Original research

The reliability of musculoskeletal screening tests used in cricket

Rebecca J. Dennis^{a,b,*}, Caroline F. Finch^{a,b}, Bruce C. Elliott^c, Patrick J. Farhart^d

^aNSW Injury Risk Management Research Centre, University of New South Wales, Sydney, NSW 2052, Australia

^bSchool of Human Movement and Sport Sciences, University of Ballarat, P.O. Box 663, Ballarat, Vic. 3353, Australia

^cSchool of Human Movement and Exercise Science, University of Western Australia, M40835 Stirling Highway, Crawley, WA 6009, Australia

^dCricket New South Wales, P.O. Box 333, Paddington, NSW 2021, Australia

Abstract

Objectives: To determine the inter and intra observer reliability of a field based musculoskeletal screening protocol used to measure potential injury risk factors in cricket fast bowlers.

Design: Test retest reliability study.

Setting: High performance Australian cricket.

Participants: Ten volunteers. Two sports physiotherapists conducted the testing.

Main outcome measures: Participants completed the following tests: knee extension; modified Thomas test (hip extension and abduction); hip internal and external rotation; combined elevation; ankle dorsiflexion lunge; bridging hold; prone four point hold; and calf heel raises.

Methods: For each of the tests, the participants were tested by each physiotherapist twice, and the inter and intra observer reliability were concurrently assessed.

Results: The inter observer reliability of the tests was generally poor, with only four of the ten tests having an intraclass correlation coefficient (ICC) greater than 0.80 (range of ICCs 0.27–0.99). The intra observer reliability of the tests was considerably higher, with nine tests having an ICC greater than 0.80 (range of ICCs 0.56–0.99).

Conclusions: With the exception of the bridging hold, all tests would be considered acceptable where only one observer was conducting the testing. However, only the ankle dorsiflexion lunge, combined elevation test, calf heel raise test and prone four point hold have acceptable reliability when there are multiple physiotherapists recording measurements.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Musculoskeletal; Reliability; Cricket; Fast bowler; Intraclass correlation coefficient

1. Introduction

The development of successful sports injury prevention strategies relies on a solid evidence base. Sports injury prevention is a staged process which consists of a series of steps (Finch, 2006; van Mechelen, Hlobil, & Kemper, 1992). The second step involves developing an understanding of the aetiology of injury, which includes identifying intrinsic injury risk factors to understand

why an athlete may be predisposed to injury (Bahr & Krosshaug, 2005; Finch, 2006).

Preparticipation (or baseline) screening is commonly used to measure potential intrinsic injury risk factors (DiFiori, 1999), by identifying characteristics of the musculoskeletal system that may predispose an athlete to injury, or to identify incomplete recovery from a previous injury. In prospective studies, measurements are made on uninjured athletes at the start of a playing season and these are related to injury outcomes during the course of the season.

This approach was taken with a prospective epidemiological study of 91 cricket fast bowlers. Studies have

identified variations in technique, physical characteristics and workload as injury risk factors for this population. However, a limitation of prior studies is that many have used specialised equipment and laboratory facilities to test the fast bowlers, and the degree to which the testing procedures can be adopted in the cricket “real world” is unclear. The advantages of using a field-based screening protocol are that it can be conducted with easily obtainable and inexpensive equipment, as well as with a large number of athletes at multiple locations. Therefore, the aim of our earlier epidemiological study (Dennis, 2006) was to identify intrinsic risk factors for injury to fast bowlers using a field-based screening protocol, which included a musculoskeletal assessment.

A major issue encountered by studies of intrinsic risk factors, such as those described above, is the reliability of the screening measures used. The ability for studies to clearly identify injury risk factors is highly dependent on the accuracy with which measurements are made (Bahr & Holme, 2003), and it is vital that the screening protocols are reliable (Gabbe, Bennell, Wajswelner, & Finch, 2004; Harvey, 1998b). Measurements need to be reproducible over time and by different observers, as well as being repeatable within a given individual (Hayen, Dennis, & Finch, 2007). Poor reproducibility limits the ability of researchers to reach conclusions about whether a measured variable is indeed a risk factor for injury, because it is difficult to differentiate participants with or without the variable of interest if there is large random measurement error (Haas, 1995; Hayen et al., 2007).

Unfortunately, the reliability of many of the tests included in the musculoskeletal assessment of fast bowlers has not previously been established for this group of athletes. Therefore, this paper presents a reliability assessment of the tests, to inform the development of an appropriate field-based preparticipation musculoskeletal screening protocol for use in cricket and potentially in a range of other sports.

2. Methods

2.1. Development of the protocol

The screening protocol was developed in collaboration with a sports physiotherapist (PF, with 15 years of experience working with cricketers) and consisted of a number of tests measuring flexibility, strength, and stability. The majority of injuries reported among fast bowlers are to the back, trunk and lower limb (Dennis, Farhart, Goumas, & Orchard, 2003; Dennis, Finch, & Farhart, 2005; Orchard, James, Alcott, Carter, & Farhart, 2002), and hence the identification of injury risk factors for these particular body regions was the

primary focus when selecting tests for inclusion in the protocol. The tests chosen are commonly used in cricket (although they are not specific to the sport) and were primarily based on clinical assessment protocols used with athletes (Harvey, 1998b) and incorporated procedures described in other relevant literature (Bennell, Talbot, Wajswelner, Techovanich, Kelly, & Hall, 1998; Berryman Reese & Bandy, 2002; Gabbe et al., 2004; Harvey, 1998a; Norkin & White, 1985).

2.2. Participants and observers

The reliability assessment was conducted with 10 volunteers. Two sports physiotherapists (observers), with extensive clinical experience in cricket, conducted the testing, with an assistant. A training session for the observers and assistants was conducted by the lead author before the reliability assessment, and a detailed procedures manual for the screening protocol was provided. The study was approved by the University of New South Wales Human Ethics Review Committee and participants provided written consent.

2.3. Procedure for completing the musculoskeletal tests

For tests requiring goniometric measurement, a modified goniometer was used. A spirit level was attached to one arm of the goniometer, so that the degree of motion relative to the vertical could be determined. The procedures for each of the tests are described below.

1. *Knee extension*: This test was used to assess hamstring muscle length and the range of assisted active knee extension in a position of hip flexion (Harvey, 1998b). While the participant lay in a supine position with the hip of their testing leg flexed, the observer provided support behind the posterior thigh to hold the hip in 90° flexion. The participant relaxed their foot and slowly extended their knee. Once the limit of extension was reached, the observer provided support behind the calf while the assistant recorded knee extension (x°), relative to the vertical, to the nearest degree. The final angle was calculated as $90^\circ - x^\circ$. In cases where the participant was able to achieve full knee extension, the observer flexed the knee and moved the thigh to 30° past the vertical. The knee was again straightened until the limit of knee extension (x°). The final angle was calculated as $120^\circ - x^\circ$.
2. *Modified Thomas test (MTT) (hip extension)*: The MTT assesses the flexibility of the hip flexors. The procedure used was the same as that described previously (Harvey, 1998a). The participant perched on the end of the bench and rolled back into a supine position, while holding both knees firmly to their chest. They held their contralateral hip in maximal

Fig. 1. Modified Thomas test (hip abduction).

flexion with both arms, while their testing leg was lowered towards the floor. An assistant supported this position by pushing gently on the contralateral knee. The participant was asked to relax the hip and thigh muscles of their testing leg so that a passive end point position was obtained due to gravity alone. The axis of the goniometer was placed over the greater trochanter, with the fixed axis directed vertically using the spirit level. The moveable arm of the goniometer was pointed toward the lateral knee joint line, representing the line of the femur. The observer then assessed the hip angle relative to the horizontal, or 0° -axis, to the nearest degree. A negative angle represented flexion above the horizontal and a positive represented extension below the horizontal.

3. *MTT (hip abduction)*: The MTT was also used to measure the degree of hip abduction. The same starting position as the previous test was used and the test procedure was modified from previous research (Harvey, 1998a). When the participant was in the supine position, the assistant held the contralateral hip in maximal flexion and added external rotation in order to move the anterior superior iliac spine (ASIS) of the left and right side into a parallel alignment (Fig. 1). Hip abduction angle was measured with the centre of the goniometer over the ipsilateral ASIS,

Fig. 2. Prone four point hold.

with the flexible arm positioned to the opposite ASIS. Hip abduction was recorded to the nearest degree, with the line perpendicular to that of the ASIS considered as 0° . A negative angle represented hip adduction and a positive represented hip abduction.

4. *Hip internal and external rotation*: This test was used to assess the range of hip rotation in a neutral hip position (Harvey, 1998b). The participant lay in a prone position with both knees bent to 90° , chin resting on the bench and arms by their sides. The participant then straightened their contralateral knee. To determine internal rotation, the participant was asked to let the ankle of the testing leg move laterally as far as possible, while the observer ensured that pelvic motion and/or hip flexion did not occur. To determine external rotation, the participant was asked to let the ankle of their testing leg drop towards the opposite side of the body as far as possible. For both internal and external rotation, the assistant measured the angle formed by the line of the tibia, relative to the vertical, to the nearest degree.
5. *Combined elevation test*: The combined elevation test assesses combined thoracic extension (strength and range of motion), shoulder girdle flexion and scapula retraction (Harvey, 1998b). The participant lay in a prone position on the floor, with both arms outstretched. They kept elbows extended, thumbs locked together and palms facing towards the floor. For the duration of the test, the participant kept their feet, hips, chest, and chin on the ground. They were asked to take a breath in and hold, and then raise both arms off the floor as high as possible, without flexing their elbows. A tape measure was used to measure the perpendicular distance from the base of the metacarpal of the thumb to the floor, to the nearest 0.5 cm.
6. *Prone four point hold*: This test was chosen to assess the length of time that the participant could maintain a neutral lumbopelvic position in a four point hold position. The participant lay in a prone position on the floor, resting on their forearms, with elbows flexed at 90° . Their forearms were held parallel to each other, shoulder width apart. Participants were not permitted to grip their hands together at any

stage during the test. Their feet were placed in a push-up position, with ankles flexed and the base of their toes on the floor. Participants were instructed to pull their navel in towards their spine and hold. They then raised their trunk off the floor, resting on their forearms and toes, holding a neutral lumbopelvic position as long as possible (Fig. 2). The test was terminated if the participant experienced back pain, if they could no longer hold the neutral position (as determined by the observer) or if they could simply not sustain the prone hold any longer. A stopwatch was used to determine the length of time the participant held a neutral position, to the nearest second.

7. *Bridging hold*: The bridging hold assesses gluteal musculature strength and endurance. The participant lay supine on the floor, with both knees bent at 90°. They raised their hips and pelvis off the floor in order to attain a continuous alignment from the shoulder, through the hip, through to the knee. They then raised their contralateral foot off the floor and with their foot in a neutral position, fully extended their knee (Fig. 3). The participant was asked to use the gluteal muscles to maintain the same height of the ASIS of the left and right side, and stop the lower back from arching. They held this position for as long as possible, until the position was no longer held correctly (as determined by the observer), the participant experienced low back or hamstring pain or could simply not hold the position any longer. Duration of the hold was recorded to the nearest second.

8. *Ankle dorsiflexion lunge*: This test measures the range of dorsiflexion at the ankle joint. A tape measure was fixed along the floor with the 0cm point at the junction of the floor and wall. The participant positioned their foot beside the tape so that their heel line and big toe were aligned beside the tape measure. The observer held the participant's heel to

Fig. 3. Bridging hold.

Fig. 4. Ankle dorsiflexion lunge.

prevent it from lifting off the floor and manually locked the subtalar joint so it remained in a neutral position throughout the test. The participant lunged forward until their knee touched the wall (Fig. 4). The maximum distance from the great toe to the wall was recorded to the nearest 0.1 cm.

9. *Calf heel raises*: This test was used to assess endurance of the ankle plantarflexor muscles in a weightbearing task. With bare feet, the participant stood with the balls of their feet on the edge of a step with their heels off the step. The participant was instructed to stand on one foot and to rise onto the ball of that foot as high as possible, then slowly lower their heel, while maintaining knee extension throughout the movement. The heel raise/lower cycle was repeated continuously (without a rest) until the participant was unable to rise through full range with the knee extended. They were asked to perform the heel raises at a rate of approximately one cycle per second.

2.4. Procedure for conducting the reliability assessment

2.4.1. Range of motion tests

The participants were required to attend one appointment, in which they were tested by each observer twice, in the order of Observer 1, Observer 2, Observer 1, Observer 2. The tests were conducted in the same order each time, with 10-min rest breaks between each session for the following tests: knee extension; MTT (hip extension and abduction); hip internal and external rotation; combined elevation; and ankle dorsiflexion lunge. The assistant recorded the measurements for each

test and at the end of each session they filed the recording sheet. This was done so that the observers could not see the results of the first session when they were completing the second session.

2.4.2. Muscular endurance tests

A different method was used to assess the reliability of the bridging hold, prone four point hold and calf heel raise tests, because they are measures of muscular endurance and require maximal effort. The scores for multiple trials could be vastly different and this would result in poor inter- and intra-observer reliability results. However, these differences would not necessarily be a result of measurement error by the observer. Therefore, the participants only completed one trial for these tests, which was examined simultaneously by both observers. A video camera was positioned side-on to the participant to record the trial (sagittal plane). The footage was later viewed by each observer to determine his or her scores for a second “trial”, to allow the calculation of inter- and intra-observer reliability. For the bridging hold and prone four point hold, the video camera was positioned approximately 30 cm off the ground (in line with the height of the pelvis). For the calf heel raises, the video camera was approximately 100 cm off the ground.

For the four point prone hold and the bridging hold, the observers each had a stopwatch. The participant held the position for as long as possible until failure. The observers did not stop the participant during the test if they were no longer holding the correct position. Instead, they noted the time at which they would usually terminate the test. A similar process was used for the calf heel raises. Both observers observed the trial while the video footage was being captured and did not stop the participant if they were no longer raising the heel through a full range with the knee extended. Instead, they noted the number of correctly completed trials. The observers did not communicate during the test, and on completion of the trial recorded their results on separate blank recording sheets. Approximately, 15 min after the muscular endurance tests had been completed the observers individually reviewed the video footage and recorded their assessments.

2.5. Statistical analyses

A method for the concurrent assessment of inter- and intra-observer reliability was used to calculate intraclass correlation coefficients (ICCs) (Eliaszew, Young, Woodbury, & Fryday-Field, 1994; Hayen et al., 2007). A review paper discussing the application of this statistical method to sports injury research has been published (Hayen et al., 2007).

The results presented include the ICCs, confidence intervals (CIs) for the ICCs, and the standard error of measurement (SEM), the definitions of which have been

provided previously (Hayen et al., 2007). The SEM, also known as measurement error, was expressed in the same units as the test was measured (e.g. degrees). It is possible to translate the test statistic for the SEM into a standard normal critical value, which represents the minimum difference between measurements that needs to be exceeded to be confident that a real change has occurred (Hayen et al., 2007).

The classification system for determining acceptable ICCs adopted in this paper was that of Landis and Koch (1977). This classifies ICCs as slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00). Finally, to determine if there was any “learning” effect because the participants completed each test four times, or observer order effect, a general linear repeated measures model was conducted using SPSS for Windows.

3. Results

The inter-observer reliability of the tests is presented in Table 1 and the intra-observer reliability is shown in Table 2. Generally, the inter-observer reliability of the tests was low, with only four of the ten tests being considered almost perfect with an ICC greater than 0.80. The intra-observer reliability of the tests was considerably higher than inter-observer reliability, with nine tests in the category of almost perfect. The 95% CIs were wide, especially for inter-observer reliability. It should also be noted that the SEM critical value, which represents the minimum difference between measurements that needs to be exceeded to be confident that a real change has occurred, was high for many of the tests when considering inter-observer reliability. Whereas, it was considerably lower for intra-observer reliability.

As presented in Table 3, the repeated measures analysis identified two of the ten tests as having a significant association for the trial number and one test as having a significant association for the observer number. None of the tests demonstrated a significant interaction between the observer and trial number.

4. Discussion

Pre-participation screening is a commonly used method for the collection of data relating to potential intrinsic risk factors. So that the data collected during these pre-participation screenings can be used to identify injury risk factors and subsequently inform the development of effective injury prevention strategies, it is vital that the screening protocols are reliable. This study has demonstrated that the intra-observer reliability of a musculoskeletal assessment used with cricket fast bowlers is generally high. However, the inter-observer

Table 1
Inter observer reliability of the musculoskeletal screening protocol

Screening test	Inter observer reliability				
	ICC	95% CI	SEM	SEM critical value	Classification
Knee extension (deg)	0.40	0.00, 0.79	5.5	15.1	Fair
MTT hip extension (deg)	0.27	0.00, 0.76	6.6	18.4	Fair
MTT hip abduction (deg)	0.29	0.00, 0.71	2.5	7.2	Fair
Hip internal rotation (deg)	0.30	0.00, 0.78	8.9	24.6	Fair
Hip external rotation (deg)	0.66	0.25, 0.90	5.0	13.8	Substantial
Ankle dorsiflexion lunge (cm)	0.96	0.89, 0.99	0.4	1.1	Almost perfect
Combined elevation test (cm)	0.87	0.63, 0.97	1.9	5.4	Almost perfect
Calf heel raises	0.99	0.98, 1.00	0.3	0.8	Almost perfect
Bridging hold (s) ^a	0.56	0.42, 0.88	24.8	68.7	Moderate
Prone four point hold (s) ^a	0.89	0.79, 0.97	15.7	43.6	Almost perfect

MTT: Modified Thomas test; ICC: intraclass correlation coefficient; CI: confidence interval; SEM: standard error of measurement.

^aVideo footage was used to determine the reliability of these tests.

Table 2
Intra observer reliability of the musculoskeletal screening protocol

Screening test	Intra observer reliability				
	ICC	95% CI	SEM	SEM critical value	Classification
Knee extension (deg)	0.86	0.43, 0.95	2.6	7.2	Almost perfect
MTT hip extension (deg)	0.97	0.84, 0.99	1.3	3.7	Almost perfect
MTT hip abduction (deg)	0.83	0.21, 0.92	1.3	3.5	Almost perfect
Hip internal rotation (deg)	0.94	0.68, 0.97	2.7	7.5	Almost perfect
Hip external rotation (deg)	0.88	0.54, 0.96	3.0	8.3	Almost perfect
Ankle dorsiflexion lunge (cm)	0.98	0.92, 0.99	0.3	0.8	Almost perfect
Combined elevation test (cm)	0.97	0.88, 0.99	1.0	2.7	Almost perfect
Calf heel raises	0.99	0.97, 1.00	0.3	0.8	Almost perfect
Bridging hold (s) ^a	0.56	0.00, 0.83	24.8	68.7	Moderate
Prone four point hold (s) ^a	0.89	0.62, 0.97	15.7	43.6	Almost perfect

^aVideo footage was used to determine the reliability of these tests.

Table 3
Musculoskeletal screening tests with a significant association for trial or observer number identified by repeated measures analysis

Screening test	Association for trial number					
	Trial 1 mean	SD	Trial 2 mean	SD	F statistic	p Value
Hip internal rotation (deg)	46.1	10.52	44.6	10.19	7.154	0.025
Bridging hold (s) ^a	45.8	24.36	71.7	36.17	15.077	0.004
	Association for observer number					
	Observer 1 mean	SD	Observer 2 mean	SD	F statistic	p Value
MTT hip abduction (deg)	12.4	2.99	10.2	2.23	6.839	0.028

SD = standard deviation.

^aVideo footage was used to determine the reliability of this test.

reliability of many of the tests was poor. This has implications for the use of this protocol to identify injury risk factors, as establishing reliability is a prerequisite for ascertaining the validity of the risk factor measures.

There are no universally applicable standards as to how high the ICC must be to constitute acceptable reliability in sports injury risk factor research. Therefore in the absence of clear guidelines, it would seem reasonable to consider those tests with reliability

considered almost perfect ($ICC = 0.80\text{--}1.00$) as acceptable for the purposes of large injury risk factor identification studies. The majority of the musculoskeletal screening tests demonstrated excellent intra-observer reliability, with ICCs above 0.80 and low SEM critical values. The only exception was the bridging hold test, with an ICC of 0.56. It is difficult to compare these findings with previous research, as different statistical methods to determine the ICCs have been used. Many sports injury studies have used the mean of two measurements to determine ICCs (Hayen et al., 2007). This has the effect of inflating the ICC, as the mean of multiple measurements will be higher than that based on a single measurement (Eliaszew et al., 1994; Fleiss, 1986; Hopkins, 2000). This may partially explain the apparent poor inter-observer reliability of many of the tests. Consistent with the findings of this study, previous studies have reported high intra-observer reliability of the MTT (Gabbe et al., 2004; Harvey, 1998a) and the ankle dorsiflexion lunge (Bennell et al., 1998).

The inter-observer reliability of the ankle dorsiflexion lunge, the combined elevation test, calf heel raise test and the prone four point hold was almost perfect. For the ankle dorsiflexion lunge and calf heel raise test, the SEM critical values demonstrated that only a small difference would need to be recorded to be confident there was a real variation between the observers. For example, for ankle dorsiflexion, a difference of only 1.1 cm would need to be recorded between different observers to be confident of a real difference in the measurements. When considering the test–retest reliability of a single observer, a difference of only 0.8 cm would need to be identified. The ankle dorsiflexion was the only one of these tests using the same protocol as in previous research, which also reported high inter-observer reliability of this test (Bennell et al., 1998). The remaining tests in the protocol demonstrated lower ICCs for inter-observer reliability, ranging between 0.27 and 0.66. Of these, the MTT demonstrated much lower inter-observer reliability than has been reported in previous research (Gabbe et al., 2004), which may be due to the differences in the statistical methods used (Hayen et al., 2007).

With the exception of the bridging hold test, all of the tests in the musculoskeletal screening protocol demonstrated almost perfect intra-observer reliability and would be considered acceptable tests for those occasions where only one observer was conducting multiple measurements. However, it must be noted that the tests would only be acceptable when an observer collects valid information. For example, an observer may demonstrate high intra-observer reliability when recording joint range of motion, but the results will not be valid if the observer consistently over-estimates the motion of interest. Validity was not assessed in this study.

The inter-observer reliability was generally not as high, but the ankle dorsiflexion lunge, combined elevation test, calf heel raise test and prone four point hold demonstrated almost perfect inter-observer reliability and would again be considered acceptable. This is of particular importance to large-scale injury risk factor identification studies, because multiple observers often contribute measurements to determine injury risk. The SEM critical value is of particular interest when considering the reliability of tests to be used in such research projects. For some tests, only a small difference between measurements would need to be observed to be confident of a real difference. Whereas, for other tests the SEM critical values were high, such as the bridging hold where 43.6 s difference would be needed to be sure that the measurements between multiple observers, or between different time periods for the same observer, were indeed different.

This study has contributed important information for the development of reliable screening protocols for use with fast bowlers, and potentially other athletes. However, the limitations of this reliability assessment are acknowledged. The validity of the tests was not assessed, but as described earlier, the tests were based on well-established clinical and field assessment tools for use with athletes. However, future research may consider the validity (the degree to which the test measures what it is supposed to measure) of these tests, in addition to reliability. Another limitation was that a small number of participants were recruited. This is not dissimilar to previous reliability studies that have only enrolled 10–20 participants, but ideally a larger number of participants would have been recruited. The current study was limited by wide 95% CIs surrounding the ICCs, particularly for the assessment of inter-observer reliability. Future research with a larger sample may avoid this problem. Finally, this reliability assessment used a short time interval to separate the testing sessions and the reliability of the measurements represents their reproducibility only within this particular time frame.

Analyses were conducted to determine if there was any “learning” effect during this reliability assessment. For hip internal rotation, there was a reduction in joint range of motion from the first to the second trial completed with each observer, but the magnitude of this is not likely to be clinically relevant. For the bridging hold, there was a significant increase from the first to the second trial and this is most likely due to the use of video footage, which indicates that this is not a reliable method to conduct this test of muscular endurance. Only one of the tests (the MTT, assessing hip abduction) demonstrated a significant difference for the means of observers 1 and 2, but this difference may not have been clinically relevant.

It is unclear as to why the inter-observer reliability was poor for the majority of the tests, but it may be

related to the observers using different landmarks or varying force in determining the end-point for the range of motion tests (Gajdosik & Bohannon, 1987; MacDermid, Chesworth, Patterson, & Roth, 1999; Martin & McPoil, 2005). The poor reliability of the bridging hold test may be due to the inability of the observers to accurately detect any change in the position of the hips in the second “trial”, when using video footage of the single trial conducted. The observers reported difficulty in examining the movement of the hips and pelvis when using the two-dimensional video footage, and therefore, this method of assessing performance in this test is limited. It must also be noted that because only a single trial was completed for the muscular endurance tests, only differences in the measurements recorded by the observers could be identified. In contrast, four trials were completed for the joint range of motion tests and so differences in both the measurements recorded by the observers, and differences in the performance of the participants, could be identified.

Further studies should endeavour to identify reliable field-based tests for ongoing use with fast bowlers and other athletes. It may be necessary to adopt different testing procedures, such as the positions used (e.g. prone vs. supine), and the type of measurement adopted (e.g. active vs. passive). It may be adequate to retain the same protocol, but necessary to simply conduct more comprehensive training of the observers and the assistants. Finally, it may be possible that only those tests conducted in a laboratory are able to provide both valid and reliable information for fast bowlers and that it is simply not possible to assess these measures in a field-based environment.

5. Conclusions

Sports injury prevention requires a firm evidence base. An important component of this is the accuracy and reliability of measurements taken in studies attempting to identify risk factors for injury. This study has established the reliability of a screening protocol for fast bowlers to measure potential risk factors for injury; it also has implications for screening completed with athletes in other sports. All tests (except the bridging hold) would be considered acceptable where only one observer was conducting the tests with all study participants. However, only the ankle dorsiflexion lunge, combined elevation test, calf heel raise test and prone four point hold have high enough inter-observer reliability to use where there are multiple observers recording measurements. It should be noted that for many tests, the minimum difference between measurements that needs to be exceeded to be confident that a real change has occurred is quite high. This has

implications for the use of these tests both in clinical and research environments. Further work is necessary to develop a comprehensively reliable field-based screening protocol so that it is possible to accurately identify injury risk factors for fast bowlers.

Acknowledgements

RD was supported by an NHMRC Public Health PhD scholarship during the data collection and analysis phase, and by an NHMRC Population Health Capacity Building Grant in Injury Prevention, Trauma and Rehabilitation during the reporting and publication phase. CF was supported by an NHMRC Principal Research Fellowship. Funding for the project was provided by Cricket Australia. The authors wish to acknowledge the valuable contributions of Andrew Hayen, Andrew Nealon, Kate Watson, Kevin Sims, Danny Redrup, Marc Portus, Andrew McIntosh, and the participants in the study.

Conflict of interest

There are no conflicts of interest.

References

- Bahr, R., & Holme, I. (2003). Risk factors for sports injuries: A methodological approach. *British Journal of Sports Medicine*, 37, 384–392.
- Bahr, R., & Krosshaug, T. (2005). Understanding injury mechanisms: A key component of preventing injuries in sport. *British Journal of Sports Medicine*, 39, 324–329.
- Bennell, K. L., Talbot, R. C., Wajswelner, H., Techovanich, W., Kelly, D. H., & Hall, A. J. (1998). Intra rater and inter rater reliability of a weight bearing lunge measure of ankle dorsiflexion. *Australian Journal of Physiotherapy*, 44, 175–180.
- Berryman Reese, N., & Bandy, W. D. (2002). *Joint range of motion and muscle length testing*. Sydney, Australia: W.B. Saunders Company.
- Dennis, R. J. (2006). *Risk factors for repetitive microtrauma injury to adolescent and adult cricket fast bowlers*. Ph.D. thesis, University of New South Wales, Sydney, Australia.
- Dennis, R., Farhart, P., Goumas, C., & Orchard, J. (2003). Bowling workload and the risk of injury in elite cricket fast bowlers. *Journal of Science and Medicine in Sport*, 6, 359–367.
- Dennis, R. J., Finch, C. F., & Farhart, P. J. (2005). Is bowling workload a risk factor for injury to Australian junior cricket fast bowlers? *British Journal of Sports Medicine*, 39, 843–846.
- DiFiori, J. (1999). Overuse injuries in children and adolescents. *The Physician and Sports Medicine*, 27, 75–89.
- Eliasziw, M., Young, S. L., Woodbury, M. G., & Fryday Field, K. (1994). Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Physical Therapy*, 74, 777–788.
- Finch, C. (2006). A new framework for research leading to sports injury prevention. *Journal of Science and Medicine in Sport*, 9, 3–9.
- Fleiss, J. L. (1986). Reliability of measurement. In J. L. Fleiss (Ed.), *The design and analysis of clinical experiments*. New York: Wiley.
- Gabbe, B. J., Bennell, K. L., Wajswelner, H., & Finch, C. F. (2004). Reliability of common lower extremity musculoskeletal screening tests. *Physical Therapy in Sport*, 5, 90–97.

- Gajdosik, R. L., & Bohannon, R. W. (1987). Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. *Physical Therapy*, 67, 1867–1872.
- Haas, M. (1995). How to evaluate intraexaminer reliability using an interexaminer reliability study design. *Journal of Manipulative and Physiological Therapeutics*, 18, 10–15.
- Harvey, D. (1998a). Assessment of the flexibility of elite athletes using the modified Thomas test. *British Journal of Sports Medicine*, 32, 68–70.
- Harvey, D. (1998b). *Screening test protocols. Pre-participation screening of athletes: A program developed by the Australian Institute of Sport, Olympic Athlete Program and the Australian Sports Injury Prevention Taskforce*. Canberra, Australia: Australian Sports Commission.
- Hayen, A., Dennis, R. J., & Finch, C. F. (2007). Determining the intra- and inter-observer reliability of screening tools used in sports injury research. *Journal of Science and Medicine in Sport*, 10, 201–210.
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30, 1–15.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- MacDermid, J. C., Chesworth, B. M., Patterson, S., & Roth, J. H. (1999). Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. *Journal of Hand Therapy*, 12, 187–192.
- Martin, R. L., & McPoil, T. G. (2005). Reliability of ankle goniometric measurements: A literature review. *Journal of the American Podiatric Medical Association*, 95, 564–572.
- Norkin, C. C., & White, D. J. (1985). *Measurement of joint motion: A guide to goniometry*. Philadelphia, PA: F.A. Davis Company.
- Orchard, J., James, T., Alcott, E., Carter, S., & Farhart, P. (2002). Injuries in Australian cricket at first class level 1995/1996 to 2000/2001. *British Journal of Sports Medicine*, 36, 270–274.
- van Mechelen, W., Hlobil, H., & Kemper, H. C. (1992). Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Medicine*, 14, 82–99.