Core Stability: Inter- and Intraobserver Reliability of 6 Clinical Tests

Adam Weir, MBBS,* Jennifer Darby,* Han Inklaar, MD,† Bart Koes, PhD,‡ Erik Bakker, PhD,§ and Johannes L. Tol, MD, PhD*

Objective: Core stability is a complex concept within sports medicine and is thought to play a role in sports injuries. There is a lack of reliable and valid clinical tests for core stability. The inter- and intraobserver reliability of 6 tests commonly used to assess core stability was determined.

Design: A video of the tests was shown to 6 observers. A second observation took place 5 weeks later with the same observers.

Setting: Sports medicine department of a hospital.

Participants: Forty male athletes.

Assessment of Variables: Core stability was rated as poor, moderate, good, or excellent by each observer for each of the 6 tests.

Main Outcome Measures: Inter- and intraobserver reliability.

Results: The mean score of all tests was 13.4% poor, 33.3% moderate, 40.1% good, and 13.2% excellent. The intraclass correlation coefficients (ICCs 2,1) for the interobserver reliability for frontal, sagittal, and transverse plane evaluation were 0.09, 0.32, and 0.51, respectively. The ICCs for the unilateral squat, the lateral step-down, and the bridge were 0.41, 0.39, and 0.36, respectively. The ICCs for the intraobserver reliability for frontal, sagittal, and transverse plane evaluation were 0.31, 0.40, and 0.55, respectively. The ICCs for the unilateral squat, the lateral step-down, and the bridge were 0.55, 0.49, and 0.21, respectively.

Conclusions: The 6 clinical core stability tests are not reliable when a 4-point visual scoring assessment is used. Future research on movement evaluation should be focused on more specific rating methods and training for the observers.

Key Words: functional testing, clinical agreement, dynamic movement analysis, lumbopelvic region, physical therapy

(Clin J Sport Med 2010;20:34–38)

INTRODUCTION

Core stability is a complex and very popular concept within sports medicine. One of the most common definitions was proposed by Kibler et al[1]: "The ability to control the position and motion of the trunk over the pelvis and to allow optimum production, transfer, and control of force and motion to the terminal segment in integrated athletic activities."

Although core stability is thought to play a crucial role in sports medicine, there are no widely accepted reliable tests for testing core stability in the clinic.[2] Chmielewski et al[3] were the first to study the observer reliability of 2 clinical tests. In this study, 2 testing methods of functional tasks for the lower extremity were evaluated and levels of agreement were descriptively compared. The 2 functional tasks were the lateral step-down and the unilateral squat. Both inter- and intraobserver reliability were low.

Kibler et al[1] described a three-plane testing model. Three core stability tests were described. A recent comprehensive review on core stability stated that the use of these tests could give useful information, although it was noted that the reliability and validity had not been studied.[2] The aim of this study was to investigate the inter- and intraobserver reliability of 6 clinical core stability tests described and recommended in the literature.

MATERIALS AND METHODS

Subjects

The reliability of 6 clinical tests was assessed in 40 male volunteers. To avoid bias based on kinematical differences between men and women,[4,5] only men were included. Male subjects (>18 years of age) were eligible for inclusion. All subjects were recruited in an outpatient sports medicine department in a large district general hospital. When they attended for preparticipation screening examinations, the subjects were informed about the aim and background of the study. After obtaining their written informed consent, a researcher instructed the subjects in detail on performing the core stability tests. The local medical ethics committee approved the study protocol.

Subjects were excluded if they reported pain on performing the trial of the 6 clinical tests from the testing protocol.

Testing Protocol

The 6 tests are described and recommended in the literature and are commonly used in clinical decision making and patients' follow-up.[1,3] All of the tests have a good written description of how they should be performed (and observed),

and all of these tests are thought to be related to core stability. All observers met 4 months before the start of the first observation to agree on which tests would be used so that they could practice them in a clinical setting to allow familiarization.

The tests were all performed in single-leg stance (right leg) to be sufficiently demanding for the subjects. All subjects were given verbal instructions on how to perform the test, followed by a visual demonstration. Subjects were allowed a first trial of 6 repetitions. Verbal feedback was given if the test was performed incorrectly. The subjects then performed the definite trial. Subjects all wore the same clothing during the video recordings to increase uniformity. With each test, 2 trials were performed that consisted of 6 repetitions.

The Tests

Unilateral Squat

The starting position for the unilateral squat was standing on the test leg with the hip and knee in a neutral anatomical position. The trunk was upright, without rotation or lateral flexion, and the contralateral leg was positioned with the hip in neutral position and the knee in 90° of flexion. Subjects moved at a self-selected pace into a squat position and then returned to the starting position (Figure 1).

Lateral Step-Down

For the lateral step-down, subjects stood on the test leg, which was positioned on the edge of an adjustable step, with

the hip and knee in a neutral anatomical position. The trunk was upright without rotation or lateral flexion, the iliac crests were level, and the contralateral leg was unsupported, with the hip in a slightly flexed position and the knee extended. Subjects lowered themselves at a self-selected pace until the contralateral heel contacted the ground and then returned to the starting position (Figure 2). Three plane core tests were done with the subjects standing at 8 cm from the wall.
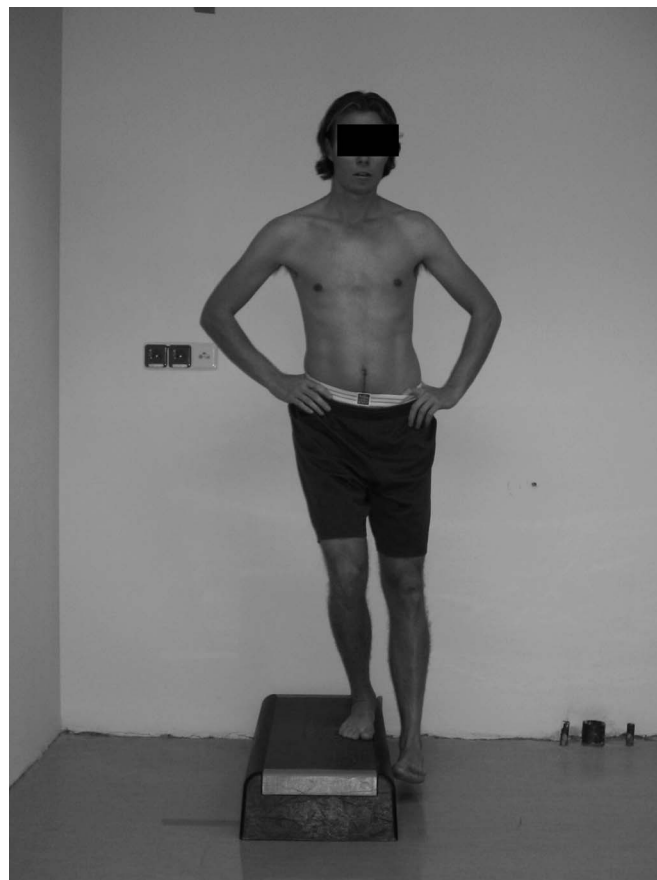
Frontal Plane Testing

The subjects stood with one side of the body toward the wall and with their shoulder 8 cm away from the wall, while standing on the inside leg, they were asked to lightly touch the wall with their shoulder. The head and pelvis were kept in the neutral position (Figure 3).

Sagittal Plane Testing

The subjects stood on one leg with their back toward the wall. The shoulders were 8 cm from the wall. They were asked to slowly move their body backward and lightly touch the wall with their head. The head and pelvis were kept in the neutral position (Figure 4).

Transverse Plane Testing

The subjects stood with their back toward the wall with their shoulders 8 cm away, in a single-leg position, and

FIGURE 1. The unilateral squat.

FIGURE 2. The lateral step-down.

FIGURE 3. Frontal plane testing.

FIGURE 4. Sagittal plane testing.

alternately lightly touched one shoulder and then the other against the wall. The head and pelvis were kept in the neutral position (Figure 5).

The Bridge

The bridge was performed with the body in horizontal prone position, supported by the underarms, with the arms directly under the shoulders and the toes of the feet. A straight line from head to toe had to be formed and maintained for 10 seconds (Figure 6).

The tests were recorded with a digital camcorder (JVC Digital Handycam GR-DVL167EG; JVC, Japan). Both trials were recorded, and the second trial was used for the observation.

Observation

There were 6 experienced observers: 4 experienced sport physicians who worked with athletes at an international level and 2 experienced sport physical therapists. Before the observation, they received an instruction on scoring the test performance. All the observers used a number of the tests in their current clinical practice when assessing core stability. The criteria for scoring the tests are shown in Table 1.

For each separate test, all subjects were subsequently scored according to the 4-point scale.

The observers scored the recordings of all 40 subjects in a random order for each single test. After they had evaluated

and scored 1 test, they were shown the recordings for all 40 subjects in a different random order for the second test.

The second observation was carried out in the same manner as the first. The time window between the 2 observations was 5 weeks.

Statistical Analysis

Data for each clinical test were analyzed separately. The statistical analysis performed was done by a 2-way random model to calculate the interobserver reliability for general use, not only to investigate the observer reliability between colleagues. For the inter- and intraobserver reliability, the intraclass correlation coefficient (ICC 2,1) was calculated. Frequency ratings were calculated for all the tests and all the observers per scoring category. Data analysis was done using SPSS 15.0 (SPSS Inc, Chicago, Illinois).

RESULTS

Forty subjects were included. The mean age of the subjects was 25.4 years (range, 18-44 years). The average height was 182.0 cm (SD, 7.3 cm), and the mean weight was 74.9 kg (SD, 12.1). The percentage of subjects who preferred to stand on the left leg was 75.7%. The mean weekly sports participation time was 7.9 hours per week (SD, 4.8). Thirty-one subjects (86%) were involved in soccer.

The mean score of all tests was 13.4% poor, 33.3% moderate, 40.1% good, and 13.2% excellent.

FIGURE 5. Transverse plane testing.

Percent agreement of all observers between the 2 observations was 0.46, 0.51, and 0.58 for the frontal, sagittal, and transverse plane tests, respectively, and 0.49, 0.47, and 0.53 for the unilateral squat, the lateral step-down, and the bridge, respectively.

The ICCs for the interobserver are shown in Table 2. The results of the intraobserver reliability are shown in Table 3.

FIGURE 6. The bridge.

TABLE 1. Score and Criteria for Core Stability Tests[3]

Score	Criteria
Excellent	No deviation from neutral alignment
Good	A small magnitude* or barely observable movement out of a neutral position and/or low frequency of segmental oscillation†
Moderate	A moderate or marked movement out of a neutral position and/or moderate-frequency segmental oscillation
Poor	Excessive or severe magnitude of movement out of a neutral position and/or high-frequency segment oscillation

*A single movement out of the neutral alignment.
†Multiple movements out of the neutral alignment.

The data were also analyzed with the score dichotomized (poor and moderate compared with good and excellent), which did not lead to an improved intra- and interobserver reliability.

DISCUSSION

This study showed a poor inter- and intraobserver reliability of the 6 clinical core stability tests when assessed with a 4-point visual evaluation score. The 6 tests examined in this study are widely used and have been recommended in the literature as being suitable to assess core stability.[1,2] The results of this study indicate that the use of these tests in clinical practice should be questioned.

Other investigators have also shown poor reliability of clinical tests used for core stability. Chmielewski et al[3] examined inter- and intraobserver reliability for the lateral step-down and the unilateral squat tests. In their study, 25 uninjured subjects were scored, on 2 occasions 5 weeks apart, using a specific and a general scoring method. In the specific method, trunk, pelvis, and hip were scored separately. The general method used a 3-point scoring system similar to that used in this study. The interobserver reliability using both the general method (weighted kappa, 0-0.55) and the specific method (weighted kappa, 0.23-0.53) was poor. The intraobserver reliability for the unilateral squat and the lateral step-down was also poor (0.13-0.68). The intraobserver reliability was better when the specific scoring system was used (0.38-0.68) when compared with the general scoring method (0.13-0.50). The slightly better results of this study should be interpreted with caution as a weighted kappa was used to express the intraobserver reliability, which can result in a better outcome.

TABLE 2. ICC (2,1) for Interobserver Reliability of 6 Tests Rated by 6 Observers With a 4-Point Scale

Test	Interobserver Reliability (ICC 2,1)	95% Confidence Interval
Unilateral squat	0.41	0.26–0.58
Lateral step-down	0.39	0.23–0.57
Frontal plane evaluation	0.09	0.01–0.21
Sagittal plane evaluation	0.32	0.19–0.49
Transverse plane evaluation	0.51	0.35–0.66
Bridge	0.36	0.22–0.53

ICC, intraclass correlation coefficient.

TABLE 3. ICC for Intraobserver Reliability of 6 Tests Rated by 6 Observers With a 4-Point Scale

Test	Intraobserver Reliability (ICC 2,1)	95% Confidence Interval
Unilateral squat	0.55	0.45–0.64
Lateral step-down	0.49	0.39–0.59
Frontal plane evaluation	0.31	0.17–0.43
Sagittal plane evaluation	0.40	0.29–0.51
Transverse plane evaluation	0.55	0.46–0.64
Bridge	0.21	0.07–0.35

ICC, intraclass correlation coefficient.

Piva et al[6] investigated the interobserver reliability for movement quality assessment during a lateral step-down. Thirty patients with patellofemoral pain were scored by 4 observers. A special rating system was created for this study. Five rating criteria included trunk, pelvis, and knee position; use of arms for balance; and the loss of balance. Each criterion was scored dichotomously, except for knee position, for which severity in deviation was rated based on anatomical reference points. Total scores were then categorized into 3 groups. A kappa coefficient of 0.67 was reported for the interobserver reliability.

At present, there are unfortunately no other studies available on clinical core stability tests. There are at present no reliable clinical tests with which core stability can be assessed.

Other studies have looked at the clinical assessment of movement patterns in other areas. Hayes et al[7] investigated the reliability of 5 methods for assessing shoulder range of motion in 8 patients with shoulder complaints. A visual estimation of passive range of motion was done using 3 static tests and 2 dynamic tests. The tests were scored by 4 observers. For the intraobserver reliability, only 1 observer was used. The time between the observations was within 48 hours, and 9 patients were included. The interobserver reliability was calculated with the ICC ranging from 0.57 to 0.70 for the static tests. The 2 dynamic tests had poor interobserver reliability (0.26 and 0.39). The author explained the poor reliability as a reflection of the complexity of the movement itself. Harrison et al[8] described interobserver reliability for evaluating single-leg stance in 78 uninjured subjects and 17 anterior cruciate ligament patients. A 3-point scale was used, and specific guidelines were provided to the 2 observers. A weighted kappa of 0.70 was found for this static test.

It would seem that a static test results in a better reliability when compared with dynamic tests. It may be the case that reliable assessment of complex dynamic movement patterns is not possible using clinical judgment and the naked eye alone and that other objective tests are needed.[2] Many studies on core stability have used complex objective tests that require specialized apparatus and are time consuming to perform.[2,9]

A shortcoming of this study that needs discussion was the use of video for the interobserver reliability. It is possible that important visual information is lost by observing the subject 2 dimensionally and only from 1 viewpoint. The choice to use video, however, was based on creating less bias in intraobserver reliability and for logistical reasons. The use of video ensured that exactly the same movement was observed at both moments by all the observers. Any differences observed must have been due to differences in the way the observer scored the test and cannot have been due to 2 observers seeing different movement patterns. The video was projected onto a screen life-sized to make the observation as lifelike and detailed as possible. The possibility that fatigue of the observers during the long duration of the assessment may have affected the reliability cannot be excluded.

In this study, the ICC was used to examine the inter- and intraobserver reliability. In many studies, the kappa is used or the weighted kappa in those with multiple observers. It has been noted that there is no real difference between the ICC and weighted kappa for multiple observers.[10]

The tests were scored separately and not after seeing the whole battery. As such, there was no general score given to the whole battery of tests. It may be that this would lead to an improved reliability, but as the scores were not recorded in this manner, it was not possible to analyze this.

In this study, no attempt was made to measure the core stability objectively using more complex movement analysis systems or electromyography as this was not the aim. The study also provides no insight as to what these tests do measure as the test performance was too poor to go on to examine the validity of the tests.

CONCLUSIONS

This study shows that all 6 clinical tests for core stability have a poor inter- and intraobserver reliability when assessment is done with visual evaluation and the use of a 4-point scoring system. Based on these results, the clinical tests are not reliable enough to be used in the clinical setting. This indicates a need to develop more reliable clinical tests for evaluating core stability.

REFERENCES

1. Kibler WB, Press J, Sciascia A. The role of core stability in athletic function. *Sports Med*. 2006;36:189–198.
2. Borghuis J, Hof AL, Lemmink KA. The importance of sensory-motor control in providing core stability: implications for measurement and training. *Sports Med*. 2008;38:893–916.
3. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinical agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. *J Orthop Sports Phys Ther*. 2007;37:122–129.
4. Jacobs C, Mattacola C. Sex differences in eccentric hip-abductor strength and knee-joint kinematics when landing from a jump. *J Sport Rehabil*. 2005;14:346–355.
5. Zeller BL, McCrory JL, Kibler WB, et al. Differences in kinematics and electromyographic activity between men and women during the single-legged squat. *Am J Sports Med*. 2003;3:449–456.
6. Piva SR, Fitzgerald K, Irrgang JJ, et al. Reliability of measures of impairments associated with patellofemoral pain syndrome. *BMC Musculoskelet Disord*. 2006;31:7–33.
7. Hayes K, Walton JR, Szomor ZL, et al. Reliability of five methods for assessing shoulder range of motion. *Aust J Physiother*. 2001;47:289–294.
8. Harrison EL, Duenkel N, Dunlop R, et al. Evaluation of single-leg standing following anterior cruciate ligament surgery and rehabilitation. *Phys Ther*. 1994;74:245–252.
9. Cholewicki J, Silfies SP, Shah RA, et al. Delayed trunk muscle reflex responses increases the risk of low back injuries. *Spine*. 2005;34:2614–2620.
10. Norman GR, Streiner DL. *Biostatistics: The Bare Essentials*. 2nd ed. Toronto, Canada: BC Decker; 2000:200.