Bakken, A., Targett, S., Bere, T., Eirale, C., Farooq, A., Tol, J. L., Whiteley, R., Witvrouw, E., Khan, K. M., Bahr, R. (2016). Interseason variability of a functional movement test, the 9+ screening battery, in professional male football players. *British Journal of Sports Medicine, under utgivelse*. doi: 10.1136/bjsports-2016-096570

Title

Inter-season variability of a functional movement test, the 9+ screening battery, in professional male football players

Author list

Bakken A[1,2], Targett S[1], Bere T[1,2],Eirale C[1], Farooq A[1], Tol JL[1,3,4], Whiteley R[1], Witvrouw E[5], Khan KM[6], Bahr R[1,2]

Affiliations

[1]Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar

[2]Oslo Sports Trauma Research Center, Norwegian School of Sport Sciences, Oslo, Norway,

[3]The Sports Physician Group, Department of Sports Medicine OLVG, Amsterdam, the Netherlands

[4]Academic Center of Evidence Based Sports Medicine, Academic Medical Center, Amsterdam, the Netherlands

[5]Department of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[6]Center for Mobility and Hip Health, University of British Columbia, Vancouver, Canada

Correspondence to: Arnhild Bakken, Department of Sports Medicine, Aspetar Orthopaedic and Sports Medicine Hospital, Sport City street, PO Box 29222, Doha, Qatar, Phone: + 974 66 17 23 61, E-mail: arnhild.bakken@aspetar.com

Keywords: Functional movement test, screening, injury prevention, athletes, measurement properties

ABSTRACT

Background: The Nine Plus screening battery test (9+) is a functional movement test intended to identify limitations in fundamental movement patterns predisposing athletes for injury. However, the inter-season variability is unknown. **Aim:** To examine the variability of the 9+ test between two consecutive seasons in professional male football players. **Methods:** Asymptomatic Qatar Star League players (n=220) completed the 9+ at the beginning of 2013 and 2014 seasons. Time-loss injuries in training and matches were obtained from the Aspetar Injury and Illness Surveillance Program. No intervention was initiated between test occasions. **Results:** A significant increase in mean total score of 1.6 points (95% CI 1.0 to 2.2, P<0.001) was found from season 1 (22.2±4.1 (SD)) to season 2 (23.8±3.3). The variability was large, as shown by an intraclass correlation coefficient (ICC) of 0.24 (95% CI 0.11 to 0.36) and a minimal detectable change (MDC) of 8.7 points. Of the 220 players, 136 (61.8%) suffered a time-loss injury between the two tests. There was an improvement in mean total scores in the injured (+2.0±0.4 (SE), P<0.001) group but not in the uninjured group (+0.9±0.5, P=0.089). The variability from season 1 to 2 was large in both the injured (ICC: 0.25, 0.09 to 0.40, MDC: 8.3) and uninjured group (ICC: 0.24, 0.02 to 0.43, MDC: 9.1). **Conclusion:** The 9+ demonstrated substantial intra-individual variability in the total score between two consecutive seasons, irrespective of injury. A change above 8 points is necessary to represent a real change in the 9+ test between seasons.

What are the new findings?

- There was a substantial intra-individual variability of the 9+ total score between two consecutive seasons, irrespective of injury and severity status.
- A change above 8 points is necessary to represent a real change in the 9+ test between seasons, irrespective of injury
- There was a small but systematic improvement in the 9+ total score among injured and uninjured players.

How might it impact on clinical practice in the near future?

Practitioners should consider the large intra-individual variability of the Nine Plus screening battery test (9+), and the high minimal detectable change (MDC) necessary to represent a real change in the 9+ test (irrespective of injury), when interpreting the 9+ total score or similar functional movement screen scores between seasons.

INTRODUCTION

Injuries are common in football causing substantial morbidity, and may have long-term health consequences to the player.[1-3] One strategy to prevent injuries is the use of a periodic health evaluation (PHE) or screening exam to identify the athlete at risk for injury, with a view to implementing targeted prevention measures.[4] For a PHE to be effective in detecting injury risk or be clinically useful, it is essential that the screening tools or tests used are reliable, valid and reproducible, and have acceptable measurement error.[5-9]

Functional movement tests have become popular components of musculoskeletal screening examinations, and are also used for clinical assessments to determine treatment response and assist in return to play decision making.[10] The Nine Plus screening battery test (9+) is a functional movement test attempting to identify limitations in fundamental movement patterns predisposing athletes for injury.[11] This relatively recently developed tool comprises of six tests with modified criteria from the Functional Movement Screen (FMS) (deep squat, in-line lunge, shoulder mobility, trunk stability push-up, active hip flexion and diagonal lift); in addition, Frohm *et al*[11] included five additional tests (one-legged squat, deep one-legged squat, drop jump test, seated rotation and straight leg raise) to fill the gap for tests challenging dynamic trunk flexors, rotation of the spine, and knee control and strength.[11, 12]

There is limited evidence for the measurement properties of the 9+. An initial study by Frohm *et al*[11] found good inter-rater (ICC: 0.80) and intra-rater reliability (ICC: 0.75) of the 9+ in a sample of elite football players. The validity of the 9+ in predicting injury is still unknown. However, athletes with scores below 67% of the total score on the FMS have shown a significant higher injury risk compared with athletes who score above 67%.[13] For 9+ to be clinical useful as a potential predictor, it is important to document the normal variation, in the absence of any intervention or injury, to be able to meaningfully interpret differences in a test result.[14]

Therefore we aimed to examine the season-to-season variability of the 9+ in a group of professional male football players. We hypothesised that in the absence of any prevention or performance intervention or injury, the 9+ score would be stable (i.e. low variability) between seasons.

METHODS

Study Design and Participants

We analysed prospectively collected data from a PHE of male professional football players in Qatar.[15] All players eligible to compete in the Qatar Stars League (QSL), the professional first division of football in Qatar, were invited to participate as they presented for their annual PHE at Aspetar Orthopaedic and Sports Medicine Hospital in Doha (Qatar) at the beginning of the 2013 and 2014 seasons, which the majority (66.6%) completed during the pre-season period (July through September). A smaller group (23.8%) completed the tests during the early/mid competition phase (October through December 2013 or 2014) and a few (9.7%) did the testing during the 2014 post-season (May through June).

As part of the musculoskeletal component of the PHE, all players underwent the 9+ test in the rehabilitation department of the hospital each year. Players presenting with 9+ data from both season examinations (2013 and 2014) were included for analyses. Players reporting a current injury or physical complaint limiting training or match play at the time of testing were excluded from analyses. Ethical approval was obtained from the Institutional Review Board, Anti-Doping Lab Qatar. All players signed a written informed consent form at inclusion, allowing their data being utilised for research.

Examiners

The 9+ was performed by experienced sports physiotherapists working at the study institution. In total, 14 physiotherapists were involved in performing the 9+ testing during the study period (7 performed both seasons, 7 in one of the two seasons only). Prior to testing, all physiotherapists underwent a two-day course with the inventors of the 9+,[11] in addition to performing the 9+ in their clinical practice.

We measured the inter-tester reliability of the 14 physiotherapists in a sub-group of 63 randomly chosen players during the screening setting in the 2014 season. The inter-tester reliability for the total score and each tests was examined with two testers from a randomly selected pool of 8 of the 14 physiotherapists (4 of these were involved in testing both seasons, 4 in the 2014 tests only). The testers were blinded for each other's 9+ score.

Procedures

The 9+ screening battery was performed as described by Frohm *et al*[11, 12] on both test occasions (2013 and 2014). The 9+ consist of 11 functional and complex movement exercises to assess stability, mobility and neuromuscular control in the kinetic chain. Each player performed the 11 tests and they completed each test in the same order on both test occasions. Seven of the 11 tests are assessed bilaterally, looking for asymmetries. For these tests, the left extremity was tested first and the lower of the two scores for the left and right side was used for data analysis. Each movement test was scored on a four-point scale (3-0), with 3 representing correct completion of the task with no compensatory movements, 2 correct but with presence of compensatory movement, 1 not correct despite compensatory movements and 0 if pain was present. Thus, the player could reach a maximum score of 33 points. A more detailed description of the 9+ movements is provided by Frohm *et al*.[11, 12]

All players performed the tests barefoot, with shorts and a t-shirt, except for the drop jump test. As described by Frohm *et al*,[12] the players wore their own training shoes for this test. Due to equipment availability, the participants performed the drop-jump test from a 30-cm box, in contrast to 40-cm box height as described by Frohm *et al*.[12] The physiotherapists gave a standardised verbal instruction, and showed the player a photo of the starting and finishing position of an optimally performed exercise. Each player performed each test 3 times, and the maximum score achieved was recorded and used for evaluation of test performance. Verbal corrections were given during the three trials in order to achieve the most optimal performance. All testers and participants were blinded to the player's score from test occasion 1 on test occasion 2. The 9+ took 20-30 min to complete.

Between-season data collection

After the completion of the initial 9+ in 2013, a report form with the total 9+ score along with the results from the other PHE tests was given to the respective team doctor.[15] Other than that, no specific intervention was advised based on 9+ score from test 1. Data on injuries in training and matches during the intervening football season were obtained from the Aspetar Injury and Illness Surveillance Program (AIISP).[2]

The AIISP is based on prospective injury recording from all 14 QSL teams. An injury was recorded if the player was unable to fully participate in future football training or match play (time-loss injury).[2, 16] The player was considered injured until declared fit for full participation in training and available for match selection by medical staff.

The team physician (or head physiotherapist when no physician was available) for each team recorded all injuries daily throughout the intervening season. For each injury recorded, the team physician/physiotherapist completed a standardised injury card containing information on the body part injured, injury type and specific diagnosis. In addition, the injury card included questions related to re-injury, overuse or trauma, injury mechanism (contact or collision), as well as information on whether the injury occurred during training or match play. Injury severity was determined by the number of days absent from matches or training sessions due to injury and was classified as mild (1-3 days), minor (4-7 days), moderate (8-28 days) or severe (>28 days). Injury data were requested from the clubs every month. We maintained regular communication with the clubs to encourage timely and accurate reporting.

Statistical analyses

Data were analysed with IBM SPSS statistics, version 21 (IBM Corp., Armonk NY). We used a paired t-test to assess for systematic differences in the 9+ total score between test occasions. Significance level was set at $p < 0.05$. The variability (random error) of the 9+ total score between tests was assessed using the intraclass correlation coefficient ($ICC_{1.1}$) with 95% confidence intervals (CI), and standard error of measurement (SEM).[17, 18] The SEM was calculated from the square root of the mean square of the residual (MSr) term derived from the analysis of variance (ANOVA). The minimal detectable change (MDC) with 95% certainty was calculated as SEM x 1.96 x $\sqrt{2}$.[17, 19]

Systematic differences and the variability of each movement tests between test occasions were also examined. As each movement test is measured on an ordinal scale, a non-parametric test (Wilcoxon signed rank test) and weighted kappa (κ_w) were used. The weighted kappa was calculated using STATA (version 11.0, StataCorp, College Station (Texas), USA).

The inter-tester reliability for the total score was analysed using $ICC_{1.1}$ with scores between 0.75 and 1.00 interpreted as good, 0.50 to 0.74 as moderate, and those below 0.50 as poor.[20] The κ_w was used to analyse the inter-tester reliability for each movement test with scores interpreted as follows: <0.20 as poor, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.00 as excellent.[21]

Data are presented as means with standard deviations (SD) or 95% CI unless otherwise stated.

RESULTS

Participants

A total of 247 male footballers completed the 9+ during both the 2013 and 2014 seasons. Of these, 27 players were excluded from analyses because of current injury, no consent or missing injury registration (figure 1). Thus, the final sample included 220 players (age: 25.3±4.6 years; height: 176±7 cm; body mass: 71±9 kg; body mass index: 22.8±2.0 kg/m^2). The players represented 35 nationalities, the majority from the Middle East (71.8%). By ethnicity, 57.7% were Arabic, 29.5% black, Caucasian 3.6%, East Asian 0.9%, Persian 6.8% and 1.4% other. There were no missing items of the 9+ among any of the players included in the inter-season variability and inter-tester reliability analyses.

Examiner inter-tester reliability

The inter-tester reliability for the total score was moderate (ICC=0.68), while the inter-tester reliability for each test ranged from fair to excellent (κ_w=0.31 to 0.81) (Table 1). For eight of the eleven exercises (72%), reliability was fair or moderate.

Table 1 Inter-tester reliability for each movement test of the 9+ (n=63 athletes tested by 8 physiotherapists).

	κ_w (95% CI)	**Strength of agreement**
Deep squat	.52 (.30 to .69)	Moderate
One-legged squat	.38 (.12 to .59)	Fair
Deep one-legged squat	.36 (.11 to .56)	Fair
In-line lunge	.31 (.10 to .54)	Fair
Active hip flexion	.81 (.69 to .91)	Excellent
Straight leg raises	.57 (.38 to .76)	Moderate
Push up	.47 (.19 to .73)	Moderate
Diagonal lift	.39 (.13 to .64)	Fair
Seated rotation	.68 (.51 to .84)	Substantial
Shoulder mobility	.81 (.67 to .92)	Excellent
Drop jump	.46 (.20 to .70)	Moderate

Inter-season variability of the 9+

The mean time between the two 9+ test scores was 359.7±65.4 days. We observed a statistically significant increase in mean total score of the 9+ test of 1.6 (95% CI 1.0 to 2.2, P<0.001) from season 1 (22.2±4.1) to season 2 (23.8±3.3). However, the variability was large (ICC: 0.24, 95% CI 0.11 to 0.36) (figure 2 and Table 2)

Table 2 Inter-season characteristics of the 9+ total score (0-33) for all players and for injured versus uninjured groups.†

	n	Mean ±SD Test 1	Mean difference test 2 to test 1 (95% CI)	ICC (95% CI)	SEM‡	MDC‡
All	220	22.2±4.1	1.6 (1.0 to 2.2)*	.24 (.11 to .36)	3.1	8.7
Any injury						
Yes	136	22.1±4.0	2.0 (1.3 to 2.7)*	.25 (.09 to .40)	3.0	8.3
No	84	22.4±4.3	0.9 (-.1 to 1.9)	.24 (.02 to .43)	3.3	9.1
Lower extremity injury						
Yes	124	22.1±4.1	2.0 (1.2 to 2.7)*	.25 (.08 to .41)	3.1	8.5
No	96	22.3±4.2	1.1 (.1 to 2.0)*	.23 (.03 to .41)	3.2	8.8
Any severe injury						
Yes	40	21.5±4.1	2.9 (1.4 to 4.4)*	.16 (-.15 to .45)	3.3	9.1
No	180	22.3±4.1	1.3 (.6 to 1.9)*	.27 (.13 to .40)	3.1	8.5
Severe lower extremity injury						
Yes	36	21.6±4.2	3.0 (1.3 to 4.6)*	.13 (-.20 to .44)	3.4	9.5
No	184	22.3±4.1	1.3 (.7 to 1.9)*	.27 (.14 to .40)	3.0	8.4

†Mean ± SD for test 1 (season 1), mean inter-season difference and intraclass correlation coefficient (ICC) with 95% CI, and measurement error (SEM and MDC) from test 1 (season 1) to test 2 (season 2) are reported.
‡ Expressed in same unit as the measurement (9+ points)
* Significant at *P*< 0.05 (paired t-test)
CI, confidence interval
SEM, standard error of measurement
MDC, minimal detectable change
Severe injury: > 28 days absence

Among the 220 players, 136 (61.8%) players had a ≥1 time-loss injury between the two 9+ tests, predominantly to the lower extremity (n=124, 91.2% of all injured players). We observed a consistent improvement in 9+score across all subgroups, which tended to be greater for the injured than the uninjured groups, as seen in Table 2. The variability between season 1 and 2 was large across all injured (ICC=0.13 to 0.25) and uninjured groups (ICC=0.23 to 0.27). Players with a severe injury (> 28 days absence) displayed the greatest increase in mean total score between season 1 and season 2 (2.9±0.7, P<0.001). Again, the variability was large in this group (ICC=0.13 to 0.16), as illustrated in figure 3.

The SEM for the total score was large (3.0 to 3.4 points) across all groups, irrespective of injury and severity. The clinical applicability of the 9+ total score is limited, as indicated by the magnitude of the MDC (8.3 to 9.5 points), again irrespective of injury and severity (Table 2).

We performed a sub-analysis of players with a mean time between 9+ test of <1SD (294.2 days, n=32) and >1SD (425.1 days, n=27) than the average, and observed similar findings as described above. There was a significant increase in mean total score for the >1SD group of 2.1 (95% CI 0.40 to 3.82, P=0.017) from season 1 (21.3±3.8) to season 2 (23.4±4.2) whereas there was no significant increase in mean total score for the <1SD (0.21±0.8, P=0.79) from season 1 (22.8±4.2) to season 2 (23.0±2.8). However, the variability was again large for both >1SD group (ICC=0.35, 95% CI -0.03 to 0.64, SEM=3.1) and <1SD group (ICC=0.19, 95% CI -0.16 to 0.50, SEM=3.2).

Inter-season variability of each movement tests

There was a significant increase in score for each movement tests between season 1 and 2, apart from one-legged squat, deep one-legged squat, seated rotation and shoulder mobility (Table 3). However, the variability was large for all movement tests (κ_w=-0.003 to 0.63), irrespective of injury and severity.

Table 3 Weighted Kappa for each movement test for test 1 (season 1) to test 2 (season 2) of the 9+ for all players, and for injured versus uninjured groups.†

	All	**Any injury**		**LE injury**		**Severe Any injury**		**Severe LE injury**	
		Yes	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Exercises	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**	**κ_w (95% CI)**
Deep squat	_.16* (.07 - .26)_	_.15* (.03 - .28)_	.18 (.02 - .35)	_.16* (.04 - .30)_	_.16* (.00 - .31)_	_.11* (-.04 - .27)_	_.19* (.07 - .30)_	_.14*(-.02 - .35)_	_.17*(.05 - .28)_
One-legged squat	.24 (.13 - .34)	.25 (.09 - .40)	_.24* (.10 - .40)_	.26 (.11 - .41)	.21 (.05 - .35)	_.12 (-.15 - .38)_	.26 (.14 - .39)	.11 (-.14 - .40)	.26 (.15 - .38)
Deep one-legged squat	.26 (.16 - .36)	.19 (.06 - .32)	_.37* (.22 - .54)_	.20 (.07 - .33)	_.34* (.20 - .48)_	.15 (-.05 - 37)	.28 (.18 - .39)	.15 (-.08 - .36)	.28 (.18- .38)
In-line lunge	_.11* (.01 - .20)_	_.10* (-.02 - .21)_	_.12* (-.02 - .29)_	.10 (-.04 - .23)	_.12*(-.02 - .27)_	_.06* (-.10 - .26)_	_.12*(.01 - .22)_	_.08*(-.10 - .30)_	_.11*(.001- .21)_
Active hip flexion	_.34* (.24 - .43)_	_.39* (.27 - .50)_	.24 (.07 - .41)	_.38* (.26 - .52)_	.27 (.11 - .42)	_.31* (.10 - .49)_	.34 (.24 - .44)	_.29*(.10 - .50)_	_.34 (.24 - .45)_
Straight leg raises	_.32* (.22 - .43)_	.39 (.25 - .53)	_.20* (.04 - .37)_	_.35* (.21 - .50)_	_.26 (.10 - .40)_	.51 (.26 - .74)	_.27*(.16 - .40)_	.63 (.39 - .84)	_.25*(.14 - .37)_
Push up	_.08*(-.02 - .19)_	_.14* (.02 - .27)_	-.003 (-.18 - .18)	_.12* (-.01 - .25)_	.03 (-.12 - .20)	_.24* (.08 - .47)_	_.05*(-.07 - .16)_	_.23*(.07 - .46)_	_.06*(-.06 - .17)_
Diagonal lift	_.08*(-.01 - .18)_	_.07* (-.04 - .19)_	.10 (-.06 - .28)	_.03* (-.06 - .15)_	.14 (-.01 - .30)	_.11* (-.06 - .32)_	_.08*(-.02 - .19)_	_.09*(-.08 - .32)_	_.08*(-.02 - .20)_
Seated rotation	.35 (.24 - .46)	.41 (.27 - .56)	.25 (.09 - .44)	.41 (.27 - .57)	.27 (.09 - .49)	.25 (.01 - .49)	.37 (.24 - .49)	.25 (-.04 - .52)	.37 (.24 - .49)
Shoulder mobility	.37 (.28 - .49)	.33 (.20 - .45)	.45 (.28 - .61)	.34 (.20 - .47)	.42 (.26 - .57)	.20 (-.04 - .43)	.42 (.30 - .52)	.26 (.02 - .49)	.40 (.29 - .52)
Drop jump	_.19* (.08 - .30)_	_.26* (.12 - .40)_	_.07* (-.10 - .24)_	_.25* (.13 - .39)_	_.11*(-.05 - .25)_	_.13 (-.07 - .38)_	_.20* (.08 - .31)_	.17 (-.08 - 41)	_.19*(.08 - .32)_

† Weighted Kappa (κ_w) with 95% confidence interval (CI) from test 1 to test 2 is reported with significant results (P<0.05) from Wilcoxon signed rank test
Significant results (P< 0.05) are marked with "* and italics"

DISCUSSION

The main finding of this study was that there was a substantial intra-individual variability in the 9+ mean total score between the two consecutive seasons, irrespective of injury and severity status, and MDC was high across all groups. The inter-tester reliability was moderate. Additionally, there was a small but systematic improvement from one season to the next across all injured and uninjured groups.

The variability of the 9+ test

Frohm *et al*[11] examined the inter-rater and intra-rater reliability among eight trained observers of the 9+ in a group of male elite football players (n=26). They reported good intra-rater reliability (ICC 0.75, based on data from 18 players) with no systematic change when players were re-tested after 7 days, indicating that player and tester performance was stable across test sessions. Similarly good ICC scores have also been reported in several studies investigating the intra-rater and test-retest reliability of the FMS in physically active populations and college athletes re-tested after 2 to 7 days.[22-24]

We therefore assumed that in the absence of any intervention or injury, the 9+ total score would be stable (i.e. low variability) between seasons. The remarkably low ICC observed in our study, across injured and uninjured groups, suggests that the ability of the 9+ test to detect changes in functional movement patterns is very limited, largely because of the sizable measurement error. A similar tendency was also observed for each movement test, displaying consistently poor κ_w for all tests across all injured and uninjured groups.

An error in a measurement includes both rater variation, variation by chance and between-session variability in player performance.[25] The inter-tester reliability of our testers (overall ICC: 0.68) was lower on all of 9+ movement tests than those reported by Frohm *et al*,[11] except for seated rotation and shoulder mobility. Frohm *et al*[11] examined the inter-tester reliability in a small group of male football players (n=26) in a controlled research setting, using eight physiotherapists who were all experienced on the 9+. Our results may differ from those of Frohm *et al*[11] given that our testing was undertaken in a busy clinical screening setting using multiple testers (n=14) with less 9+ experience (than in Frohm's *et al* study). However, studies on the FMS have reported good inter-tester reliability for testers with varying experience.[22-24, 26-28] It is possible that in our screening setting some of the detailed movement criteria may have been missed, although all of our testers received the same initial 9+ training, and had similar clinical and 9+ experience. On the other hand, this increases the external validity of our findings.

The SEM in this study was large, ranging from 3.0 to 3.4 points across all groups independent of injury status, indicating that the 9+ total score has a normal variation (measurement error) of 3 to 4 points from season-to-season. Furthermore, the MDC ranged from 8.3 to 9.5 points indicating that a minimum improvement of 8 to 10 points is required to represent a real change in the 9+ test, again irrespective of injury and severity. Given our large SEM and MDC, it suggests that the 9+ total score inter-season variation is too large for the 9+ to detect change attributed to injury or clinical interventions.[19, 29] In other words, the large variability in the 9+ is mainly attributed to variability in player performance and chance rather than variability between testers.[17, 19] This view is substantiated by the difference in ICC values, which for the total score was 0.68 for between testers but only 0.24 between seasons.

There are several potential sources of random error that may help explain the observed variability in the 9+, including the motivation of the player, interpretation of the test instructions by the player or a learning effect.[30] Another possible explanation may be the ambiguity of the scoring criteria. The difficulty in assessing and performing the more complex tests involving multiple joints and complex physical qualities such as balance, coordination and core stability (i.e. the diagonal lift, in-line lunge, one-legged squat test) make scoring and performance uncertain, and subsequently will cause variability in both athlete performance and in the scoring (tester variability).[11, 27]

Clinical implications

Functional movement tests, including the 9+, are growing in popularity as an injury screening tool. Our results shows that there is a large variability in the 9+ total score and a change of above 8 points is necessary to represent a real change in a player’s 9+ test between seasons. Practitioners should consider this when interpreting the 9+ or similar functional movement screen scores. Our inter-tester reliability, using multiple testers, was moderate and practitioners are advised to perform their own reliability tests on their target population before considering the 9+ for clinical use.

The ability of the 9+ to predict injury is still unknown. However, the validity of the FMS as an injury prediction tool has been scrutinized recently, and with conflicting results.[8, 13, 31, 32] Based on the initial study by Kiesel *et al*[33] on the FMS in professional American football players, a total score below 67% was believed to represent an increased risk of injury.[33] However, a recent meta-analysis revealed that a cut-off of 67% only provided a sensitivity of 24.7% and a specificity of 85.7%, with an area under the curve (AUC) of 0.58, indicating that the overall predictive validity of the FMS is only slightly better than a 50/50 chance.[32]

Nevertheless, based on the study by Kiesel *et al*, a 9+ score below 67% (22 points) has been suggested as a possibly cut-off point for identifying players at increased risk of injury.[11] Given our SEM of 3 to 4 points, a player may be considered at risk one season and not the next season without any injury or intervention occurring. We therefore anticipate that the 9+ test will have limited value in predicting injury. Practitioners should therefore exercise caution using a 67% cut-off value when interpreting the 9+ total score as an injury-screening tool. Further studies are needed to confirm (or refute) the predictive validity of the 9+ test.

Methodological considerations

A major strength of this study was that it was undertaken in a real clinical athlete screening setting with a large group of male professional football players in one sports medicine hospital. A further strength of our study was the use of multiple testers. This provides good generalisability, but also might have influenced the inter-tester reliability adversely.

Limitations include that we did not record any prevention interventions occuring between the two test occasions. Also, this study was performed in a multi-national and multi-language setting. Although most of our testers spoke the same language as the players and we used pictures of the tests as described by Frohm *et al,*[11] it is possible that players did not understand the instructions given. This may have influenced the variability in the player performance of the 9+ score.[30] Finally, our study participants consisted of a homogeneous group of male professional football players in a specific setting which limits the generalisability of the findings to other sports, settings, age groups or women.

CONCLUSION

There was a substantial intra-individual variability of the 9+ total score between two consecutive seasons irrespective of injury and severity status. A change above 8 points between seasons is necessary to represent a real change in the 9+ test. Additionally, there was a small but systematic improvement from one season to the next across all injured and uninjured groups.

Twitter: Arnhild Bakken at @phbakken

ACKNOWLEDGMENTS

The authors would like to sincerely thank all the Aspetar staff involved in this study, especially the physiotherapists at the Rehabilitation Department and the Qatar National Sports Medicine Program (NSMP) who participated in the numerous 9+ functional movement assessments.

FUNDING

None.

ETHICAL APPROVAL

The study has been reviewed and approved by the Institutional Review Board, Anti-Doping Lab Qatar (ADLQ), Doha, Qatar

CONTRIBUTORSHIP STATEMENT

AB designed the study, contributed in data collection, analysed and interpreted the data, and drafted the article. RB designed the study, interpreted the data, revised the article and approved the final revision of the article. AF and RW contributed in data analysis, interpreted the data, revised the article and approved the final revision of the article. KMK, ST, TB, CE, JLT, and EW interpreted the data, revised the article and approved the final revision of the article.

COMPETING INTERESTS

One declared. Karim M Khan is Editor in Chief of BJSM and was at arms length (and blinded) from the review process in BJSM.

REFERENCES

1. Ekstrand J, Hagglund M, Walden M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med* 2011 Jun;45(7):553-8.

2. Eirale C, Farooq A, Smiley FA, et al. Epidemiology of football injuries in Asia: a prospective study in Qatar. *J Sci Med Sport* 2013 Mar;16(2):113-7.

3. Kuijt MT, Inklaar H, Gouttebarge V, et al. Knee and ankle osteoarthritis in former elite soccer players: a systematic review of the recent literature. *J Sci Med Sport* 2012 Nov;15(6):480-7.

4. Ljungqvist A, Jenoure PJ, Engebretsen L, et al. The International Olympic Committee (IOC) consensus statement on periodic health evaluation of elite athletes, March 2009. *Clin J Sport Med* 2009 Sep;19(5):347-65.

5. Hegedus EJ, McDonough S, Bleakley C, et al. Clinician-friendly lower extremity physical performance measures in athletes: a systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests. *Br J Sports Med* 2015 May;49(10):642-8.

6. Hegedus EJ, McDonough SM, Bleakley C, et al. Clinician-friendly lower extremity physical performance tests in athletes: a systematic review of measurement properties and correlation with injury. Part 2--the tests for the hip, thigh, foot and ankle including the star excursion balance test. *Br J Sports Med* 2015 May;49(10):649-56.

7. Dennis RJ, Finch CF, Elliott BC, et al. The reliability of musculoskeletal screening tests used in cricket. *Phys Ther Sport* 2008 Feb;9(1):25-33.

8. Wright AA, Stern B, Hegedus EJ, et al. Potential limitations of the Functional Movement Screen: a clinical commentary. *Br J Sports Med* 2016 Mar 31.

9. Beckerman H, Roebroeck ME, Lankhorst GJ, et al. Smallest real difference, a link between reproducibility and responsiveness. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2001;10(7):571-8.

10. Hegedus EJ, Cook CE. Return to play and physical performance tests: evidence-based, rough guess or charade? *Br J Sports Med* 2015 Oct;49(20):1288-9.

11. Frohm A, Heijne A, Kowalski J, et al. A nine-test screening battery for athletes: a reliability study. *Scand J Med Sci Sports* 2012 Jun;22(3):306-15.

12. Frohm A, Flodstrom F, Kockum B. 9+ Screening Batteri - plusswebb. 2013 [cited 2016 14 August]; Available from: https://utbildning.sisuidrottsbocker.se/sisu/generell/idrottsskador/9-screening-batteri/tester/.

13. Bonazza NA, Smuin D, Onks CA, et al. Reliability, Validity, and Injury Predictive Value of the Functional Movement Screen: A Systematic Review and Meta-analysis. *The American Journal of Sports Medicine* 2016 April 29, 2016.

14. Moran RW, Schneiders AG, Major KM, et al. How reliable are Functional Movement Screening scores? A systematic review of rater reliability. 2015 Aug 27.

15. Bakken A, Targett S, Bere T, et al. Health conditions detected in a comprehensive periodic health evaluation of 558 professional football players. *Br J Sports Med* 2016 Mar 24.

16. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Clin J Sport Med* 2006 Mar;16(2):97-106.

17. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005 Feb;19(1):231-40.

18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 Feb 8;1(8476):307-10.

19. Koumantakis GA, Winstanley J, Oldham JA. Thoracolumbar proprioception in individuals with and without low back pain: intratester reliability, clinical applicability, and validity. *J Orthop Sports Phys Ther* 2002 Jul;32(7):327-35.

20. Portney LG, Watkins MP. Foundations of clinical research. 2nd ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc; 2000.

21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159-74.

22. Shultz R, Anderson SC, Matheson GO, et al. Test-retest and interrater reliability of the functional movement screen. *J Athl Train* 2013 May-Jun;48(3):331-6.

23. Teyhen DS, Shaffer SW, Lorenson CL, et al. The Functional Movement Screen: a reliability study. *J Orthop Sports Phys Ther* 2012 Jun;42(6):530-40.

24. Smith CA, Chimera NJ, Wright N, et al. Interrater and Intrarater Reliability of the Functional Movement Screen. *J Strength Cond Res* 2012 Jun 11.

25. Holmefur M, Aarts P, Hoare B, et al. Test-retest and alternate forms reliability of the assisting hand assessment. *J Rehabil Med* 2009 Nov;41(11):886-91.

26. Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. *J Strength Cond Res* 2012 Feb;26(2):408-15.

27. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. *J Strength Cond Res* 2010 Feb;24(2):479-86.

28. Gribble PA, Brigle J, Pietrosimone BG, et al. Intrarater reliability of the functional movement screen. *J Strength Cond Res* 2013 Apr;27(4):978-81.

29. Waldron M, Gray A, Worsfold P, et al. The reliability of Functional Movement Screening (FMS) and in-season changes in physical and performance among elite rugby league players. *J Strength Cond Res* 2014.

30. Frost DM, Beach TA, Callaghan JP, et al. FMS Scores Change With Performers' Knowledge of the Grading Criteria-Are General Whole-Body Movement Screens Capturing "Dysfunction"? *J Strength Cond Res* 2015 Nov;29(11):3037-44.

31. McCunn R, Aus der Funten K, Fullagar HH, et al. Reliability and Association with Injury of Movement Screens: A Critical Review. *Sports Med* 2015 Dec 31.

32. Dorrel BS, Long T, Shaffer S, et al. Evaluation of the Functional Movement Screen as an Injury Prediction Tool Among Active Adult Populations: A Systematic Review and Meta-analysis. *Sports Health* 2015 Nov-Dec;7(6):532-7.

33. Kiesel K, Plisky PJ, Voight ML. Can Serious Injury in Professional Football be Predicted by a Preseason Functional Movement Screen? *N Am J Sports Phys Ther* 2007 Aug;2(3):147-58.

FIGURE LEGENDS

Figure 1 The flow of players included in the study. MSK, musculoskeletal
Figure 2 Bubble plot presenting the total score for all players (n=220) on season 1 (test 1) and season 2 (test 2). The bubble size depicts the number of players with identical total score; the smallest points represent one player, the largest seven players. The hatched line represents the identity line.
Figure 3 Bubble plot presenting the total score for players with lower extremity injury (n=124) and severe lower extremity injury (n=36) on season 1(test 1) and season 2 (test 2). The bubble size depicts the number of players with identical score; the smallest points represent one player, the largest five players. The hatched line represents the identity line.

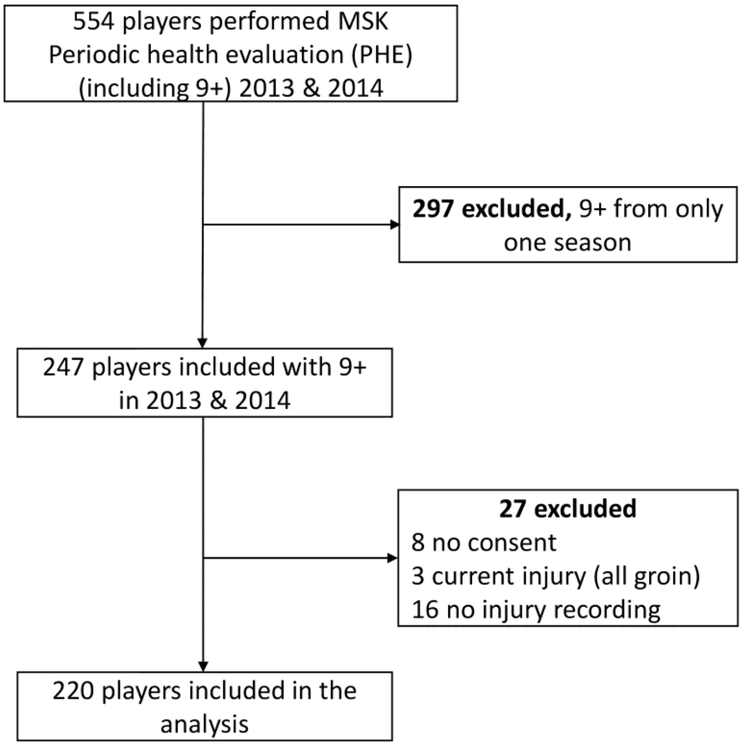

Figure 1 The flow of players included in the study. MSK, musculoskeletal

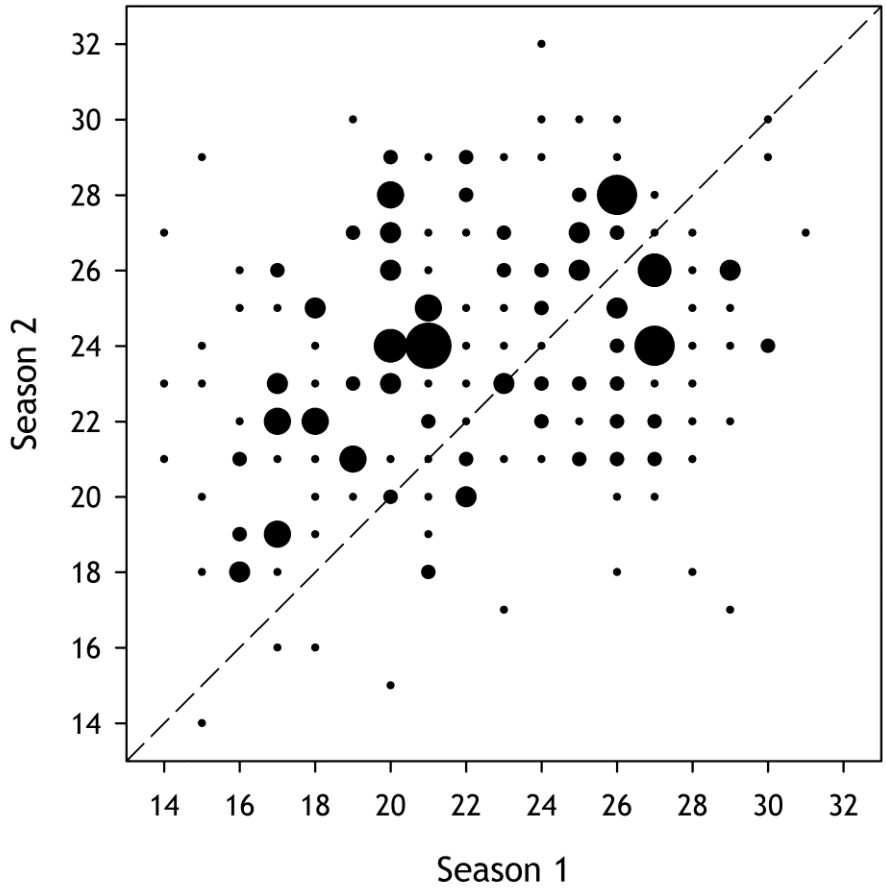

Figure 2 Bubble plot presenting the total score for all players (n=220) on season 1 (test 1) and season 2 (test 2). The bubble size depicts the number of players with identical total score; the smallest points represent one player, the largest seven players. The hatched line represents the identity line.

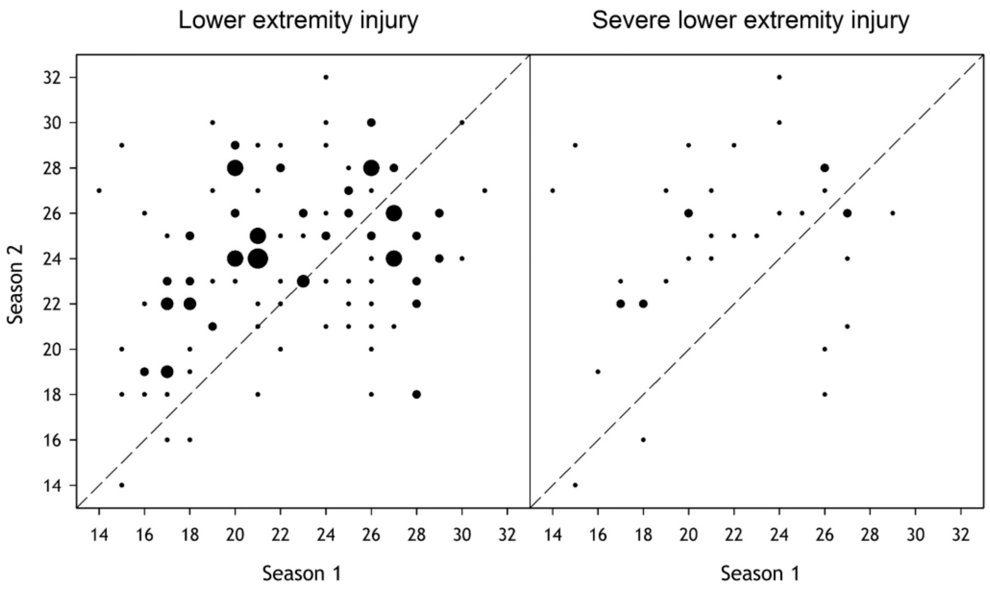

Figure 3 Bubble plot presenting the total score for players with lower extremity injury (n=124) and severe lower extremity injury (n=36) on season 1(test 1) and season 2 (test 2). The bubble size depicts the number of players with identical score; the smallest points represent one player, the largest five players. The hatched line represents the identity line.