Journal of Athletic Training 2012:47(2):221–223

www.nata.org/jat

Psychometric Properties of Self-Report Concussion Scales and Checklists

Tamara C. Valovich McLeod, PhD, ATC, FNATA; Candace Leach, MS, ATC

Arizona School of Health Sciences, A.T. Still University, Department of Interdisciplinary Health Sciences, Mesa

Reference/Citation: Alla S, Sullivan SJ, Hale L, McCrory P. Self-report scales/checklists for the measurement of concussion symptoms: a systematic review. *Br J Sports Med*. 2009;43 (suppl 1):i3–i12.

Clinical Question: Which self-report symptom scales or checklists are psychometrically sound for clinical use to assess sport-related concussion?

Data Sources: Articles available in full text, published from the establishment of each database through December 2008, were identified from PubMed, Medline, CINAHL, Scopus, Web of Science, SPORTDiscus, PsycINFO, and AMED. Search terms included *brain concussion, signs or symptoms,* and *athletic injuries,* in combination with the AND Boolean operator, and were limited to studies published in English. The authors also hand searched the reference lists of retrieved articles. Additional searches of books, conference proceedings, theses, and Web sites of commercial scales were done to provide additional information about the psychometric properties and development for those scales when needed in articles meeting the inclusion criteria.

Study Selection: Articles were included if they identified all the items on the scale and the article was either an original research report describing the use of scales in the evaluation of concussion symptoms or a review article that discussed the use or development of concussion symptom scales. Only articles published in English and available in full text were included.

Data Extraction: From each study, the following information was extracted by the primary author using a standardized protocol: study design, publication year, participant characteristics, reliability of the scale, and details of the scale or checklist, including name, number of items, time of measurement, format, mode of report, data analysis, scoring, and psychometric properties. A quality assessment of included studies was done using 16 items from the Downs and Black checklist[1] and assessed reporting, internal validity, and external validity.

Main Results: The initial database search identified 421 articles. After 131 duplicate articles were removed, 290 articles remained and were added to 17 articles found during the hand search, for a total of 307 articles; of those, 295 were available in full text. Sixty articles met the inclusion criteria and were used in the systematic review. The quality of the included studies ranged from 9 to 15 points out of a maximum quality score of 17. The included articles were published between 1995 and 2008 and included a collective total of 5864 concussed athletes and 5032 nonconcussed controls, most of whom participated in American football. The majority of the studies were descriptive studies monitoring the resolution of concussive self-report symptoms compared with either a preseason baseline or healthy control group, with a smaller number of studies (n = 8) investigating the development of a scale.

The authors initially identified 20 scales that were used among the 60 included articles. Further review revealed that 14 scales were variations of the Pittsburgh Steelers postconcussion scale (the Post-Concussion Scale, Post-Concussion Scale: Revised, Post-Concussion Scale: ImPACT, Post-Concussion Symptom Scale: Vienna, Graded Symptom Checklist [GSC], Head Injury Scale, McGill ACE Post-Concussion Symptoms Scale, and CogState Sport Symptom Checklist), narrowing down to 6 core scales, which the authors discussed further. The 6 core scales were the Pittsburgh Steelers Post-Concussion Scale (17 items), Post-Concussion Symptom Assessment Questionnaire (10 items), Concussion Resolution Index postconcussion questionnaire (15 items), Signs and Symptoms Checklist (34 items), Sport Concussion Assessment Tool (SCAT) postconcussion symptom scale (25 items), and Concussion Symptom Inventory (12 items). Each of the 6 core scales includes symptoms associated with sport-related concussion; however, the number of items on each scale varied. A 7-point Likert scale was used on most scales, with a smaller number using a dichotomous (*yes/no*) classification.

Only 7 of the 20 scales had published psychometric properties, and only 1 scale, the Concussion Symptom Inventory, was empirically driven (Rasch analysis), with development of the scale occurring before its clinical use. Internal consistency (Cronbach α) was reported for the Post-Concussion Scale (.87), Post-Concussion Scale: ImPACT 22-item (.88–.94), Head Injury Scale 9-item (.78), and Head Injury Scale 16-item (.84). Test-retest reliability has been reported only for the Post-Concussion Scale (Spearman $r = .55$) and the Post-Concussion Scale: ImPACT 21-item (Pearson $r = .65$). With respect to validity, the SCAT postconcussion scale has demonstrated face and content validity, the Post-Concussion Scale: ImPACT 22-item and Head Injury Scale 9-item have reported construct validity, and the Head Injury Scale 9-item and 16-item have published factorial validity.

Sensitivity and specificity have been reported only with the GSC (0.89 and 1.0, respectively) and the Post-Concussion Scale: ImPACT 21-item when combined with the neurocognitive component of ImPACT (0.819 and 0.849, respectively). Meaningful change scores were reported for the Post-Concussion Scale (14.8 points), Post-Concussion Scale: ImPACT 22-item (6.8 points), and Post-Concussion Scale: ImPACT 21-item (standard error of the difference = 7.17; 80% confidence interval = 9.18).

Conclusions: Numerous scales exist for measuring the number and severity of concussion-related symptoms, with most evolving from the neuropsychology literature pertaining to head-injured populations. However, very few of these were created in a systematic manner that follows scale development processes and have published psychometric properties. Clinicians need to understand these limitations when choosing

and using a symptom scale for inclusion in a concussion assessment battery. Future authors should assess the underlying constructs and measurement properties of currently available scales and use the ever-increasing prospective data pools of concussed athlete information to develop scales following appropriate, systematic processes.

Key Words: mild traumatic brain injuries, evaluation, reliability, validity, sensitivity, specificity

COMMENTARY

Identifying and understanding the symptoms of sport-related concussion are important for a number of reasons, including the diagnosis, evaluation, and management of the injury. Clinical symptoms reported by athletes are often the primary grounds on which a concussion is initially diagnosed. Once recognized, the evaluation of self-report symptoms has become a common means of concussion assessment, with approximately 85% of athletic trainers (ATs) using some form of symptom assessment to evaluate concussion.[2] Symptom evaluation is also used by 80% of ATs to assist with return-to-play (RTP) decisions, with nearly 15% of ATs indicating that the assessment of symptoms is their primary method for aiding RTP decisions, second only to the clinical examination (60%).[2] These rates are not surprising considering that symptom scales are easily available and inexpensive to use. Furthermore, symptoms have demonstrated the greatest effect after injury,[3] meaning that self-report symptom scores increased to a greater extent than neurocognitive or balance test scores during the initial evaluation and within 14-days postconcussion follow-up. Additionally, sport concussion position and consensus statements[4,5] recommend the evaluation of self-report symptoms as a component of a comprehensive concussion assessment and management strategy. One common recommendation regarding concussion is that no athlete begin an RTP progression until he or she is asymptomatic, further highlighting the need for clinicians to use some form of self-report symptom assessment.

The use of self-report symptom scales has been criticized as unreliable because athletes subjectively report their symptoms and may be motivated to not report symptoms in order to hasten RTP.[6] However, an even bigger criticism is that highlighted by Alla et al,[7] who noted that most scales have not been psychometrically validated before their use in clinical practice. Having sound psychometric properties for the symptom scales is important because these tools are often used to track symptom resolution and aid the clinician in making the important and complex RTP decision. When used for RTP decision making, the symptom scale is administered serially at multiple time points during the recovery period. Therefore, it is important for clinicians to be able to conclude that positive changes in patient self-report symptom scores are the result of symptom recovery rather than low instrument stability. However, literature evaluating the psychometric properties of self-report scales is limited. In this systematic review, only 7 of the 20 scales had any published psychometric properties, and none of the symptom scales were supported by a complete set of published psychometric properties, including item selection, reliability, validity, sensitivity, specificity, and change scores.

A valid scale is one that lists items (symptoms) that are important to the clinician when evaluating a concussed athlete and provides information that may be useful in the management or RTP decision-making process. Although not every scale has been formally evaluated for validity, most that are used clinically do include a wide variety of concussion-related symptoms. Clinicians should choose scales that contain items meaningful to their clinical practice until additional studies of validity are published. Of the studies evaluated by Alla et al,[7] only the Post-Concussion Scale: ImPACT-22 item (construct), SCAT Post-Concussion Symptom Scale (face, content), Head Injury Scale 16-item (factorial), and Head Injury Scale 9-item (factorial, construct) have been evaluated for validity.

Multiple measures of reliability are important to understand when choosing and using a scale. Internal consistency, or the ability of the instrument to measure the same construct (eg, concussion-related symptoms), and test-retest reliability (stability over time) are important psychometric variables. Good test-retest reliability increases the likelihood that changes over time are the result of symptom resolution and not variability due to error or chance. Furthermore, scales must be reliable to ensure validity and change scores (eg, reliable change index).[8] Numerous scales have published internal consistency, including the Post-Concussion Scale, Post-Concussion Scale: ImPACT 22-item, and Head Injury Scale 16-item and 9-item.[7] The Post-Concussion Symptom Scale: ImPACT 21-item has reported moderate test-retest reliability values of $r = 0.65$ over a 5.8-day test-retest interval.[7]

Finally, the sensitivity and specificity of the instrument are important to understand when looking to classify an athlete as concussed. *Sensitivity* is the ability of the instrument to correctly identify athletes with a concussion, whereas *specificity* is the instrument's ability to correctly identify those without the condition (eg, healthy controls).[9] Both of these properties are important to differentiate concussed athletes from healthy controls. Instruments with low sensitivity or specificity may increase the chance for false positives or false negatives. The Graded Symptom Checklist (GSC) 17-item and the Post-Concussion Scale: ImPACT 21-item have published sensitivity and specificity values; however, for the latter, the values were combined with the cognitive scores from ImPACT.[7]

With the widespread use of these scales in clinical practice, it is not feasible to discontinue their use for lack of knowledge about their psychometric properties. However, clinicians should be aware of the potential shortfalls of a scale, and researchers should continue to conduct studies that assess these properties.

The process of scale selection is also important, and understanding psychometric properties is an essential aspect of that process. Alla et al[7] briefly described a variety of similarities and differences among 20 symptom scales frequently used in concussion assessment. The most common type uses a Likert-style grading scale as the rating system. Likert scales offer a more precise measurement of symptom variability and are better in noting small changes in status over time as opposed to a dichotomous scale with simple *yes*/*no* responses. Although in this article we have used the term *scale* to describe both formats, the choice of a scale (Likert style) or checklist (*yes*/*no*) depends on what the clinician intends to gain from the instrument's use. If the question is really whether or not an athlete is reporting symptoms, a dichotomous scale may suffice. However, the *yes*/*no* responses will not allow the clinician to interpret small, but perhaps meaningful, changes in the athlete's symptom severity.

In contrast, if the clinician is interested in objectifying the symptom data by calculating a total symptom score, or if tracking the changes in specific symptom severity is important, then a symptom scale is recommended.

The GSC is an example of a Likert-style questionnaire recommended for use by ATs in the "National Athletic Trainers' Association Position Statement: Management of Sport-Related Concussion."[5] Sensitivity (.89) and specificity (1.0) values for the GSC have been published, with sensitivity values being lower when the scale is used to evaluate concussions further from the time of injury. The decrease in sensitivity values over time is thought to be the result of an athlete's progressively improving symptoms, which reduce variability on subsequent assessment days.[10] The tool allows clinicians to calculate a total symptom score (sum total of all response values) and the number of symptoms endorsed. Clinicians can then track changes in the total symptom score and individual symptoms as the athlete recovers from the concussion.

Although Alla et al[7] presented some important questions about symptom scale use in concussion management, their systematic review has limitations. The authors included only articles to which they had full-text access. Inclusion of only articles to which they had full access may bias their results because they excluded 12 articles that were not available in full text. Information from these studies may be important or might have provided additional details for this review. Also, the authors chose to include review articles that discussed symptom scales. These 5 review articles were not original research studies but rather narrative or clinical reviews. They should not be judged by the same criteria as original research, because they would not include concussed or control participants, describe original results, or contain original analyses of psychometric properties. The systematic review also lacked important details about how the authors identified the 6 core scales. Finally, although they stated that the quality scores of included studies ranged from 9 to 15 points, it would have been helpful if the authors included the quality scores for each study in their systematic review and the criteria for determining the quality score.

Although there are limitations, this systematic review does provide insight into the relationship between research and clinical practice, especially with patient populations. It would be ideal for clinicians to include only psychometrically sound scales in clinical practice, but the speed at which concussion evaluation and management have changed in the past decade has not allowed that to occur. Researchers have developed scales and used them in studies of concussed athletes without understanding all the underlying psychometrics. In the future, researchers should continue to evaluate existing databases to determine the measurement properties of existing symptom scales and conduct additional studies of scale measurement properties. Clinicians should continue to use a multifactorial approach for concussion assessment that includes not only an evaluation of symptoms but also mental status, cognitive function, and postural stability.

REFERENCES

1. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377–384.
2. Notebaert AJ, Guskiewicz KM. Current trends in athletic training practice for concussion assessment and management. *J Athl Train*. 2005;40(4):320–325.
3. Broglio SP, Puetz TW. The effect of sport concussion on neurocognitive function, self-report symptoms and postural control: a meta-analysis. *Sports Med*. 2008;38(1):53–67.
4. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on Concussion in Sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *J Sci Med Sport*. 2009;12(3):340–351.
5. Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association position statement: management of sport-related concussion. *J Athl Train*. 2004;39(3):280–297.
6. Kaut KP, DePompei R, Kerr J, Congeni J. Reports of head injury and symptom knowledge among college athletes: implications for assessment and educational intervention. *Clin J Sport Med*. 2003;13(4):213–221.
7. Alla S, Sullivan SJ, Hale L, McCrory P. Self-report scales/checklists for the measurement of concussion symptoms: a systematic review. *Br J Sports Med*. 2009;43(suppl 1):i3–i12.
8. Gioia GA, Schneider JC, Vaughan CG, Isquith PK. Which symptom assessments and approaches are uniquely appropriate for paediatric concussion? *Br J Sports Med*. 2009;43(suppl 1):i13–i22.
9. Loong TW. Understanding sensitivity and specificity with the right side of the brain. *BMJ*. 2003;327(7417):716–719.
10. McCrea M, Barr WB, Guskiewicz K, et al. Standard regression-based methods for measuring recovery after sport-related concussion. *J Int Neuropsychol Soc*. 2005;11(1):58–69.

Address correspondence to Tamara C. Valovich McLeod, PhD, ATC, FNATA, Arizona School of Health Sciences, A.T. Still University, Department of Interdisciplinary Health Sciences, 5850 East Still Circle, Mesa, AZ 85206. Address e-mail to tmcleod@atsu.edu.