrumentation, conference proceedings, dissertations and theses, case studies, and case series. All articles were read in full by two authors (EJH and DTT) to further include or exclude them for scoring of methodological quality and quality of measurement properties. Any discrepancies were resolved via consensus discussion or consultation with a third author (LKF) if consensus could not be reached.

Data extraction, summaries and best evidence synthesis

Included studies were summarised based on population, injury classification, sport, PPT description and study results. PPTs were grouped based on testing procedures to determine whether the naming of PPTs and their conduct was consistent across studies. Finally, we summarised the methodological quality of the literature and the quality of the measurement properties of the PPTs and combined them using a best evidence synthesis.⁹ One author scored the methodological quality and measurement properties (DTT).

A modified version of the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) tool^{10 11} was used to assess the quality of methodology of all included articles. Our version of the COSMIN tool consisted of five categories (reliability, agreement, hypothesis testing, criterion validity and responsiveness). The methodological quality of each categorical property was individually scored on a scale of *excellent*, *good*, *fair* and *poor*. To determine the quality of PPT measurement properties, the Terwee Scale of positive (+), indeterminate (?) and negative (-) was used.¹²

For the best evidence synthesis, we combined the results of the COSMIN tool assessment¹³ and the Terwee Scale¹² for each article. Since the COSMIN scoring guidelines were originally designed for health-related patient-reported outcomes

(HR-PROs)¹⁰ and adapted for performance measures,¹⁴ final COSMIN quality scores were not based on sample size to avoid automatic exclusion of small ($n < 30$) studies.¹⁵ The COSMIN's original scoring guidelines sets a minimal threshold for adequate sample size at > 30 (fair quality).¹³ Automatically disqualifying small studies (< 30) as poor quality may unnecessarily exclude studies that report large effect sizes for physical performance measures.¹⁴ Following the recommendation by Bartel *et al*,¹⁴ studies scored as poor quality based solely on small sample sizes ($n < 30$) and without formal power analysis were retained, and accounted for in the best evidence synthesis as limited evidence. The lowest score methodology was used; poor quality studies were eliminated, as recommended previously, to contextualise best evidence.^{4 14} The grading key⁴ for the best evidence summary was:

- ▶ Unknown: investigated in studies of exclusively poor methodology or not investigated in any study.
- ▶ Conflicting: contradictory findings.
- ▶ Limited: one study of fair methodological quality.
- ▶ Moderate: multiple studies of fair methodological quality or one study of good methodological quality.
- ▶ Strong: multiple studies of good methodological quality or at least one study of excellent methodological quality.

RESULTS

Description of included studies and PPTs

The database searches identified 1021 studies for preliminary title/abstract screening. A majority ($n = 854$) of studies were eliminated for not meeting the inclusion criteria of PPT, athletic population or upper extremity (UE) body region. In addition, 4 summary abstracts, 114 case reports, 17 narrative reviews and 19 systematic reviews were excluded. Thirteen studies were eligible for full-text review. The preliminary article selection process yielded a moderate¹⁶ correlation ($\kappa = 0.55$) with 99.2% agreement. The full-text article screening process eliminated 11 studies for not meeting inclusion criteria as a feasible PPT or being administered in an athletic population. Hand searching identified nine additional articles for inclusion. The outline of study selection and inclusion is in figure 1. The preliminary analysis consisted of 11 studies (table 1). Online supplementary appendix B provides the COSMIN scoring tables.

Summary of methodology quality (COSMIN) of included studies

Reliability

Based on two reliability studies of upper extremity PPTs, there is poor²⁵ to good²⁰ evidence of test-retest reliability. Both studies reported good to excellent interclass coefficient correlation (ICC) values (> 0.75). A small sample size was used for the one-arm hop test,²⁵ which resulted in a poor quality rating. There was good methodological quality evidence for the reliability of the closed kinetic chain upper extremity stability test (CKCUEST).²⁰

Agreement/measurement error

Measurement error of UE PPTs was analysed in one study of good methodological quality, which reported measurement error (SE measurement (SEM)) and minimal detectable change (MDC) with 95% CIs for the CKCUEST.²⁰

Hypothesis testing/construct validity

There was conflicting evidence from two studies for hypothesis testing^{24 27} and consistent evidence from two studies of construct validity.^{20 21} Hypothesis testing of the unilateral seated

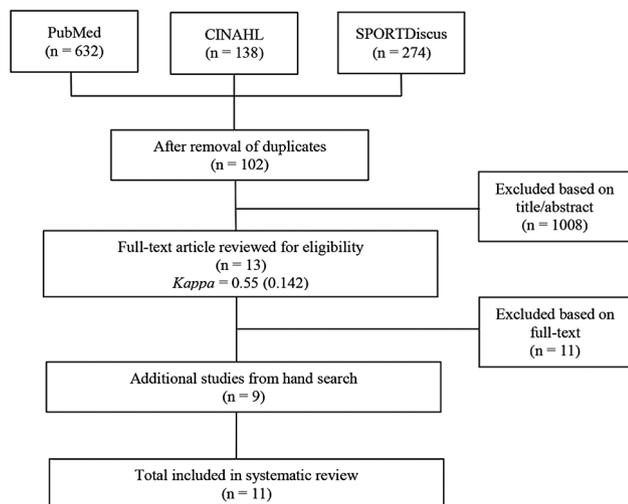


Figure 1 Flow chart of study selection and inclusion.

shot put test was of excellent quality,²⁴ while the CKCUEST was of fair quality.²⁷ Methodological quality was excellent for hypothesis testing of the unilateral seated shot put test.²⁴ The construct validity was established for the CKCUEST and evidence was of fair quality.^{20 21}

Criterion/predictive validity

Two studies, of fair methodological quality, investigated the criterion validity of PPTs of the shoulder.^{22 23} Both correlated One-repetition maximum (1RM) bench press values to the 4.5 kg two-handed seated shot put test distance. Predictive validity for the CKCUEST was reported in one study, but was of poor quality due to a small sample size (n=26) and a prediction model based on only six injury cases.¹⁸

Responsiveness

Responsiveness of upper extremity PPTs was reported in two studies.^{17 26} The CKCUEST was used to quantify improvement for SICK (Scapular malposition, Inferior medial border prominence, Coracoid pain and malposition, and dysKinesis) scapula in asymptomatic overhead athletes following a 3-week open kinetic chain exercise intervention and was rated as poor quality because of low sample size (n<30).¹⁷ The seated medicine ball throw was used to quantify improvement following traditional and maximal concentric acceleration strength training programmes and was rated to be of fair methodological quality.²⁶

Summary of quality of measurement properties (Terwee Scale) for included PPTs

Ten studies reported measurement properties. One study²⁷ did not report any measurement properties included to the COSMIN tool or Terwee Scale, and was excluded from further analysis. Measurement properties were reported for the following six PPTs (online supplementary appendix C provides the Terwee Scale scoring tables):

- ▶ CKCUEST;^{17 18 20 21}
- ▶ Seated shot put (two hands);^{22 23}
- ▶ Unilateral seated shot put;²⁴
- ▶ Medicine ball throw;²⁶
- ▶ Modified push-up test;²⁷
- ▶ One-arm hop test.²⁵

The seated shot put (two hands), unilateral seated shot put and medicine ball throw were not grouped due to differences in

testing procedures (see table 1). The quality of PPT measurements, based on the Terwee Scale, is described below.

Reliability

Test-retest reliability was reported for two of the six PPTs. The studies reporting on one-arm hop test²⁵ and CKCUEST²⁰ were of positive quality and reported ICC values >0.75.

Agreement/measurement error

One study reported data consistent with agreement/measurement error (SEM and MDC for the CKCUEST).²⁰ Since the Terwee Scale scores this category based on reported minimal important change (MIC) and smallest detectable change (SDC) values,¹² we calculated the MIC and SDC. There was positive evidence for the quality of the measurement properties of agreement/measurement error for the CKCUEST.

Hypothesis testing/construct validity

Hypothesis testing of unilateral seated shot put yielded positive quality.²⁴ Conversely, the modified push-up test²⁷ was of negative quality. Construct validity of the CKCUEST was of positive quality for discriminate validity²⁰ in distinguishing between healthy, active UE recreational sport athletes and sedentary individuals with a positive history of subacromial impingement syndrome. With respect to convergent validity, the CKCUEST was rated as negative quality in correlating to upper quarter y balance test (UQYBT).²¹

Criterion/predictive validity

The two-hand seated shot put^{22 23} demonstrated negative quality criterion validity when referenced to 1-RM bench press values. Criterion validity, in this context, is an indication of how a PPT reflects a 'gold' or reference standard based on statistical correlation.¹³ There was positive quality for predictive validity: the CKCUEST prospectively predicted in-season shoulder injury in male, collegiate football players.¹⁸

Responsiveness

The responsiveness of the CKCUEST¹⁷ and the medicine ball throw²⁶ were both of positive quality, with demonstrated score improvement after following 3-week¹⁷ and 14-week²⁶ resistive exercise programmes, respectively.

Best evidence synthesis

The results of the best evidence synthesis are summarised by test in table 2.

Closed kinetic chain upper extremity stability test

Several studies reported on the CKCUEST. However, two articles^{17 18} were eliminated for poor quality evidence and one for not reporting measurement properties.¹⁹ The two remaining studies^{20 21} provided varied evidence based on measurement properties. There was moderate evidence for reliability and agreement,²⁰ fair evidence for discriminate construct validity,²⁰ and limited evidence for convergent construct validity.²¹ The CKCUEST discriminated between young (21.7–23.1 years), active, recreational athletes and sedentary adults (45.1–49.8 years). With respect to convergent validity, the relationship between the UQYBT and CKQUEST had a positive weak correlation for the dominant (r=0.43) and non-dominant (r=0.49) sides.^{16 21} The responsiveness, criterion validity and injury prediction in athletic populations is unknown.

Table 1 Summary of included studies in preliminary analysis

Study	Population/UE injury status	Description	Results
CKQUEST			
Jayesh <i>et al</i> ¹⁷	n=20 healthy; 15 male and 5 female; asymptomatic SICK scapula	From a push-up position with hands 36 inches apart, shoulders over hands, one submaximal trial; 45 s rest; three maximal effort trials with 45 s rest intervals; scored based on average of maximal effort trials; score was normalised based on body height	CKQUEST pretest and post-test scores showed significant ($p \leq 0.05$) improvement and were 21.6±3.9 and 25.8±4.8, respectively
Pontillo <i>et al</i> ¹⁸	n=26, division I college football players; 1 (n=6) shoulder injuries	From start position, one hand is brought over the opposite tape line, returned and repeated with the opposite hand; highest number of touches within 15 s; two trials with 1 min rest	CKQUEST was significant at predicting injury with a cut score of <21; Sn=0.83, Sp=0.79, +LR=4.47 and -LR=0.25
Roush <i>et al</i> ¹⁹	n=77 healthy division III collegiate baseball players	Two 1.5 in tape strips, parallel, 36 inches apart. Start in a push-up position with one hand on each piece of tape. Perform alternating cross-over touches under the opposing hand for 15 s followed by 45 s rest. Scored average of three trials	No difference in CKQUEST scores between playing position. Average CKQUEST 30.4±3.87 touches
Tucci <i>et al</i> ²⁰	n=108 participants; 40 UE sport-specific recreational athletes; I; (n=28) Hx of subacromial impingement syndrome	Two tape markers 36 inches apart: from push-up position, the number of supporting hand touches with the swing hand during 15 s. For female participants: test performed from the modified push-up position, from the knees. Scored average of three trials	Intersession reliability: ICC=0.85 to 0.96; Intrasession reliability: ICC=0.86 to 0.96. In males and females, touches were greater in active (24.5) vs sedentary (22.6) vs SIS (10.1) and active (27.9) vs sedentary (24.7) vs SIS (12.2), respectively; SEM ranged from 1.45 to 2.76 touches; MDC ranged from 2.05 to 3.91 touches
Westrick <i>et al</i> ²¹	n=30 healthy USMA cadets; 24 males, 6 females; healthy	Two tape markers 36 inches apart. Start in a push-up position with one hand on each piece of tape. Perform alternating hand crosses under the opposing hand. Number of touches for 15 s was recorded. Scored average of three trials	Correlation between dominant and non-dominant UQYBT and CKQUEST, respectively ($r=0.43$ and $r=0.49$). Average CKQUEST 18.8±3.3 (males) and 15.9±4.5 (females), no statistical difference based on gender
Seated shot put			
Mayhew <i>et al</i> ²²	n=40 division II college football players; injury status not reported	Performed with 4.5 kg medicine ball; seated on the floor with back against a support; feet flat on the floor with knees at 90 degree angle; two hand push from centre of chest. Three trials with 30 s rest, average distance thrown used for analysis	Seated shot put correlated to the 1-RM bench press, $r=0.57$
Mayhew <i>et al</i> ²³	NR, n=64 college female athletes; Injury status not reported		Seated shot put correlated to the 1-RM bench press, $r=0.38$
Unilateral seated shot put			
Chmielewski <i>et al</i> ²⁴	n=125 healthy college-aged athletes	Performed with a 2.72 kg medicine ball; ball starts at shoulder height; participant pushes (not throws) the ball. Scored based on average of three trials	Allometric scaling (exponent of 0.35) accounted for a 5–10% greater distance on the dominant side. Males produced greater seated shot put distances compared with women (119.0±9.3 vs 73.3±8.2 cm/kg ^{0.35})
One-arm hop test			
Falsone <i>et al</i> ²⁵	n=26 healthy, 13 wrestlers, 13 football players	Participants allowed a warm-up and practice trial(s); from a one arm push-up position, participant completed five proper technique one-arm hops onto a 10.2 cm step and scored for time to complete task	Test-retest reliability was ICC 0.81 in the wrestlers and 0.78 in the football players. No statistically difference between dominant and non-dominant arms, with non-dominant arm values ranging from 1.1% faster to 9.8% slower than the dominant arm
Medicine ball throw			
Jones <i>et al</i> ²⁶	n=40 college-aged football players; injury status not reported	Seated on a 20-inch bench with back against the wall; using two hands, subjects tossed 12 lb medicine ball at a distance. The back and scapula had to maintain contact with the wall; three warm-up throws followed by three max effort throws. Scored based on longest throw	Significant training effect for all tests. No significant difference between experimental and control groups. Pre-test and post-test medicine ball throw: 7.9±0.5 vs 8.1±0.6 m
Modified push-up			
Ambegaonkar <i>et al</i> ²⁷	n=32 healthy college-aged female dancers; n=15 female physically active non-dancers	Participant performed a modified push-up from knees with lower legs together in contact with the floor with ankles in dorsiflexion; down position—chin to floor, up position— full elbow extension; required to maintain strict form	No difference between groups. College-aged dancers 22.2±8.6; physically active college-aged non-dancers 19.9±8.2

CKQUEST, closed kinetic chain upper extremity stability test; ICC, interclass coefficient correlation; LR, likelihood ratio; MDC, minimal detectable change; SEM, SE measurement; SIS, subacromial impingement syndrome; SICK, Scapular malposition, Inferior medial border prominence, Coracoid pain and malposition, and dyskinesia; Sp, specificity; Sn, sensitivity; UQYBT, upper-quarter y balance test; UE, upper extremity Hx, history; USMA, United States Military Academy.

Unilateral seated shot put test

There was moderate evidence for hypothesis testing based on a good methodology, and positive quality measurement properties with hypothesis testing in one study.²⁴ Chmielewski *et al*²⁴ reported that the unilateral seated shot put test should be scored on an allometric scale (exponent of 0.35), and in healthy athletes, a 5–10% great difference can be accounted for when

comparing dominant to non-dominant arm. The reliability, agreement, criterion validity and responsiveness are unknown. There is no evidence pertaining to athletes with UE injury.

Seated shot put (two hands)

Two studies^{22, 23} reported fair evidence on the criterion validity of the seated shot put test in relationship to 1-RM bench press.

Table 2 Best evidence synthesis by PPT

Measurement properties	Unknown (???)	Conflicting (±)	Limited (+)(-)	Moderate (++)(--)	Strong (+++)(---)
CKQUEST					
Reliability				++	
Agreement				++	
Hypothesis testing				++	
Criterion validity	???				
Responsiveness	???				
Unilateral seated shot put					
Reliability	???				
Agreement	???				
Hypothesis testing				++	
Criterion validity	???				
Responsiveness	???				
Seated shot put					
Reliability	???				
Agreement	???				
Hypothesis testing	???				
Criterion validity				--	
Responsiveness	???				
Medicine ball throw					
Reliability	???				
Agreement	???				
Hypothesis testing	???				
Criterion validity	???				
Responsiveness			+		
One-arm hop test					
Reliability			+		
Agreement	???				
Hypothesis testing	???				
Criterion validity	???				
Responsiveness	???				
Modified push-up					
Reliability	???				
Agreement	???				
Hypothesis testing			-		
Criterion validity	???				
Responsiveness	???				

+, positive rating; -, negative rating; CKQUEST, closed kinetic chain upper extremity stability test; conflicting, contradictory findings; grading key, unknown (???)=investigated in studies of exclusively poor methodology or not investigated in any study; limited, one study of fair methodological quality; moderate, multiple fair methodological studies or one study of good methodology; PPT, physical performance test; strong, multiple studies of good methodological rating or at least one study of excellent methodology.

The correlation between the PPT and strength measure was weak ($r=0.38$) in female athletes and male ($r=0.17$ to 0.57) athletes. There was no evidence for reliability, agreement, construct validity, responsiveness or injury prediction.

Medicine ball throw

There was limited evidence for the measurement properties of the seated medicine ball throw. Jones *et al*²⁶ reported 2.8–9.4% improvement in throwing distance following traditional and maximal concentric acceleration strength training programmes. These findings reflect positive quality measurement properties for the PPT, but were only demonstrated in healthy athletes. The evidence relating to reliability, agreement, hypothesis testing/construct validity, and criterion validity for the seated medicine ball throw are unknown.

One-arm hop test

Limited quality evidence is available for the one-arm hop test. One study reported measurement properties for this PPT in a

small sample ($n=26$) of uninjured athletes.²⁵ Test-retest reliability was established in subset of 13 wrestlers ($ICC=0.8$) and 13 football players ($ICC=0.78$). One-arm hop test scores were 4.4% lower in the non-dominant versus dominant arm. The evidence relating to agreement, hypothesis testing/construct validity, criterion validity and responsiveness for the one-arm hop test is unknown.

Modified push-up test

The modified push-up test had fair evidence in discriminate construct validity, but was unable to distinguish between healthy university-level modern dancers and non-dancers.²⁷ No evidence was available for reliability, agreement, criterion validity, responsiveness and injury prediction.

DISCUSSION

This study provided a systematic review of upper extremity PPTs in response to a broader research recommendation specific to the use of PPTs in sports medicine.⁵ The methodology

quality (COSMIN) of eight studies and the quality of measurement properties (Terwee Scale) of six PPTs were included in the synthesis of best evidence. Consistent with our hypothesis, the evidence relating to upper extremity PPTs is mostly unknown or limited.

The use of upper extremity PPTs as assessment tools must be grounded in the tenants of reliability and validity. In general, the available studies that report UE PPTs in athletic populations fail to provide adequate reliability and validity evidence based on COSMIN criteria due to a number of factors. The first issue relates to studies relying on poor ($n < 30$)^{17 18 19 25} and fair ($n = 30-49$)^{21 22 26 27} quality sample sizes. Small sample sizes affect statistical power and limit the generalisability of study results. The exceptions to this were studies reporting on the CKCUEST,²⁰ seated shot put²³ and unilateral shot put.²⁴ However, the use of a small subgroup of healthy ($n = 40$) upper extremity sport-specific athletes ($n = 40$) and upper extremity-injured athletes ($n = 28$) to evaluate the reliability, agreement and construct validity (discriminate) of the CKCUEST weakened the methodological quality of the study examining the CKCUEST. Collectively, these findings highlight a problematic issue relating to the low quality of research availability on UE PPTs.

Studies reporting strong evidence relating to reliability, agreement, hypothesis testing/construct validity, criterion validity and responsiveness of upper extremity PPTs are lacking. Research designs consistent with moderate evidence relating reliability, agreement and construct validity (discriminate) are only available based on two studies reporting on CKCUEST.^{20 21} One study presented moderate evidence of criterion validity for the unilateral shot put.²⁴ Limited evidence from two studies was available for the one-arm hop test reliability²⁵ and medicine ball throw responsiveness.²⁶ Aside from these investigations, the quality of research designs for the remaining studies was mostly rated as unknown due to an absence of research focused on the constructs of reliability and validity.

Beyond the COSMIN criteria, the majority of studies included in the best evidence synthesis examined healthy athletes^{21 24 27} or failed to report upper extremity injury history.^{22 23 26} Tucci *et al*²⁰ included an upper extremity-injured group with a history of subacromial impingement syndrome to determine construct (discriminate) validity for CKCUEST scores. However, comparisons were made between an older (41.5–49.8 years) sedentary injured group and a younger (21.7–23.1 years) upper extremity athletic population. The difference in age and athletic status is a confounder when interpreting the discriminate validity of the CKCUEST for injury status, and calls into question the validity of this test. The absence of studies that evaluate the measurement properties of PPTs in injured athletes is problematic for sports medicine practitioners to interpret PPTs in the context of injury screening, evaluation and rehabilitation responsiveness.

The naming and testing procedures for comparative tests were not a matter of concern because of the low number of PPTs in the best evidence synthesis. The CKCUEST was performed and measured consistently across studies^{20 21} and was reflective of the original test description.²⁸ In one study, the CKCUEST was modified for females by allowing a push-up start position for the knees.²⁰ Although this modification to the original test may seem like a reasonable accommodation, it is a deviation that may influence scoring and limit generalisability, and warrants separate reliability and validity assessment.

The unilateral seated shot put,²⁴ seated shot put^{22 23} and medicine ball throw²⁶ are PPTs designed to measure upper extremity power. As these tests may appear to measure the same construct, it is tempting for clinicians to group them together

for comparative analysis. However, this should not be done based on differences in testing procedures relating to the weight of the ball and alignment of the trunk and upper extremity. Despite modest evidence that supports the relationship of these tests to other measures of strength, there has yet to be any data supporting their inter-relationship. Thus, each test should be considered independently. For example, the unilateral seated shot put may be the best test to quantify performance and outcomes in overhead sport athletes (eg, tennis, volleyball, baseball) who rely on dominant arm performance.

Limitations

Systematic reviews are prone to a number of limitations²⁹ and there are several studies related to this review worthy of consideration. The quality of our systematic review is dependent on the availability and quality of primary literature catalogued in electronic databases. There is no standard search strategy for PPTs in athletic populations,⁴ nor is there a MeSH (Medical Subject Headings) term to facilitate the cataloguing and searching of PPTs. The absence of a MeSH term may compromise identification of all relevant studies.³⁰ However, to account for this, we included empirical³¹ and manual³⁰ search strategies in addition to the electronic database searches. Our search was confined to English language studies due to language translation limitations. Therefore, we may have missed articles that were published in languages other than English.

Our operational definition for qualifying PPTs was restricted to tests that were portable, affordable and easy to administer in diverse settings, including clinics, courtside or field side. This eliminated studies and PPTs that required technology (force plates, three-dimensional motion capture) for data collection or scoring.

Methodological quality and PPT measurement properties were assessed with the COSMIN and Terwee Scale, respectively. These tools were originally designed to evaluate the quality of HR-PROs with respect to reliability, validity and responsiveness.^{12 13 32} Recently, the COSMIN and Terwee Scale have been adapted for systematic reviews relating to PPTs.^{4 5 14 15 33} Although their measurement properties have been questioned,^{14 15 33} these tools provide an objective means for determining quality and evidence relating to PPTs. Lastly, the review was based on small number of available PPTs.

Summary

There are a limited number of upper extremity PPTs that can be used as part of a musculoskeletal screening examination or an outcome measure in athletic populations. Based on current evidence, the CKCUEST, unilateral seated shot, one-arm hop test and medicine ball throw may have promise as PPTs. However, the utility of these PPTs in injured populations is unsubstantiated in literature and warrants further investigation. Higher quality research needs to focus on the measurement properties of upper extremity PPTs across a spectrum of healthy and injured athletic populations. Considering this, future research needs to consider standardisation of PPT names, descriptions, and testing procedures to ensure best methodological quality and appropriate comparisons of measurement properties. This is a critical link for reporting PPTs in sports medicine literature and to support the clinical feasibility of tests by physiotherapists, athletic trainers, coaches and trainers. In the absence of sound evidence relating to reliability and validity, upper extremity PPTs provide little utility in determining the scope of functional performance or injury status in athletic populations.

What are the findings?

- ▶ Six upper extremity physical performance tests (PPTs) were investigated and provided modest evidence pertaining to reliability and validity.
- ▶ The available studies on upper extremity PPTs should use larger sample sizes to provide higher quality evidence and generalisability of findings.
- ▶ The closed kinetic chain upper extremity stability test (CKCUEST) and unilateral seated shot put were associated with moderate quality evidence relating to reliability, agreement and hypothesis testing.
- ▶ The evidence pertaining to one-arm hop test reliability and medicine ball throw responsiveness was of limited quality.

How might it impact on clinical practice in the future?

- ▶ Clinicians should exercise caution in using upper extremity PPTs in athletic populations as musculoskeletal screening tools or outcome measures, as current tests lack evidence to suggest their widespread use in clinical practice.
- ▶ The CKCUEST has moderate test-retest reliability and may demonstrate discriminate validity in subacromial impingement syndrome.
- ▶ The unilateral shot put test may be used to quantify dominant versus non-dominant (−5% to 10%) upper extremity power differences.
- ▶ The medicine ball throw is best used to quantify responsiveness to upper extremity strength training programmes in healthy athletes.

Contributors DTT, EJM and LKF planned the study, executed the searches, examined the articles for quality and edited the manuscript. DTT and JBT wrote the initial version of the manuscript. JBT edited and contributed to all other versions of the manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement There are additional figures from the meta-analysis that the authors are happy to share with receipt of a written request by the corresponding author.

REFERENCES

- 1 PPE preparticipation physical evaluation. Bernhardt DT, Roberts WO, eds. 4th edn. Elk Grove Village, IL: American Academy of Pediatrics, 2010.
- 2 Ljungqvist A, Jenoure P, Engebretsen L, et al. The International Olympic Committee (IOC) Consensus Statement on periodic health evaluation of elite athletes March 2009. *Br J Sports Med* 2009;43:631–43.
- 3 Hegedus EJ, Cook CE. Return to play and physical performance tests: evidence-based, rough guess or charade? *Br J Sports Med* 2015;49:1288–9.
- 4 Hegedus EJ, McDonough S, Bleakley C, et al. Clinician-friendly lower extremity physical performance measures in athletes: a systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests. *Br J Sports Med* 2015;49:642–8.
- 5 Hegedus EJ, McDonough SM, Bleakley C, et al. Clinician-friendly lower extremity physical performance tests in athletes: a systematic review of measurement properties and correlation with injury. Part 2—the tests for the hip, thigh, foot and ankle including the star excursion balance test. *Br J Sports Med* 2015;49:649–56.
- 6 Robinson TW, Corlette J, Collins CL, et al. Shoulder injuries among US high school athletes, 2005/2006–2011/2012. *Pediatrics* 2014;133:272–9.
- 7 Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res* 1985;00:43–9.
- 8 Reiman MP, Manske RC. The assessment of function: How is it measured? A clinical perspective. *J Man Manip Ther* 2011;19:91–9.
- 9 Furlan AD, Pennick V, Bombardier C, et al. Editorial Board, Cochrane Back Review Group. 2009 Updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* 2009;34:1929–41.
- 10 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- 11 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- 12 Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- 13 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
- 14 Bartels B, de Groot JF, Terwee CB. The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties. *Phys Ther* 2013;93:529–41.
- 15 Dobson F, Hinman RS, Hall M, et al. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage* 2012;20:1548–62.
- 16 Portney LG, Watkins MP. *Foundations of clinical research: application to practice*. Upper Saddle River, NJ: Pearson Education, Inc, 2009.
- 17 Jayesh PN, Muragod AR, Motimath B. Open kinematic chain exercises for SICK scapula in competitive asymptomatic overhead athletes for 3 weeks. *Int J Physiother Res* 2014;2:608.
- 18 Pontillo M, Spinelli BA, Sennett BJ. Prediction of in-season shoulder injury from preseason testing in division I collegiate football players. *Sports Health* 2014;6:497–503.
- 19 Roush JR, Kitamura J, Waits MC. Reference values for the Closed Kinetic Chain Upper Extremity Stability Test (CKCUEST) for collegiate baseball players. *N Am J Sports Phys Ther* 2007;2:159–63.
- 20 Tucci HT, Martins J, Sposito Gde C, et al. Closed Kinetic Chain Upper Extremity Stability Test (CKCUES test): a reliability study in persons with and without shoulder impingement syndrome. *BMC Musculoskelet Disord* 2014;15:1.
- 21 Westrick RB, Miller JM, Carow SD, et al. Exploration of the y-balance test for assessment of upper quarter closed kinetic chain performance. *Int J Sports Phys Ther* 2012;7:139–47.
- 22 Mayhew JL, Bembem MG, Piper FC, et al. Assessing bench press power in college football players: the seated shot put. *J Strength Cond Res* 1993;7:95–100.
- 23 Mayhew JL, Bembem MG, Rohrs DM, et al. Specificity among anaerobic power tests in college female athletes. *J Strength Cond Res* 1994;8:43–7.
- 24 Chmielewski TL, Martin C, Lentz TA, et al. Normalization considerations for using the unilateral seated shot put test in rehabilitation. *J Orthop Sports Phys Ther* 2014;44:518–24.
- 25 Falsone SA, Gross MT, Guskiewicz KM, et al. One-arm hop test: reliability and effects of arm dominance. *J Orthop Sports Phys Ther* 2002;32:98–103.
- 26 Jones K, Hunter G, Fleisig G, et al. The effects of compensatory acceleration on upper-body strength and power in collegiate football players. Les effets de l'acceleration compensatoire sur la force et la puissance du corps de joueurs de football universitaires. *J Strength Cond Res* 1999;13:99–105.
- 27 Ambegaonkar JP, Caswell SV, Winchester JB, et al. Upper-body muscular endurance in female university-level modern dancers: a pilot study. *J Dance Med Sci* 2012;16:3–7.
- 28 Goldbeck TG, Davies GJ. Test-retest reliability of the closed kinetic chain upper extremity stability test: a clinical field test. *J Sport Rehabil* 2000;9:35–45.
- 29 Hemmingway P, Breerton N. *What is a systematic review?* 2nd edn. Covent Garden, London: Hayward Medical communications, 2009.
- 30 Whiting P, Westwood M, Burke M, et al. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *J Clin Epidemiol* 2008;61:357–64.
- 31 Montori VM, Wilczynski NL, Morgan D, et al. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68.
- 32 Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
- 33 Kroman SL, Roos EM, Bennell KL, et al. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage* 2014;22:26–39.