J Shoulder Elbow Surg (2019) 28, 178–185

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

REVIEW ARTICLES

The reliability, validity, and methodologic quality of measurements used to quantify posterior shoulder tightness: a systematic review of the literature with meta-analysis

Paul A. Salamh, PT, DPT, PhD[a,*], Xinliang Liu, PhD[b], Morey J. Kolber, PT, PhD[c], William J. Hanney, PT, DPT, PhD[d], Eric J. Hegedus, PT, DPT, PhD[e]

[a]*Krannert School of Physical Therapy, University of Indianapolis, Indianapolis, IN, USA*
[b]*Department of Health Management and Informatics, University of Central Florida, Orlando, FL, USA*
[c]*Department of Physical Therapy, Nova Southeastern University, Fort Lauderdale, FL, USA*
[d]*Doctor of Physical Therapy Program, University of Central Florida, Orlando, FL, USA*
[e]*Department of Physical Therapy, High Point University, High Point, NC, USA*

Hypothesis and background: Posterior shoulder tightness (PST) has been linked to numerous shoulder pathologies in both the general and athletic populations. Several methods for documenting PST have been described in the literature, which may lend to variability in clinical practice and research. The purpose of this study was to perform a systematic review with meta-analysis to investigate the reliability, validity, and methodologic quality of methods used to quantify PST.

Methods: Relevant studies were assessed for inclusion, and selected studies were identified from the PubMed, Embase, Cochrane, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) databases. Data were extracted from the selected studies and underwent methodologic quality assessment and meta-analysis.

Results: The search resulted in 1006 studies identified, with 18 ultimately retained. Intrarater reliability was reported in 12 studies with a summary intraclass correlation coefficient of 0.93 (95% confidence interval, 0.90-0.95), whereas inter-rater reliability was reported in 6 studies with a summary intraclass correlation coefficient of 0.89 (95% confidence interval, 0.80-0.94). Validity was reported in 10 studies, all using internal rotation as the convergent standard, and was found to be significant in all but 1 study.

Conclusion: Current methods used to quantify PST have good reliability but are primarily limited to measures of horizontal adduction of the glenohumeral joint with scapular stabilization. Limitations in using a single measurement technique exist particularly as there may be multiple contributing factors to PST. A more comprehensive approach for quantifying PST is necessary, and suggested components include a cluster of techniques composed of horizontal adduction, internal rotation, and total glenohumeral joint range of motion.

This study was exempt from required ethical approval.

*Reprint requests: Paul A. Salamh, PT, DPT, PhD, Krannert School of Physical Therapy, University of Indianapolis, 1400 E Hanna Ave, Indianapolis, IN 46227, USA.

E-mail address: salamhp@uindy.edu (P.A. Salamh).

1058-2746/$ - see front matter
https://doi.org/10.1016/j.jse.2018.07.013

Level of evidence: Level IV; Systematic Review and Meta-Analysis

Keywords: Flexibility; internal rotation; mobility; range of motion; shoulder; horizontal adduction

Shoulder pain affects up to 67% of the adult population throughout their life span.[23,24] The etiology of shoulder pain is multifactorial and inclusive of numerous impairments, including but not limited to restricted mobility. Posterior shoulder tightness (PST), in particular, has been associated with more common diagnoses such as labral tears, impingement syndrome, and postoperative arthrofibrosis among both the general and athletic populations, with a predilection for the overhead athlete.[5,6,14,27,28,32,35,36] Owing to the well-documented association of PST and shoulder pain in athletes, sports medicine professionals often seek to improve PST as a key element of their interventions. PST has been defined as a limitation of the extensibility within the posterior soft tissue of the shoulder including both contractile and noncontractile elements as well as osseous changes as seen in the form of humeral torsion within the overhead athlete through training adaptations.[13] Moreover, PST has been associated with restricted internal rotation, horizontal adduction (HA), and flexion range of motion. From a biomechanical perspective, evidence suggests that PST may influence obligate translation, which is a source of subacromial impingement syndrome.[38]

Tyler et al[37] were the first authors to report a technique that set out to quantify PST without the use of internal rotation. Their technique was performed in the side-lying position, using a carpenter square to measure the distance between the medial epicondyle and the plinth during HA with scapular stabilization. Since then, a variety of techniques and instruments have been developed and used to quantify PST including the use of inclinometers (digital and bubble), goniometers, smartphones, diagnostic ultrasound, myotonometers, and anthropometers. Given the potential association between PST and shoulder pathology, a measurement technique that is both reliable and valid should provide clinicians and researchers with the necessary information for documenting impairments, determining intervention strategies, and measuring the amount of change as a result of interventions. Thus, the purpose of this study was to perform a systematic review and meta-analysis of the current literature that has investigated the reliability and validity of measurement techniques used to quantify PST. A review of the literature on PST was performed in 2011; however, the review did not include a meta-analysis and did not consider potential risk of bias, which limits the overall utility of the study results.[30] The aforementioned review concluded that methods for quantifying PST at the time showed poor to good reliability; however, it did not investigate the methodologic quality of the included studies. Furthermore, new techniques have been established and published since the previously published review.

Materials and methods

Guidelines

This systematic review used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines during the search and reporting phase of the research process. The PRISMA statement includes a 27-item checklist designed to improve reporting of systematic reviews and meta-analyses.[25]

Literature search

An online literature search was conducted using the PubMed, Embase, Cochrane, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) databases from the dates of their origin until August 2017. The search strategy was created and performed by a biomedical librarian. An example of the search strategy used for the PubMed database is shown in Supplement 5 and similar strategies were used for the remaining databases. This study was registered using the international prospective register of systematic reviews, PROSPERO, in March 2016 (reference No. CRD42016036281).

Study selection

All titles were independently appraised during the initial online literature search for studies by 2 authors (P.S. and M.K.) who are researchers with a clinical background in musculoskeletal orthopedics. The abstracts of these titles were then read to determine whether the studies met the inclusion criteria. Studies with abstracts that met the inclusion criteria were read in their full-text format to determine eligibility. Two authors undertook the study selection process for this review, with discrepancies being decided by a third author if necessary (no discrepancies were present that needed to be brought to a third author). The inclusion criteria for studies included in this review consisted of the following: (1) studies that investigated the reliability and/or validity of measurements used to quantify PST; (2) human participants of any age; (3) any study design, including case studies; (4) measurement techniques that were clearly described and applied directly to the glenohumeral joint; (5) studies that used within- or between-rater analysis; and (6) no restriction on date published. Studies were excluded if they were not published in the English language as we would be unable to interpret the findings.

Data extraction

Data and results from the studies selected for this review were extracted using a standardized Population, Intervention (Treatment),

Comparison, Outcome, and Setting (PICOS) format. The PICOS framework includes the characteristics of the population studied, treatments performed, comparative treatments, primary and secondary outcomes, and setting in which the data were collected. Data were extracted, reviewed, and analyzed by 2 authors. A single author extracted data, and this was verified by a second author. Discrepancies in data collection were resolved through discussion.

Quality assessment and best-evidence synthesis

Each of the articles was assessed for both methodologic quality and quality of measurement properties by 2 authors (P.S. and E.H.) who are researchers with a clinical background in musculoskeletal orthopedics with experience in quality scoring, using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist and the Terwee scale, respectively.[26,33] Disagreement was resolved by discussion between the 2 independent scorers. The COSMIN checklist was originally created to evaluate patient-reported outcomes but has been modified previously for use in assessing the measurement properties of physical tests.[1,12] We modified the COSMIN checklist to assess the quality of articles examining reliability, measurement error and/or agreement, hypothesis testing and/or construct validity, criterion validity, and responsiveness. We used the Terwee scale[33] to examine the quality of the measurement properties of various tests used to assess PST. To produce a best-evidence synthesis, the findings from both quality-assessment tools were combined. At this best-evidence stage, only studies of fair or better methodologic quality were considered and the level of evidence was rated as strong, moderate, limited, conflicting, or unknown. Also at this stage, those studies without an a priori power analysis in which the sample size was less than 30 were rated as having limited evidence.[1]

Data analysis

The meta-analysis was performed for inter-rater and intrarater reliability. Intraclass correlation coefficient (ICC) values and sample size data were extracted from any study that assessed inter-rater or intrarater reliability using an ICC. To account for the non-normal distribution of ICC data, Fisher variance-stabilizing z transformation was conducted following the tradition for correlational meta-analyses.[2,10,15] Because significant heterogeneity was observed between included studies, Fisher z-transformed reliability values were combined using the DerSimonian and Laird random-effects model.[8] Homogeneity analysis was performed using the Q statistic after the meta-analysis.[2,11] The Q statistic follows a χ^2 distribution with $n - 1$ df. A small P value ($P < .05$) for the Q statistic indicates that heterogeneity is present and the meta-analysis model has some unaccounted-for bias. We converted Fisher z values back to ICC values after completing meta-analyses to make interpretation of the results more standard. The summary ICC value for each random-effects model was interpreted based on the guidelines offered by Portney and Watkins[29]: poor, 0.00-0.25; fair, 0.26-0.50; moderate, 0.51-0.75; or good, 0.76-1.00. The meta-analyses were completed using Stata statistical software (version 12.1; StataCorp, College Station, TX, USA) with the "metan" command. All data analysis was performed by X.L., an academician with a specialty in biostatistics and health informatics.

Results

Search results

The search of the databases yielded 1006 independent study titles, resulting in 898 once duplicates were removed. After the title and abstract search, 874 studies were removed because they included objective measures other than those related to PST, they did not include reliability or validity analysis, and/or they were not published in the English language. After review of the full-text studies, 18 were ultimately included in the analysis (Supplement 1). Intrarater reliability values were reported in 13 of the studies (Supplement 2), whereas inter-rater reliability values were reported in 7 (Supplement 3) and validity in 10 (Supplement 4).

Summary of methodologic quality of included studies using COSMIN checklist

Reliability

The methodologic quality of studies examining the reliability of methods to quantify PST was generally poor regardless of the method used. There were a handful of studies that received a rating of good methodologic quality[17,21,22,28]; however, all but 1 of these studies[21] had a sample size of less than 30 subjects. There were no studies that received a rating of excellent.

Agreement and/or measurement error

Three studies reported agreement and/or measurement error,[18,19,31] all regarding the side-lying method for HA in the form of the minimal detectable change. All 3 studies received a methodologic rating of poor.

Hypothesis testing and/or construct validity

Evidence of hypothesis testing and/or construct validity was present in 13 studies.[7,9,16,19,20,22,28,31,34,36,37,39,40] For HA, the methodologic quality was mixed. Of the 11 studies that investigated hypothesis testing and/or construct validity related to HA,[7,9,18,19,22,28,31,36,37,39,40] 5 (examining the supine and/or horizontal method) received a poor methodologic quality rating[7,9,22,31,37] and 6 (examining the supine and/or side-lying method) received a methodologic quality rating of fair.[19,20,28,36,39,40] One of the studies examined both the supine and side-lying methods.[28] The 2 remaining studies investigated hypothesis testing and/or construct validity with regard to the use of diagnostic ultrasound to measure posterior capsule thickness, both of which showed poor methodologic quality.[16,34]

Criterion validity

There were no studies examining the criterion validity of methods used to quantify PST.

Responsiveness

A single study reported on the responsiveness of a myotonometer for quantifying PST and received a methodologic quality rating of poor.[17] This study was performed in 15 asymptomatic subjects.

Summary of quality of measurement properties using Terwee scale

Reliability

Eleven studies examined the reliability of the HA method, including the supine and/or side-lying technique,[4,7,18-22,28,31,37,39] with 3 studies examining both methods.[4,21,28] All of the aforementioned studies (5 supine and 9 side lying) examining the reliability of the HA method scored a positive measurement property. In addition, all studies examining both the supine and side-lying HA methods scored a negative measurement property.[7,28] Reliability examination for internal rotation,[4,20] low-flexion,[3,7] and myotonometer[7,17] methods was performed within 2 studies for each method, all of which received a positive measurement property score. A single study examining reliability for both the extension–with–internal rotation method and the scapular-plane abduction method received a positive measurement property score for both.[7]

Agreement and/or measurement error

Agreement was examined in 3 studies, all with regard to the side-lying HA method.[19,20,31] Two studies received an indeterminate measurement property score,[19,31] and one received a positive measurement property score.[20]

Hypothesis testing and/or construct validity

The quality rating of hypothesis testing for studies examining the HA method was predominately positive,[18,19,22,31,36,37,40] with only a single study receiving a negative measurement property score.[9] The quality rating in 2 studies examining diagnostic ultrasound yielded both positive and negative measurement property scores.[16,34]

Criterion validity

No studies investigated criterion validity.

Responsiveness

A single study investigated responsiveness with regard to the use of a myotonometer and received a positive measurement property rating.[17]

Intrarater reliability

Intrarater reliability was reported in 12 of the 18 studies using 7 different measurement techniques (low flexion, extension with internal rotation, HA, internal rotation, diagnostic ultrasound, scapular-plane abduction, and myotonometer) and various instruments to quantify PST even among similar techniques. The overall ICC values ranged from 0.40 to 0.99 among all measurement techniques.[4,20] The low-flexion technique was evaluated in 2 studies, both using a digital inclinometer, with ICC values ranging from 0.92[3] to 0.96.[7] Extension with internal rotation was examined by Dashottar et al[7] using a digital inclinometer and reporting ICC values of 0.93 and 0.98 after fatigue and before fatigue, respectively. HA (with the scapula stabilized) was examined in 9 studies using both side-lying and supine techniques with a variety of measurement instruments (carpenter square, goniometer, digital inclinometer, bubble inclinometer, and smartphone).[4,7,18-22,30,37] The overall ICC values ranged from 0.40[4] to 0.97[7,21] for those techniques using HA. Diagnostic ultrasound was used to measure posterior capsule thickness in 1 study, which yielded an ICC value of 0.95.[16] Dashottar et al were the only authors to use scapular-plane abduction and reported ICC values of 0.93 and 0.96 after fatigue and at 24 hours, respectively. Finally, a myotonometer was investigated by 2 studies, reporting a range of ICC values from 0.33[17] to 0.97.[7]

The meta-analysis for intrarater reliability yielded a pooled Fisher z of 1.65 (95% confidence interval [CI], 1.49-1.81) (Supplementary Fig. S1, *A*). The corresponding summary ICC value was 0.93 (95% CI, 0.90-0.95), indicating good intrarater reliability. The homogeneity statistic (Q statistic) was 265.40 ($df = 37$, $P < .001$), indicating that significant heterogeneity in intrarater reliability was present between studies and the variability in the reliability values was greater than that expected by sampling error alone.

Inter-rater reliability

Inter-rater reliability was reported in 6 of the 18 studies using 3 different measurement techniques (HA, internal rotation, and posterior capsule thickness) and various instruments to quantify PST among similar techniques. Subjects in all studies were asymptomatic with the exception of those measured by Lin and Yang,[22] whose patients were identified as having stiff shoulders. The overall ICC values ranged from 0.69[28] to 0.98[16] among all measurement techniques. HA was the most used technique among studies reporting inter-rater reliability. Positions for measuring HA included both supine and side lying, with ICC values for each ranging from 0.79[21] to 0.94[28] and from 0.69[28] to 0.94,[21] respectively. Finally, a single study, by Ishigaki et al,[16] using diagnostic ultrasound to measure posterior capsule thickness among asymptomatic collegiate baseball players yielded an ICC of 0.98.

The meta-analysis for inter-rater reliability yielded a pooled Fisher z of 1.43 (95% CI, 1.10-1.77) (Supplementary Fig. S1, *B*). The corresponding summary ICC value was 0.89 (95% CI, 0.80-0.94), indicating good inter-rater reliability. The homogeneity statistic (Q statistic) was 67.05 ($df = 7$, $P < .001$), indicating that significant heterogeneity in inter-rater reliability was present between studies and the variability in the reliability values was greater than that expected by sampling error alone.

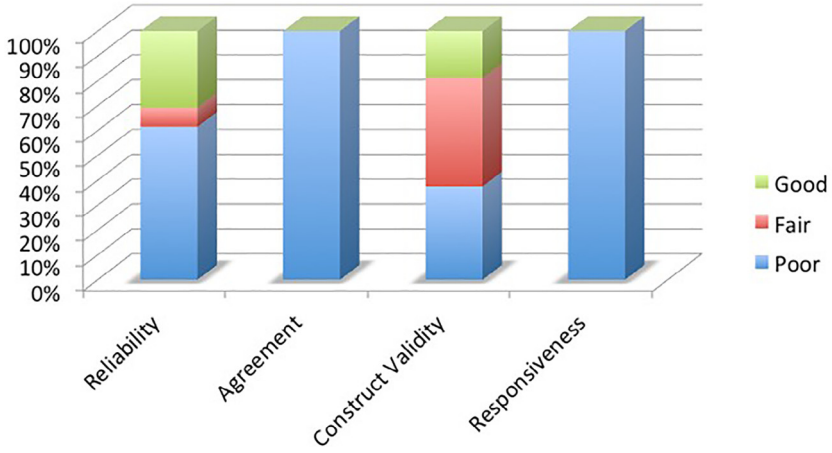

Figure 1 The methodologic quality (reliability, agreement, construct validity, responsiveness) of most studies was poor, with the exception of those examining construct validity, in which the quality was most often fair.

Validity

No studies examined criterion validity. Construct validity was reported in 10 of the 18 studies, using HA (8 studies) and posterior capsule thickness via diagnostic ultrasound (2 studies). All studies used internal rotation for convergent validity and external rotation for divergent validity. Convergent validity values among all studies using HA showed statistical significance in those reporting it, with the exception of the study of Myers et al,[28] who reported $P = .295$ using an anthropometer within an asymptomatic population. Divergent validity values among all studies using HA showed no statistical significance. The 2 studies using diagnostic ultrasound to measure posterior capsule thickness reported statistical significance with respect to convergent validity.[16,34] However, in 1 of the 2 aforementioned studies, Ishigaki et al[16] reported $P = .343$ with regard to divergent validity.

Methodologic quality using COSMIN scale

The methodologic quality of most studies was poor, with the exception of those examining construct validity, in which the quality was most often fair (Fig. 1). All of the studies rated as good examined reliability.

Best-evidence synthesis combining COSMIN and Terwee scales

The best-evidence synthesis was created by combining the results of the review of the quality of the measurement properties and the methodologic quality of the articles. It was at this stage that only articles receiving a fair or better rating were considered and that articles with fewer than 30 subjects were rated as having evidence that was "unknown."[1] These 2 steps were taken because all studies have been reviewed and reported on in the separate methodologic quality and quality–of–measurement properties sections of this article and because recommendations in a best-evidence synthesis should be of higher quality. Adhering to these tenets eliminated low flexion, extension with internal rotation, internal rotation, diagnostic ultrasound, scapular-plane abduction with low flexion, and myotonometer; this left only HA in both the supine and side-lying positions available for synthesis. The grading in the best-evidence synthesis was as follows:

- Strong: multiple studies of good methodologic rating or at least 1 study of excellent methodology
- Moderate: multiple fair methodologic studies or 1 study of good methodology
- Limited: 1 study of fair methodologic quality
- Conflicting: contradictory findings
- Unknown: investigated in studies of exclusively poor methodology or not investigated in any study

HA was the only method remaining that met all the aforementioned tenets put forth for the synthesis of best evidence (Table I). The studies remaining for best-evidence synthesis provided a grade of unknown evidence for agreement, criterion validity, and responsiveness. Much of this was because of the low sample sizes of many of the studies. No studies remaining for best-evidence synthesis regarding HA investigated criterion validity or responsiveness. Reliability within the synthesis of best evidence was conflicting. Hypothesis testing and/or construct validity received a grade of moderate given the positive methodologic property scores and the fair methodologic quality rating of those studies with adequate sample sizes.

Discussion

The purpose of this study was to perform a systematic review of the literature with meta-analysis to investigate the reliability and validity of techniques used to quantify PST. We sought to evaluate the current literature by investigating the

Table I Synthesis of best evidence by method used to quantify posterior shoulder tightness

Quantification method	Grade
Horizontal adduction in supine position	
Reliability	Conflicting
Agreement	Unknown
Hypothesis testing	Moderate
Criterion validity	Unknown
Responsiveness	Unknown
Horizontal adduction in side-lying position	
Reliability	Conflicting
Agreement	Unknown
Hypothesis testing	Moderate
Criterion validity	Unknown
Responsiveness	Unknown

methodologic quality of each study as well as available components through meta-analysis. We believe that this combined assessment approach allows for a more critical appraisal of each study selected.

The results of the best-evidence synthesis were a bit surprising to us. Of the several methods purported by the literature to quantify PST, only HA met the necessary criteria to be included in the best-evidence synthesis. Reliability, as examined for the HA method of quantifying PST from the synthesis of best evidence, revealed conflicting results. Furthermore, evidence to support many of the necessary clinimetric properties of a measurement method were nonexistent (criterion validity and responsiveness) both within those studies in the best-evidence synthesis and within those not included. Given the overall clinical and research use of the various methods to quantify PST, the lack of methodologic quality that exists to support their efficacy is somewhat alarming.

The results of the intrarater reliability meta-analysis revealed that there is excellent intrarater reliability (summary ICC, 0.93; 95% CI, 0.90-0.95) among the methods used to quantify PST. Although the 95% CI of the summary ICC value falls into the upper range of the good interpretation category, there was substantial variation in the ICC values reported. One study reported 3 ICC values that are in the moderate interpretation category and 1 ICC value in the fair interpretation category.[4]

The authors of the aforementioned study suggested that the inclusion of symptomatic subjects, the use of less skillful raters, and the long interval between test-retest assessments (8-12 weeks) all contributed to the low intrarater reliability values.[4] Another study reported an intrarater ICC value in the moderate interpretation category, which may be the result of the rater having only 1 year of clinical practice in orthopedic physical therapy.[21] These findings suggest that the level of experience of the raters, whether the subject is symptomatic, and the interval between measurements could play a role in the intrarater reliability of the current methods used to quantify PST.

The results of the inter-rater reliability analysis revealed that there is good inter-rater reliability (summary ICC, 0.89; 95% CI, 0.80-0.94) when using the current methods to quantify PST. In addition to the 95% CI of the summary ICC value not crossing into the moderate interpretation category, all but 1 ICC value reported by the included studies fall into the good interpretation category.[28]

One study reported a moderate inter-rater ICC value (0.69) with a low level of inter-rater reliability when the side-lying position was used, which may be due to the difference in stature of the 2 testers (height of 197 cm and mass of 92 kg vs height of 163 cm and mass of 61 kg). By contrast, the inter-rater reliability was good (ICC value, 0.94) when the supine position was used by the same 2 raters, suggesting that the supine position may be less susceptible to testers' statures and sizes.[28]

HA with the supine position or side-lying position was the most common measurement technique used in the included studies to assess intrarater reliability, as well as inter-rater reliability. When the summary ICC values calculated for measurement techniques are further investigated, HA in the supine position appears to show better intrarater and inter-rater reliability than HA in the side-lying position. More studies should be conducted to assess the reliability of other measurement techniques.

A majority of the studies focused on HA as the means to quantify PST. However, it is critical to note that every study that has investigated the validity of techniques used to quantify PST has used internal rotation as the reference criterion for convergent validity and external rotation for divergent validity. There is an inherent flaw in the assumption that glenohumeral joint internal rotation and HA are restricted by the same elements of soft tissue or other limiting structures. There has been evidence to support the theory that the restriction in HA and internal rotation at the glenohumeral joint is a result of restrictions of different structures. Tyler et al[35] used a multimodal treatment approach and showed an increase in HA range of motion and a reduction in symptoms among symptomatic individuals; however, they reported no significant gains in internal rotation. Understanding that limitations in HA and internal rotation may be a result of restrictions found within different structures not only helps to guide examination in this area but also is imperative when developing intervention strategies to address these limitations. In addition, when considering the overhead athlete, there is the potential for the presence (in varying degrees) or absence of asymmetrical humeral torsion that is in no way taken into account by quantifying either glenohumeral HA or internal rotation. In some regard, there is an attempt to consider humeral torsion when looking at total range of motion (TROM), which is the combined motion of internal and external rotation, of the glenohumeral joint in the overhead athlete and comparing it with the nondominant shoulder. However, there is no definitive way to determine by measurement of TROM alone whether the decrease in range of motion is a result of soft-tissue restrictions or osseous changes as a result

of adaptive training and to what degree each may be contributing.

Directions for future research

It is our recommendation that a combination of glenohumeral joint HA and internal rotation be considered when one is quantifying PST, as restrictions for each of these elements may arise from different structures. Understanding which independent elements may be contributing to a loss of motion in these areas will allow for a more comprehensive assessment of what should be considered as PST moving forward. In addition, when a clinician is quantifying PST in the overhead athlete, measurements of glenohumeral joint HA, internal rotation, TROM, and when possible, the amount of humeral torsion through diagnostic imaging should be considered. Each of these elements may contribute in varying amounts and must be considered in a comprehensive approach both when attempting to quantify PST and when developing an intervention program attempting to address these limitations. Finally, further research of higher quality needs to be performed to better establish the clinimetric properties for quantifying PST.

Limitations

There are several limitations to this study including the inclusion of only those studies published in the English language. We also did not include a search of the gray literature as we believed that, within the area being investigated, this type of search would not lend any further results beyond our current search.

Conclusion

Current techniques used to quantify PST appear to have acceptable inter-rater and intrarater reliability, with HA showing the greatest overall methodologic quality relative to other methods. However, nearly all studies focused on HA as the sole means for quantifying PST. It is imperative that clinicians begin to collectively consider multiple components that together comprise PST, including HA, internal rotation, TROM, and humeral torsion when appropriate. This comprehensive approach will allow for a more thorough assessment of PST and, in turn, a more specific plan for interventions.

Acknowledgments

The authors thank Leila Ledbetter.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2018.07.013.

References

1. 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. http://dx.doi.org/10.2522/ptj.20120210
2. Borenstein M, Hedges LV, Higgins J, Rothstein HR. Introduction to meta-analysis. West Sussex, UK: John Wiley & Sons; 2009.
3. Borstad JD, Dashottar A, Stoughton T. Validity and reliability of the Low Flexion measurement for posterior glenohumeral joint capsule tightness. Man Ther 2015;20:875-8. http://dx.doi.org/10.1016/j.math.2015.08.007
4. Borstad JD, Mathiowetz KM, Minday LE, Prabhu B, Christopherson DE, Ludewig PM. Clinical measurement of posterior shoulder flexibility. Man Ther 2007;12:386-9. http://dx.doi.org/10.1016/j.math.2006.07.014
5. Burkhart SS, Morgan CD, Kibler WB. The disabled throwing shoulder: spectrum of pathology Part I: pathoanatomy and biomechanics. Arthroscopy 2003;19:404-20. http://dx.doi.org/10.1053/jars.2003.50128
6. Burkhart SS, Morgan CD, Kibler WB. The disabled throwing shoulder: spectrum of pathology. Part II: evaluation and treatment of SLAP lesions in throwers. Arthroscopy 2003;19:531-9. http://dx.doi.org/10.1053/jars.2003.50139
7. Dashottar A, Costantini O, Borstad J. A comparison of range of motion change across four posterior shoulder tightness measurements after external rotator fatigue. Int J Sports Phys Ther 2014;9:498-508.
8. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
9. Downar JM, Sauers EL. Clinical measures of shoulder mobility in the professional baseball player. J Athl Train 2005;40:23-9.
10. Fisher RA. Statistical methods for research workers. Edinburgh: Oliver & Boxel; 1925.
11. Hedges LV, Olkin I. Statistical methods for meta-analysis. Orlando, FL: Academic Press; 1985.
12. Hegedus EJ, McDonough S, Bleakley C, Cook CE, Baxter GD. 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. http://dx.doi.org/10.1136/bjsports-2014-094094
13. Hibberd EE, Oyama S, Myers JB. Increase in humeral retrotorsion accounts for age-related increase in glenohumeral internal rotation deficit in youth and adolescent baseball players. Am J Sports Med 2014;42:851-8. http://dx.doi.org/10.1177/0363546513519325
14. Huberty DP, Schoolfield JD, Brady PC, Vadala AP, Arrigoni P, Burkhart SS. Incidence and treatment of postoperative stiffness following arthroscopic rotator cuff repair. Arthroscopy 2009;25:880-90. http://dx.doi.org/10.1016/j.arthro.2009.01.018
15. Hunter JE, Schmidt FL. Correcting for sources of artificial variation across studies. New York: Russell Sage Foundation; 1994.

16. Ishigaki T, Ishida T, Samukawa M, Saito H, Ezawa Y, Hirokawa M, et al. Does restriction of glenohumeral horizontal adduction reflect posterior capsule thickening of the throwing shoulder? J Phys Ther Sci 2015;27:1299-302. http://dx.doi.org/10.1589/jpts.27.1299
17. Kerins CM, Moore SD, Butterfield TA, McKeon PO, Uhl TL. Reliability of the myotonometer for assessment of posterior shoulder tightness. Int J Sports Phys Ther 2013;8:248-55.
18. Kolber MJ, Beekhuizen KS, Cheng MS, Hellman MA. Shoulder joint and muscle characteristics in the recreational weight training population. J Strength Cond Res 2009;23:148-57. http://dx.doi.org/10.1519/JSC.0b013e31818eafb4
19. Kolber MJ, Hanney WJ. The reliability, minimal detectable change and construct validity of a clinical measurement for identifying posterior shoulder tightness. N Am J Sports Phys Ther 2010;5:208-19.
20. Kolber MJ, Saltzman SB, Beekhuizen KS, Cheng MS. Reliability and minimal detectable change of inclinometric shoulder mobility measurements. Physiother Theory Pract 2009;25:572-81. http://dx.doi.org/10.3109/09593980802667995
21. Lim JY, Kim TH, Lee JS. Reliability of measuring the passive range of shoulder horizontal adduction using a smartphone in the supine versus the side-lying position. J Phys Ther Sci 2015;27:3119-22. http://dx.doi.org/10.1589/jpts.27.3119
22. Lin JJ, Yang JL. Reliability and validity of shoulder tightness measurement in patients with stiff shoulders. Man Ther 2006;11:146-52. http://dx.doi.org/10.1016/j.math.2005.05.002
23. Linsell L, Dawson J, Zondervan K, Rose P, Randall T, Fitzpatrick R, et al. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patterns of diagnosis and referral. Rheumatology (Oxford) 2006;45:215-21. http://dx.doi.org/10.1093/rheumatology/kei139
24. Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. Scand J Rheumatol 2004;33:73-81.
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264-9. http://dx.doi.org/10.1016/j.ijsu.2010.02.007
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006
27. Myers JB, Laudner KG, Pasquale MR, Bradley JP, Lephart SM. Glenohumeral range of motion deficits and posterior shoulder tightness in throwers with pathologic internal impingement. Am J Sports Med 2006;34:385-91. http://dx.doi.org/10.1177/0363546505281804
28. Myers JB, Oyama S, Wassinger CA, Ricci RD, Abt JP, Conley KM, et al. Reliability, precision, accuracy, and validity of posterior shoulder tightness assessment in overhead athletes. Am J Sports Med 2007;35:1922-30. http://dx.doi.org/10.1177/0363546507304142
29. Portney LG, Watkins MP. Foundations of clinical research: application to practice. Upper Saddle River, NJ: Pearson Prentice Hall; 2009.
30. Salamh PA, Corrao M, Hanney WJ, Kolber MJ. The reliability and validity of measurements designed to quantify posterior shoulder tightness. Phys Ther Rev 2011;16:347-55. http://dx.doi.org/10.1179/1743288X11Y.0000000038
31. Salamh PA, Kolber MJ. The reliability, minimal detectable change and construct validity of a clinical measurement for quantifying posterior shoulder tightness in the post-operative population. Int J Sports Phys Ther 2012;7:565-75.
32. Salamh PA, Kolber MJ, Hanney WJ. Effect of scapular stabilization during horizontal adduction stretching on passive internal rotation and posterior shoulder tightness in young women volleyball athletes: a randomized controlled trial. Arch Phys Med Rehabil 2015;96:349-56. http://dx.doi.org/10.1016/j.apmr.2014.09.038
33. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012
34. Thomas SJ, Swanik CB, Higginson JS, Kaminski TW, Swanik KA, Bartolozzi AR, et al. A bilateral comparison of posterior capsule thickness and its correlation with glenohumeral range of motion and scapular upward rotation in collegiate baseball players. J Shoulder Elbow Surg 2011;20:708-16. http://dx.doi.org/10.1016/j.jse.2010.08.031
35. Tyler TF, Nicholas SJ, Lee SJ, Mullaney M, McHugh MP. Correction of posterior shoulder tightness is associated with symptom resolution in patients with internal impingement. Am J Sports Med 2010;38:114-9. http://dx.doi.org/10.1177/0363546509346050
36. Tyler TF, Nicholas SJ, Roy T, Gleim GW. Quantification of posterior capsule tightness and motion loss in patients with shoulder impingement. Am J Sports Med 2000;28:668-73.
37. Tyler TF, Roy T, Nicholas SJ, Gleim GW. Reliability and validity of a new method of measuring posterior shoulder tightness. J Orthop Sports Phys Ther 1999;29:262-74.
38. Werner CML, Nyffeler RW, Jacob HAC, Gerber C. The effect of capsular tightening on humeral head translations. J Orthop Res 2004;22:194-201. http://dx.doi.org/10.1016/S0736-0266(03)00137-2
39. Witwer A, Sauers E. Clinical measures of shoulder mobility in college water-polo players. J Sport Rehab 2006;15:45-57. http://dx.doi.org/10.1123/jsr.15.1.45
40. Yang JL, Chen SY, Chang CW, Lin JJ. Quantification of shoulder tightness and associated shoulder kinematics and functional deficits in patients with stiff shoulders. Man Ther 2009;14:81-7. http://dx.doi.org/10.1016/j.math.2007.11.004