SAARLAND UNIVERSITY

Institute of Sports and Preventive Medicine

ASSOCIATION WITH INJURY OF A FOOTBALL-SPECIFIC MOVEMENT SCREEN

By

Robert McCunn

This thesis is presented for the award of a Doctor of Philosophy (Sports Medicine) from the Medical Faculty, Saarland University, Saarbrücken, Germany

PREFACE

I, Robert McCunn, declare that this thesis, is submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy in the Institute of Sport and Preventive Medicine, Saarland University, and is wholly my own work unless otherwise referenced or acknowledged. As such, I also I certify to the best of my knowledge and belief that this thesis does not:

i. incorporate any material previously submitted for a degree or diploma in any institution of higher education without acknowledgement;

ii. contain any material previously published or written by another person without acknowledgement; or

iii. contain any defamatory material

I also grant permission for the library at Saarland University to produce duplicate copies of my thesis as required.

Robert McCunn

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This thesis represents the culmination of almost 10 years of university education, many more years of schooling and several years of applied experience in the field. I highlight this only to acknowledge the fact that many individuals who have helped and supported me along the way may not be explicitly mentioned by name here; however, I am deeply grateful to everyone who has facilitated me reaching this point.

Firstly, I wish to thank Professor Tim Meyer, Dr Karen aus der Fünten and Dr Sabrina Skorski for selecting me as the successful candidate for the PhD scholarship in the first place. You provided me with an incredible opportunity and allowed me to enjoy what has ultimately proved to be an amazing and life-changing experience living and studying in Germany. Indeed, I would also like to thank everyone at the institute who has helped me along the way.

To my fellow PhD students past and present in the 'Büro': I'm not sure I could have done this without you. As you all know, when you join our international programme you become more than a peer or even a friend, it creates a bond. I wish you all the best and will always be ready with a helping hand. In particular I would like to mention Dr Hugh Fullagar, Ross Julian and Emiel Schulze: thank you for everything.

I have been so very lucky to meet, collaborate and become friends with some amazing sport science researchers and practitioners over the last few years. There are too many of you to mention here; however, I would like to especially thank Dr John Sampson and Matthew Whalan. John and Matt, I can honestly say that without you both I don't know where I'd be. Thank you so much for hosting me in Australia and helping me conduct the flagship study within this body of work. Thank you for your friendship and support: material, academic and personal. I'll never forget it. I must also make honourable mention of the entire Fullagar family for providing me with a place to truly call home while in Australia – thank you. I also wish to acknowledge Dr Alan McCall: thank you for your mentorship and friendship over the last few years as well as the incredible professional opportunities you continue to provide me with.

Throughout my academic and professional career I have been fortunate to have an unwavering and close confidant. To thank him here doesn't do him justice; however, I wish to acknowledge Neil Gibson for his unbridled support. Neil, you have been, and continue to be, a mentor and dear friend to whom I will always be grateful. Thank you for taking me under your wing all those years ago and helping me become the researcher, practitioner and man I am today.

To Mum, Dad and Sandy: in everything I do I aspire to make you proud. Thank you for your love and support. Through the highs and lows, you've always been there and I've always known I could count on you. To finally be here, at this point in the process, writing these acknowledgments, feels surreal! Thank you for *always* believing in me.

Finally, to my grandfather, Archie: the most impressive and humble man I am ever likely to meet. You were always a proponent of education despite being forced to leave school at 14 years old to help support your family. I hope this achievement is a realisation of your ideals. Thank you for instilling in me the desire to be the best I can be. Summa Petenda.

ABBREVIATIONS

+ve	Positive
16-PPM	Physical Performance Measures Screen
AAA	Athletic Ability Assessment
AC1	First order agreement coefficient
ACL	Anterior cruciate ligament
AR	Anterior reach
AT	Athletic trainer
AUC	Area under the curve
CI	Confidence interval
CL	Confidence limit
cm	Centimetres
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CSMT	Conditioning Specific Movement Tasks
DVJ	Drop vertical jump
ES	Effect size
ExpAT	Experienced athletic trainer
F	Female
FIFA	Fédération Internationale de Football Association
FMS	Functional Movement Screen
GLM	Generalized linear model
ICC	Intraclass correlation coefficient
ILL	In-line lunge
IPL	Illawarra Premier League
κ	Kappa
Kg	Kilogram
LESS	Landing Error Scoring System
LR	Likelihood ratio
M	Male
MDC	Minimal detectable change
n	Number of (participant sample size)
NA	Not applicable

NCAA	National Collegiate Athletic Association
NMST	Netball Movement Screening Tool
NPL	National Premier League
NSCA	National Strength & Conditioning Association
NSW	New South Wales
OR	Odds ratio
ROC	Receiver operating characteristic
RR	Relative risk (or risk ratio)
S&C	Strength and conditioning
SD	Standard deviation
SEBT	Star Excursion Balance Test
SIMS	Soccer Injury Movement Screen
SIGN	Scottish Intercollegiate Guidelines Network
SLDL	Single-leg deadlift
SLHD	Single-leg hop for distance
T&F	Track and field
TJ	Tuck jump
UEFA	Union of European Football Associations
UKSCA	United Kingdom Strength & Conditioning Association

TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

FIGURES

Chapter 1

Chapter 3

LIST OF PUBLICATIONS RELEVANT TO THIS THESIS

In chronological order:

1. **McCunn R**, aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 2016;46(6):763-781.

2. **McCunn R**, Meyer T. Screening for risk factors: If you liked it then you should have put a number on it. *Br J Sports Med*. 2016;50:1354.

3. **McCunn R**, Sampson JA, Whalan M, Meyer T. Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement? *Science and Medicine in Football*. 2017;1(1):86-88.

4. Newton F, McCall A, Ryan D, Blackburne C, aus der Fünten K, Meyer T, Lewin C, **McCunn R**. Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players. *Science and Medicine in Football*. 2017;1(2):102-6.

5. **McCunn R**, aus der Fünten K, Govus A, Julian R, Schimpchen J, Meyer T. The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS). *Int J Sports Phys Ther*. 2017;12(1):53-66.

Under review:

6. **McCunn R**, aus der Fünten K, Whalan M, Sampson JA, Meyer T. Soccer Injury Movement Screen (SIMS) composite score is not associated with injury among semi-professional soccer players.

ABSTRACT

INTRODUCTION: Movement screening is widely used within football to quantify players' movement quality. An established definition of movement quality does not exist; however, it encompasses aspects typically associated with safe exercise technique such as maintenance of neutral posture, balance and coordination. The underlying theory behind movement screening is that 'poor' movement quality increases the risk of injury whereas 'good' movement quality reduces the risk. Indeed, current convention advocates for the risk stratification of individuals based on movement screening scores with intervention targeted only to the high-risk group. The appeal of such practice is clear: potentially more efficient allocation of resources. However, very little evidence to support this approach exists within football. Consequently, the aims of this thesis were as follows: i) to determine what movement screens currently exist within the scientific literature, how reliable they are and their potential association with injury, ii) investigate the predictive ability of the most widely used and researched movement screen within a football population, iii) develop a new football-specific movement screen and test its reliability, and iv) establish the association with injury of the newly created football-specific movement screen.

METHODS: i) A structured literature review was conducted to identify the movement screens with supporting evidence regarding their reliability and association with injury. ii) The Functional Movement Screen (FMS™) was identified as the most widely used and researched movement screen; however, a paucity of research involving football players was observed. In order to investigate the predictive ability of the FMS™ within football 84 elite youth players from a professional club academy performed the screen during the pre-season period and were subsequently observed during the in-season period and injury incidence recorded. iii) The Soccer Injury Movement Screen (SIMS) was developed as a football-specific movement screen. The constituent movements making up the SIMS were selected based on the most common injury locations (lower-body) and types (muscle strains and ligament sprains). To assess the intra- and inter-rater reliability of the SIMS, 25 recreational athletes performed the assessment and were scored by three raters on three separate occasions. iv) The final investigation established the association with injury of the SIMS within a football population. Utilising a prospective cohort study design, 306 semi-professional football players performed the SIMS during the pre-season period and were subsequently observed during the in-season period with injury incidence and exposure time recorded.

RESULTS: i) A total of 10 movement screens were identified by the structured literature search. The majority of the identified screens demonstrated acceptable reliability for use in applied practice and future research. However, only two of the 10 identified screens had any supporting evidence regarding their association with injury. Furthermore, the limited available evidence related to association with injury was equivocal and not sufficient to justify any movement screen as 'predictive'. ii) No association with injury was observed for the FMS™ within a cohort of elite youth football players. iii) The SIMS demonstrated good to excellent intra- and inter-rater reliability. iv) However, no association with injury was observed for the SIMS composite score in relation to any of the categories investigated.

DISCUSSION: While many reliable movement screens exist none have compelling evidence supporting a strong association with injury. Despite its widespread use within football the FMS™ was not associated with injury among a cohort of elite youth players questioning its value in terms of injury prediction. The SIMS demonstrated good to excellent reliability indicating its suitability for use in applied practice and future research; however, no association with injury was observed for the composite score. The present results challenge current convention that advocates risk stratification and targeted intervention based on screening score. Implementing universal injury prevention programmes albeit with content informed by prospective studies such as the ones conducted within this body of work are recommended ahead of attempting to identify specific 'at-risk' individuals.

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Why is injury an important issue in football?

Scientific research related to injury within football is very prevalent. Much epidemiological research has been conducted at the professional level of the game revealing the lower extremities as the most common site of injury with muscle strains and ligament sprains the most prominent types.[13, 14, 31] Injury incidences reported within the scientific literature vary depending on the level of play and population investigated; however, a professional team of 25 players can realistically expect about 50 injuries each season.[14] Equally, it is commonplace to see mainstream media outlets, including national newspapers and television broadcasters, report on injuries suffered by high-profile players. Indeed, scientific studies have even used data collated by the mainstream media to establish injury incidence in professional leagues, alluding to the extensive coverage of, and public interest in, the topic.[21] However, an important question is: why is injury such an important issue?

There are numerous arguments as to why injury within football is worthy of consideration by the scientific community. Not least, there is an obvious legal and moral obligation on governing bodies, national football associations, professional clubs and the staff they employ to protect the welfare of players. As in any professional occupation, football included, health and safety law applies.[33] In order to appropriately address the risk of injury associated with playing football it is first necessary to establish the extent of the problem: injury incidence.[30] Secondly, the aetiology of injury needs established: mechanism and risk factor identification.[30] Thirdly, interventions designed to reduce the established risks need applied and finally, evaluation of their effectiveness.[30] Sports injury research has and continues to allow these steps to be applied. This on going cycle of research ultimately allows applied practitioners to be evidence-based in their efforts to reduce injury incidence.

1.1.1 Player health

Suffering a football related injury has an obvious acute impact on players' health; however, such circumstances may also result in long-term sequelae such as osteoarthritis.[25] In addition, once a player has suffered an injury their risk of sustaining it again in the future (re-injury) is increased.[1, 13] Indeed, the increased risk related to previous injury is not limited to a recurrence of the index injury but a range of subsequent injury types and locations.[15, 27] Furthermore, the effects of sports injury are not limited to physical implications. Sports injury can also elicit symptoms related to depression and anxiety.[23] Protecting players' immediate and long-term physical and mental health should be a priority for applied practitioners working within football, as both the Union of European Football Associations (UEFA) and Fédération Internationale de Football Association (FIFA) advocate.[9]

1.1.2 Competitive advantage

A competitive incentive also exists for reducing injury incidence. Hägglund et al.[17] demonstrated that lower injury burden and higher player availability were associated with an increase in UEFA season club coefficient: an indicator of success in the Champions League or Europa League. Injury incidence was also negatively correlated with team ranking, number of games won, number of goals scored and total points within Qatari first division clubs.[10] Similarly, Carling et al.[6] reported the lowest squad utilization and lower injury occurrence in a championship-winning season compared to less successful preceding and subsequent seasons in a professional team. Intuitively, maximizing the number of games a club can field its strongest team seems likely to improve its chances of success.

1.1.3 Financial incentive

The financial incentive related to reducing injury incidence is substantial. The financial cost associated with player unavailability includes not only the salary of the individual but also potential medical treatments. While the specific monetary value will vary depending on player wages and the medical treatment sought it has been estimated that the average cost of a player being absent for a month due to injury equates to €500,000 within a top-level European club.[11] Even at the non-professional level, football related injury may have wider financial implications to society through lost work days and additional burden to public healthcare systems.

1.2 Movement screening within football

Periodic health examinations are widespread throughout sport, including football.[2, 22] These regular medical assessments often include a form of movement screening. Movement screening is widely used within professional football.[24] It involves the assessment of movement quality and seeks to quantify *how well* an individual moves. An established definition of movement quality does not exist; however, it encompasses aspects typically associated with safe exercise technique such as maintenance of neutral posture, balance and coordination. As such, it differs from most fitness tests that are generally quantitative in nature and measure performance in units such as centimeters, seconds and kilograms. The judgment of 'quality' inherently introduces an element of subjectivity since the score ultimately relies on the opinion of the test rater. The underlying theory behind movement screening is that 'poor' movement quality increases the risk of injury whereas 'good' movement quality reduces the risk. However, it should be noted that very little evidence to support this assertion exists within football. Despite this lack of evidence, a survey of 44 top-division professional football clubs revealed that 82% of them used movement screening as a tool to identify injury risk.

Considering the numerous compelling reasons as to why reducing injury incidence within football is a worthwhile endeavor, the appeal of predictive tools that can alert practitioners to players who may be 'high-risk' is obvious. In theory, using the results of movement screening to create high- and low-risk (of injury) player groups allows applied practitioners to focus their attention and resources on the individuals they can help the most. The logic behind this approach is related to optimising efficiency, of both time and material resources. However, an evidence base supporting the efficacy of such practice – risk stratification leading to individualized programming – in terms of effectively reducing injury incidence as opposed to implementing universal injury prevention interventions is lacking, especially within the sport of football.[2, 29]

1.3 Thesis structure and research aims

The following thesis comprises several chapters, each addressing a specific research question or contemporary issue related to movement screening, its association with injury within football or injury data collection procedures within non-professional environments. Each of these topics were explored with the intention of addressing the overall research aims. Given the prevalence of movement screening within football and the lack of evidence supporting its usefulness as an injury prediction tool the following research aims were established:

1) Determine what movement screens currently exist within the scientific literature, how reliable they are and their potential association with injury.

2) Investigate the injury predictive ability of the most widely used and researched movement screen within a football population.

3) Develop a new football-specific movement screen and test its reliability.

4) Using a prospective cohort study design, establish the association with injury of the newly created football-specific movement screen.

5) Discuss the methodological challenges related to collecting injury data in non-professional football populations.

6) Discuss how movement screening may be useful to applied practitioners outwith the context of injury prediction.

Each chapter of this thesis represents these research aims, in sequential order.

1.4 Literature review

The following section contains the submitted manuscript pertaining to the following publication:

McCunn R, aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 2016;46(6):763-781.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of *Sports Medicine*. The numerical citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

Reliability and association with injury of movement screens: A critical review

Robert McCunn[1], Tim Meyer[1], Hugh H. K. Fullagar[1, 2], Ian McKeown[3], Karen aus der Fünten[1]

[1]*Institute of Sport and Preventive Medicine, Saarland University, Germany*

[2] *Sport & Exercise Discipline Group, UTS: Health, University of Technology, Sydney, Australia*

[3]*Port Adelaide Football Club, Adelaide, Australia*

Corresponding author

Robert McCunn

Institute of Sport and Preventive Medicine, Saarland University, Geb. B8.2

66123 Saarbrücken, Germany

Email: robert.mccunn@uni-saarland.de

Phone: +49681-302 70400 Fax: +49681-302 4296

Running head: "Movement screens: reliability and injury risk"

Abstract

Subjective assessment of athletes' movement quality is widely used by physiotherapists and other applied practitioners within many sports. One of the beliefs driving this practice is that individuals who display 'poor' movement patterns are more likely to suffer an injury than those who do not. The aim of this review was to summarize the reliability of the movement screens currently documented within the scientific literature and explore the evidence surrounding their association with injury risk. Ten assessments with accompanying reliability data were identified through the literature search. Only two of these 10 had any evidence directly related to injury risk. A number of methodological issues were present throughout the identified studies including: small sample sizes, lack of descriptive rater or participant information, ambiguous injury definitions, lack of exposure time reporting and risk of bias. These factors combined with the paucity of research on this topic make drawing conclusions as to the reliability and predictive ability of movement screens difficult. None of the

movement screens that appear within the scientific literature currently have enough evidence to justify the tag of 'injury prediction tool'.

Key Points

- Subjective assessment of athletes' movement quality is commonplace within professional sport, often in an attempt to predict injury risk.
- Of the 10 movement screens identified within the scientific literature only two have had their injury predictive ability investigated via prospective cohort studies.
- None of the movement screens present within the scientific literature currently have enough supporting evidence to justify being heralded as 'injury prediction tools', however, they may well provide practitioners with greater holistic understanding of their athletes' physical capabilities.

1 Introduction

The use of fitness assessments to profile and categorize athletes' physical capabilities is commonplace and is a central aspect of many applied practitioners' jobs [1]. The data collected from traditional fitness tests are typically objective in nature i.e. can be measured in units such as seconds, centimeters or grams. Movement screening is a type of assessment frequently used within professional soccer as well as other sports and is predominantly a subjective process that aims to measure the 'quality' of a movement pattern [2]. However, for various reasons including the subjective nature of such assessments and its relatively recent adoption by practitioners this practice has received limited attention within the scientific literature. A consensus on what defines movement quality is not available; however, the concept encapsulates the maintenance of correct posture and joint alignment in addition to balance while performing the selected movements. While some sporting institutions measure intuitively related parameters such as strength and joint range of motion, movement quality has been identified as an independent attribute [3, 4]. Therefore, a fitness testing battery that seeks to build a comprehensive profile of an athlete should incorporate an assessment of movement quality. This highlights the need for reliable and valid movement screening tools.

The foundation of a comprehensive injury prevention program is identifying individuals with a high risk of injury [5] and this is one of the key concepts underpinning the practice of movement screening. If athletes who display 'poor' movement patterns have a greater risk of injury than those who display 'good' movement patterns then screening protocols may be an important component of injury prevention strategies. However, the purpose of movement screens is not to diagnose *why* a poor movement pattern exists but simply to highlight it [6]. It is up to the judgment of the practitioner as to what course of action, if any, is taken in response to the outcome. Furthermore, elite sport is not the only environment where movement ability is important. Giblin et al. [7] stated that fundamental movement ability (core stability, balance, coordination) is related to perceived competence and confidence associated with physical activity. As such, movement quality is linked to general health as well as sports performance.

Despite movement quality being an important skill for the general population, in addition to athletes, measuring it is problematic due to its subjective nature. A variety of movement screens exist, the most well known being the functional movement screen (FMS) [6, 8]. The FMS has received attention from researchers and different aspects of this protocol such as its reliability and association with injury have been investigated. However, other screens do exist with some but not all appearing within the scientific literature. A collective critique of the movement screens detailed in the scientific literature does not currently exist and this is necessary to raise awareness of the available options. This will allow practitioners to make informed decisions about which, if any, movement screen is most appropriate for them. Accordingly, the aim of the present review was to summarize the intra- and inter-rater reliability of the available movement screens and discuss the evidence surrounding their ability to determine injury risk.

2 Literature search

In order to accomplish this critical review, a computerized literature search (figure 1) was performed with PubMed, Web of Science and ScienceDirect for articles published up until 1st July 2015 using the search terms "movement", "screen", "screening", "reliability", "injury", "prediction", "predicts", "landing error scoring system", "tuck jump assessment", "functional movement screen", "functional movement screening", "single leg squat test", "squat", "test", "drop jump", "drop vertical jump" and "movement quality" in various combinations. In addition, articles were identified manually from the reference lists of original manuscripts. A total of 51 relevant articles were identified. For the purpose of this review a movement screen was defined as a protocol designed for use with apparently healthy, uninjured individuals to primarily assess the 'quality' of a movement(s) rather than objective outcomes such as number of repetitions, distance or time achieved. The movement(s) included should rely on multiple physical qualities to execute correctly e.g. strength, balance and flexibility. It is not used to identify specific clinical conditions and does not require interpretation by a medical professional.

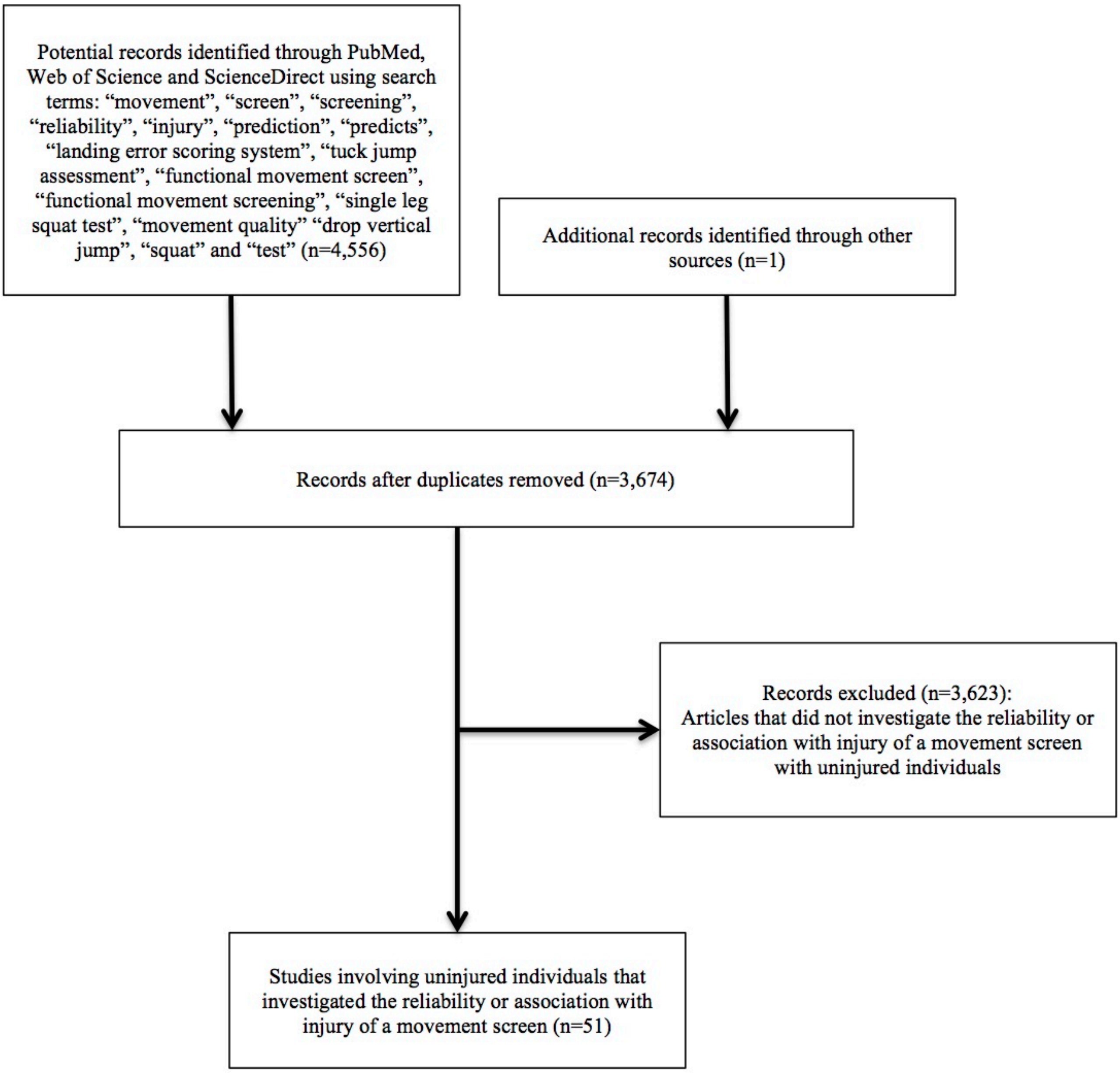


Figure 1. Flow diagram showing the movement screens present within the scientific literature.

3 Reliability of identified screens

A total of 10 movement screens that met the definition outlined above and with accompanying reliability data were identified through the literature search (figure 2). These screens consisted of the: FMS, landing error scoring system (LESS), single-leg squat screen variations, drop vertical jump screen variations, tuck jump assessment, athletic ability assessment (AAA), conditioning specific movement tasks (CSMT), the netball movement screening tool (NMST), the physical performance measures screen (16-PPM) and the star excursion balance test (SEBT) movement quality screen. A description of the exercises involved in each screen is provided in table 1.

The reliability of an assessment tool is paramount since it is a pre-requisite for test validity [9]. As such, before any given movement screen can be investigated with respect to injury prediction it must first be demonstrated that the test is reliable. Throughout the 51 articles identified by the literature search intra class correlation coefficients (ICC) were commonly reported. Atkinson and Nevill [10] stated that various qualitative interpretations of ICC values exist yet none were related to "analytical goals for research" and so it is difficult to say exactly what value constitutes 'good' or 'excellent' reliability. Some of the identified studies classified an ICC value of ≥0.75 as good [11] whereas others [12, 13] classified scores of ≥0.80 and ≥0.90 as good and excellent respectively. Shultz et al. [14] reported ICC values of 0.40-0.75 as fair to good and >0.75 as excellent. A reasonable consensus as to what can be considered good reliability appears to be an ICC ≥0.75 and thus this classification will be used throughout this review. In addition to ICCs, kappa values were also often reported and guidelines presented by Landis & Koch [15] will be used to classify these scores. Accordingly, the strength of agreement was considered poor, slight, fair, moderate, substantial or almost perfect for kappa scores of <0.00, 0.00-0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80 and 0.81-1.00 respectively.

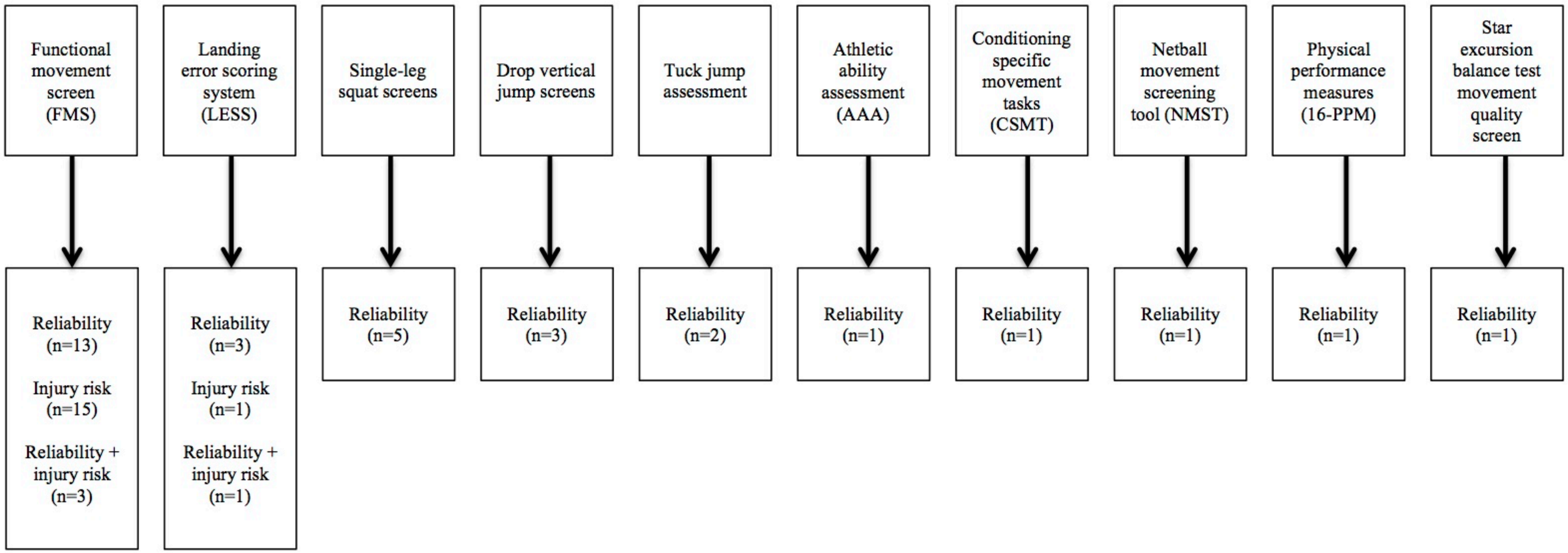

Functional movement screen (FMS)
Landing error scoring system (LESS)
Single-leg squat screens
Drop vertical jump screens
Tuck jump assessment
Athletic ability assessment (AAA)
Conditioning specific movement tasks (CSMT)
Netball movement screening tool (NMST)
Physical performance measures (16-PPM)
Star excursion balance test movement quality screen
Reliability (n=13) Injury risk (n=15) Reliability + injury risk (n=3)
Reliability (n=3) Injury risk (n=1) Reliability + injury risk (n=1)
Reliability (n=5)
Reliability (n=3)
Reliability (n=2)
Reliability (n=1)
Reliability (n=1)
Reliability (n=1)
Reliability (n=1)
Reliability (n=1)

Table 1. Content of the identified movement screens

Screen name	Number of exercises	Name of exercises	Protocol description
Functional movement screen	7	Deep squat Hurdle step In-line lunge Shoulder mobility Active straight-leg raise Trunk stability push up Rotary stability	Cook et al. (2014) [8] Cook et al. (2014) [6]
Landing error scoring system	1	Drop jump	Padua et al. (2009) [34]
Single-leg squat screens			
Single-leg squat task	1	Single-leg half squat	Crossley et al. (2011) [36]
Single-leg mini squat	1	Single-leg half squat	Ageberg et al. (2010) [38]
Unilateral lower extremity functional tasks	2	Single-leg half squat Lateral step down	Chmielewski et al. (2007) [39]
Drop vertical jump screens	1	Drop vertical jump	Nilstad et al. (2014) [44] Whatman et al. (2013) [45] Ekegren et al. (2009) [46]
Tuck jump assessment	1	Tuck jump	Myer et al. (2008) [48]
Athletic ability assessment	9	Prone hold on hands Lateral hold on hands Overhead squat Single-leg squat off box Walking lunge Single-leg forward hop Lateral bound Push ups Chin ups	McKeown et al. (2014) [53]

Table 1. Continued

Screen name	Number of exercises	Name of exercises	Protocol description
Conditioning specific movement tasks	6	Overhead squat Romanian deadlift Single-leg squat Double-leg to single-leg landing Sprint (40m) Countermovement jump	Parsonage et al. (2014) [54]
Netball movement screening tool	10	Squat Lunge & twist Bend & pull Push up Single-leg squat Vertical jump (land on both legs) Vertical jump (land on one leg) Broad jump Star excursion balance test Active straight leg raise	Reid et al. (2014) [55]
Physical performance measures	16	Broad jump Closed kinetic chain upper extremity stability test Y-balance test In-line lunge for distance Lateral lunge for distance Lumbar endurance Side plank hip abduction Side plank hip adduction Triple hop for distance Nordic hamstring Full squat Downward dog Single-leg squat Shoulder mobility test Active straight leg raise Beighton hypermobility	Tarara et al. (2014) [56]
Star excursion balance test movement quality screen	1	Anterior reach	Ness et al. (2015) [58]

3.1 Methodological quality assessment for reliability studies

The methodological quality of each paper that reported the reliability of a movement screen was assessed using the *Co*nsensus-based *S*tandards for the selection of health *M*easurement *In*struments (COSMIN) checklist [16]. This checklist utilizes a four-point scoring system (poor, fair, good, excellent) and contains sub-sections relating to numerous aspects of study design. For the purposes of this critique only the reliability section (box B), containing 14 questions, was addressed. As per the checklist instructions, if the answer to any of the 14 questions was 'poor' then, based on the 'worst score counts' principle, the study was classified as such. The outcomes of this methodological quality assessment are presented in table 2 (see Electronic Supplementary Material Appendix S1 for raw data).

Table 2. Methodological quality (according to COSMIN checklist) of reliability studies

Movement screen	Methodological quality rating			
	Poor	Fair	Good	Excellent
FMS	Teyhen et al. (2012) intra-rater [24] Hotta et al. (2015) [17] Gulgin & Hoogenboom (2014) [18] Letafatkar et al. (2014) [19] Parenteau-G et al. (2014) [12] Elias (2013) [20] Gribble et al. (2013) [21] Smith et al. (2013) [11] Frohm et al. (2012) [22] Klusemann et al. (2012) [23] Onate et al. (2012) [13] Schneiders et al. (2011) [26] Chorba et al. (2010) [27]	Shultz et al. (2013) [14] Butler et al. (2012) [25] Minick et al. (2010) [28]	Teyhen et al. (2012) inter-rater [24]	
LESS	Smith et al. (2012) [31] Onate et al. (2010) [33]	Padua et al. (2011) [32]	Padua et al. (2009) [34]	
Single-leg squat screens	Örtqvist et al. (2011) inter-rater [37] Crossley et al. (2011) [36] Ageberg et al. (2010) [38] Chmielewski et al. (2007) [39]	Örtqvist et al. (2011) intra-rater [37]	Junge et al. (2012) [35]	
DVJ screens	Whatman et al. (2013) [45] Ekegren et al. (2009) [46]	Nilstad et al. (2014) [44]		
Tuck jump assessment	Herrington et al. (2013) [51]	Dudley et al. (2013) [52]		
AAA	McKeown et al. (2014) [53]			
CSMT		Parsonage et al. (2014) [54]		

Table 2. Continued

Movement screen	Methodological quality rating Poor	Fair	Good	Excellent
NMST	Reid et al. (2014) [55]			
16-PPM	Tarara et al. (2014) [56]			
SEBT movement quality screen		Ness et al. (2015) [58]		

AAA Athletic ability assessment, *COSMIN* Consensus-based standards for the selection of health measurement instruments, *CSMT* Conditioning specific movement tasks, *DVJ* Drop vertical jump, *FMS* Functional movement screen, *LESS* Landing error scoring system, *NMST* Netball movement screening tool, *PPM* Physical performance measures, *SEBT* Star excursion balance test

3.2 Functional movement screen

The FMS comprises seven subtests including an overhead squat, hurdle step, in-line lunge, shoulder mobility assessment, active straight-leg raise, trunk stability push-up and a prone two-point rotary stability movement all performed without external load [6, 8]. As can be seen from table 3, 16 studies were identified that reported either the intra- or inter-rater reliability of the FMS or related variations [11-14, 17-28]. Seven of the eight studies that investigated the intra-rater reliability reported ICC values ≥0.75. In addition, Teyhen et al. [24] reported an ICC of 0.74 (95% CI: 0.60-0.83). As such, it would appear that the FMS has consistently demonstrated good intra-rater reliability. Of the 16 identified studies, 14 reported inter-rater reliability either in the form of ICC, weighted kappa or Krippendorf α. Twelve of these observed an ICC ≥0.75 – representing good inter-rater reliability. In addition, Minick et al. [28] used the weighted kappa statistic to measure inter-rater reliability, however, rather than use the FMS composite score they compared raters using the individual subtests. The weighted kappa values ranged from 0.79-1.0 with the authors stating that this represented substantial to excellent agreement. In contrast, Shultz et al. [14] used the Krippendorf α statistic to quantify reliability and classified a value ≥0.80 as acceptable; however, they reported a score of 0.38 (95% CI: 0.35-0.41). These authors concluded that the FMS demonstrated poor inter-rater reliability and suggested that improved rater training may have resulted in an improved reliability score. The authors also highlighted the difference between years of experience and the number of tests a rater has administered stating that the latter is likely of greater relevance to improving reliability.

Five studies [11, 18, 20, 21, 29] included information as to the raters' experience (years of clinical practice, number of FMS tests performed or level of certification) and this allowed comparison of test reliability based on these variables. Four studies [11, 18, 20, 29] suggested that the experience of the rater did not influence the inter-rater reliability. Additionally, Smith et al. [11] observed good intra-rater reliability (ICCs >0.80) for all raters regardless of experience. In contrast, Gribble et al. [21] showed that intra-rater reliability did vary depending on the experience of raters. The raters were divided into three groups: athletic training students, athletic trainers who had not previously used the FMS and athletic trainers who had at least one year of experience administering FMS tests. The greater number of raters included in the study by Gribble et al. [21] suggests a stronger experimental foundation and this may have contributed to their contrasting findings. A clear trend, highlighting the importance of rater experience, was apparent with the students, athletic trainers, and experienced athletic trainers displaying ICC values of 0.37 (95% CI: -0.79-0.78), 0.76 (95% CI: 0.32-0.92) and 0.95 (95% CI: 0.68-0.99) respectively. However, while Gribble et al. [21] employed a greater number of raters than the other four studies the ICC values reported were based on the ratings of only three participants. Therefore, the weight of the currently available evidence suggests that the experience of the rater is not a significant factor influencing scoring.

Another aspect of movement screening that is pertinent to the reliability of such assessments is whether the rater scores participants in real-time or after the test via video recording. One study sought to address this issue [14] and found that the intra-rater reliability was superior when using recorded footage to score participants. One rater assessed individuals live while they were being filmed and retrospectively assessed the footage. They also assessed the participants again one week later and scored the tests in real-time. The ICCs for the live-live re-test and live-recorded re-test were 0.60 (95% CI: 0.35-0.77) and 0.92 (95% CI: 0.85-0.96) respectively. However, only one rater's scoring was investigated in this manner hence it is difficult to draw firm conclusions on the respective merits of live and recorded FMS scoring. Nonetheless, the outcome of this study suggests that using recorded footage to assess participants may elicit greater intra- and inter-rater reliability than doing so in real-time.

Recently, it has been shown that participant knowledge of the scoring criteria can influence FMS total score [29]. Participants were assessed prior to and after having the criteria for a perfect score explained to them. Significant improvements in scores were observed simply by providing this information. This finding demonstrates that test reliability may be affected by how the test is delivered by the assessor. If the extent of task instruction and explanation differs between assessors or test occasions it is likely that intra- and inter-rater reliability will be impacted. To ensure that any changes in score are not simply due to familiarisation it is recommended that participants have the scoring criteria clearly explained to them and are allowed practice attempts before being scored. This is not to say that individuals should be coached through the movements, rather, it is imperative they know *what* is being asked of them without being told *how* to do it.

Overall, the majority of the identified studies reveal the FMS possesses good intra- and inter-rater reliability; although it should be noted that this conclusion is not unanimous throughout the literature. It should also be noted that the majority of the identified studies were classified as demonstrating poor methodological quality (table 2). The influence of rater experience on reliability appears negligible. Furthermore, the practice of scoring tests via video footage may aid reliability yet there is limited evidence on this issue so only tentative conclusions can be drawn. Test reliability is likely influenced by the performer's knowledge of the scoring criteria; as such it is advisable to provide clear instructions to participants and allow practice attempts to reduce the influence of any learning effect. The depth of research investigating the FMS is much greater than for any other movement screen yet despite this some organizations choose to use alternative tools [2, 30]. While the specific reasons are not clear, many professional football clubs seemingly do not feel the FMS meets their screening needs. However, a number of other screens appear within the scientific literature albeit with much less supporting evidence.

Table 3. Studies that reported the intra- and/or inter-rater reliability of the FMS

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Hotta et al. (2015) [17]	10 male	Middle and long distance runners (collegiate)	2 physical therapists	n/a	n/a	2	ICC=0.98 [95% CI: 0.93-1.00]
Gulgin & Hoogenboom (2014) [18]	10 male, 10 female	University students	1 expert rater, 3 physical therapy students	n/a	n/a	4	ICC=0.88 [95% CI: 0.77-0.95]
Letafatkar et al. (2014) [19]	20	Unknown	2 physical therapists	n/a	n/a	2	ICC=0.92
Parenteau-G et al. (2014) [12]	28 male	Ice-hockey players (elite youth)	1 physiotherapist, 3 physiotherapy students	2	Rater 1 ICC=0.96 [95% CI: 0.92-0.98] Rater 2 ICC=0.96 [95% CI: 0.92-0.98]	2	ICC=0.96 [95% CI: 0.92-0.98]
Elias (2013) [20]	3 male, 2 female	Squash players (elite)	20 physiotherapists	n/a	n/a	20	ICC=0.91
Gribble et al. (2013) [21]	2 male, 1 female	University students	16 students, 15 athletic trainers, 7 experienced athletic trainers	38	All raters ICC=0.75 [95% CI: 0.53-0.87] ExpATs ICC=0.95 [95% CI: 0.68-0.99] ATs ICC=0.76 [95% CI: 0.32-0.92] Student ICC=0.37 [95% CI: -0.79-0.78]	n/a	n/a
Shultz et al. (2013) [14]	18 male, 21 female	NCAA Division 1 varsity athletes	1 student, 1 physical therapist, 2 athletic trainers, 2 S&C coaches	1	Live test-retest ICC=0.60 [95% CI: 0.35-0.77] Live-recorded test-retest ICC=0.92 [95% CI: 0.85-0.96]	6	Krippendorf α=0.38 [95% CI: 0.35-0.41]

Table 3. Continued

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Smith et al. (2013) [11]	10 male, 10 female	University students	2 students, 1 faculty member, 1 FMS certified instructor	4	Rater 1 ICC=0.90 [95% CI: 0.76-0.96] Rater 2 ICC=0.81 [95% CI: 0.57-0.92] Rater 3 ICC=0.91 [95% CI: 0.78-0.96] Rater 4 ICC=0.88 [95% CI: 0.72-0.95]	4	Occasion 1 ICC=0.89 [95% CI: 0.80-0.95] Occasion 2 ICC=0.87 [95% CI: 0.76-0.94]
Frohm et al. (2012) [22]	26 male	Soccer players (elite)	8 physiotherapists	8	Rater 1 ICC=0.87 Rater 2 ICC=0.77 Rater 3 ICC=0.83 Rater 4 ICC=0.77 Rater 5 ICC=0.79 Rater 6 ICC=0.45 Rater 7 ICC=0.79 Rater 8 ICC=0.75	8	Occasion 1 ICC=0.80 Occasion 2 ICC=0.81
Klusemann et al. (2012) [23]	10	Basketball players (elite youth)	8 (unspecified combination of S&C coaches / physiotherapists)	8	ICC=0.82	n/a	n/a
Onate et al. (2012) [13]	12 male, 7 female	University students	1 athletic trainer, 1 S&C coach	1	ICC=0.92	2	ICC=0.98
Teyhen et al. (2012) [24]	53 male, 11 female	Military personnel	8 physical therapy students	4	ICC=0.74 [95% CI: 0.60-0.83]	8	ICC=0.76 [95% CI: 0.63-0.85]
Butler et al. (2012) [25]	30	Middle school students	1 FMS creator, 1 FMS certified instructor	n/a	n/a	2	ICC=0.99

Table 3. Continued

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Schneiders et al. (2011) [26]	10	Recreationally active individuals	2 academic researchers	n/a	n/a	2	ICC=0.97
Chorba et al. (2010) [27]	3 male, 5 female	University students	2 physical therapists	n/a	n/a	2	ICC=0.98
Minick et al. (2010) [28]	17 male, 23 female	University students	2 FMS creators, 2 FMS certified instructors	n/a	n/a	4	Weighted κ values for each test ranged from 0.79-1.0 when comparing novice and experienced raters

N.B. all values refer to the FMS composite score unless otherwise stated *AT* Athletic trainer, *ExpAT* Experienced athletic trainer, *CI* Confidence interval, *FMS* Functional movement screen, *ICC* Intraclass correlation coefficient, *κ* kappa, *n/a* Not applicable, *NCAA* National Collegiate Athletic Association, *S&C* Strength and conditioning

3.3 Landing error scoring system

The LESS comprises one movement; drop vertical jumps. The accompanying scoring criteria relate to observed errors in technique and result in a potential minimum and maximum score of zero to 19 respectively with a higher score indicating poorer performance. The four studies reporting either the intra- or inter-rater reliability of the LESS are presented in table 4 [31-34]. Both Smith et al. [31] and Padua et al. [34] reported excellent intra-rater reliability (ICCs >0.90). When interpreting these findings it should be noted that, in total, only three raters' data was used to calculate these ICC values. Further research is needed to establish the robustness of such findings. All four of the identified studies measured the inter-rater reliability of the LESS with the ICC values ranging from 0.72 to 0.92 indicating good repeatability. Again, caution should be employed when analyzing these results since this conclusion is based on the data from only nine raters. One study investigated the influence of rater experience on LESS scoring and found that novice (less than one year experience as a certified athletic trainer) and expert (15 years experience as a certified athletic trainer) raters displayed moderate to perfect agreement on all items [33]. The detailed scoring criteria employed by the LESS likely explain this high level of agreement between raters. The drawback to such a thorough scoring system is inevitably the time it takes to score each participant. The original LESS protocol requires video recording of tests with subsequent scoring by assessors from the footage and this methodology has associated costs, both from a financial and time perspective. A real-time scoring system to overcome these restrictive issues was developed by Padua et al. [32]. Three raters' real-time scoring of 43 participants was compared and the resulting ICC values ranged from 0.72 (95% CI: 0.42-0.88) to 0.81 (95% CI: 0.56-0.92) suggesting moderate to good inter-rater reliability for the real-time version. Taken collectively, the initial evidence is promising with regard to the reliability of the test.

Table 4. Studies that reported the intra- and/or inter-rater reliability of the LESS

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Smith et al. (2012) [31]	10	High school/collegiate athletes	2 raters (occupation unspecified)	2	ICC=0.97	2	ICC=0.92
Padua et al. (2011) [32]	19 male, 24 female	Military personnel	3 athletic trainers	n/a	n/a	3	Rater 1 vs rater 2 ICC=0.81 [95% CI: 0.56-0.92] Rater 1 vs rater 3 ICC=0.72 [95% CI: 0.42-0.88] Rater 1 vs combined rater 2 & 3 ICC=0.79 [95% CI: 0.64-0.88]
Onate et al. (2010) [33]	19 female	NCAA Division 1 soccer players	2 athletic trainers	n/a	n/a	2	ICC=0.84
Padua et al. (2009) [34]	25 male, 25 female	Military personnel	2 raters (occupation unspecified)	1	ICC=0.91	2	ICC=0.84

N.B. all values refer to the LESS composite score *CI* Confidence interval, *ICC* Intraclass correlation coefficient, *LESS* Landing error scoring system, *n/a* Not applicable, *NCAA* National Collegiate Athletic Association

3.4 Single-leg squat screens

Five studies were identified that explored either the intra- or inter-rater reliability of movement screens containing various single-leg squat tests (table 5) [35-39]. The single-leg squat task [36], the single-leg mini squat [35, 37, 38] and another assessment comprising two "functional tasks" – the single-leg squat and lateral step-down – were investigated [39]. While these three screens differ slightly in their protocols the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during single leg squatting. Three of the five identified studies investigated the intra-rater reliability with kappa values ranging from 0.13-0.80 representing poor to substantial agreement. These results are difficult to interpret since the kappa values within each individual study varied so widely. Differences in the populations observed between studies may explain some of the variance; however, comprehensive participant information was not reported in all instances. Similarly, differences in the precise protocols and scoring criteria for each screen variation may have contributed to the inconsistent results. These values were derived from the data collected by only seven raters in total and this relatively small evidence source likely contributes to the uncertain findings. Similarly, as can be seen from table 5 the inter-rater kappa values ranged from 0.00-0.92 making conclusions difficult to draw with regard to the reliability of these screening tools.

Table 5. Studies that reported the intra- and/or inter-rater reliability of the identified single-leg squat screens

References	Participation information		Rater information	Intra-rater reliability		Inter-reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Junge et al. (2012) [35]	72	Students (children)	2 physiotherapy students	n/a	n/a	2	Weighted κ values for each scoring category ranged from 0.54-0.86
Crossley et al. (2011) [36]	15	Unknown (adult)	3 physical therapists, 1 'expert panel'	3	Rater 1 κ=0.80 Rater 2 κ=0.70 Rater 3 κ=0.60	4	κ values ranged from 0.60-0.80
Örtqvist et al. (2011) [37]	33	Students (children)	2 physiotherapists	1	κ=0.48 [95% CI: 0.16-0.79]	2	κ=0.57 [95% CI: 0.30-0.85]
Ageberg et al. (2010) [38]	8 male, 17 female	Unknown (adult)	2 physical therapists	n/a	n/a	2	κ=0.92 [95% CI: 0.75-1.08]
Chmielewski et al. (2007) [39]	7 male, 18 female	Unknown (adult)	2 physical therapists, 1 athletic trainer	3	Weighted κ values for each test and each scoring method ranged from 0.13-0.68	3	Weighted κ values for each test and each scoring method ranged from 0.00-0.55

CI Confidence interval, *κ* kappa, *n/a* Not applicable

3.5 Drop vertical jump screens

The drop vertical jump, whereby an individual drops from a raised surface to the floor and immediately jumps vertically as high as possible, is a common screening test performed in order to identify movement patterns thought to be associated with risk of injury [40-42]. However, quantification of performance on such tests is typically achieved via objective analysis of joint angles and "separation distance" between the knees and ankles [40-43]. This type of assessment was considered separate to the primarily subjective process of movement screening discussed in this review. However, three studies were identified through the literature search that described the intra- and inter-rater reliability of drop vertical jump screen variations that conformed to the definition of a movement screen previously outlined here (table 6) [44-46]. Of the three studies, one used the first order agreement coefficient (AC1) statistic to analyze reliability [45]. The AC1 values can be interpreted in the same way as described above for kappa values [47]. As with the identified single-leg squat screens the drop vertical jump screens differ in their protocols; however, the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during landing. Intra-rater reliability ranged from moderate to almost perfect; however, much greater variation existed between raters. Across the three studies AC1 and kappa values ranged from 0.32 to 0.92 representing fair to almost perfect agreement. The poorest intra- and inter-rater reliability was reported by Whatman et al. [45] suggesting that perhaps the protocols and scoring criteria adopted by the other two studies are superior [44, 46]. The results from the identified studies are mixed; therefore, further research utilizing consistent test protocols and scoring criteria are required to elucidate the reliability of these screening tools.

Table 6. Studies that reported the intra- and/or inter-rater reliability of the identified DVJ screens

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Nilstad et al. (2014) [44]	60 female	Soccer players (elite)	3 physiotherapists	n/a	n/a	3	κ values ranged from 0.52-0.92
Whatman et al. (2013) [45]	12 male, 11 female	Variety of undisclosed sports (youth)	66 physiotherapists	26	All raters AC1=0.60 [range: 0.14-0.92] Raters >14 years experience AC1=0.65 [range: 0.22-0.91] Raters <10 years experience AC1=0.56 [range: 0.20-0.83]	66	All raters AC1=0.34 [95% CI: 0.22-0.47] Raters >14 years experience AC1=0.36 [95% CI: 0.22-0.50] Raters 10-14 years experience AC1=0.37 [95% CI: 0.21-0.53] Raters 5-9 years experience AC1=0.33 [(95% CI: 0.33-0.55] Raters <5 years experience AC1=0.32 [95% CI: 0.19-0.46]
Ekegren et al. (2009) [46]	40 female	Soccer players (regional level youth)	3 physiotherapists	3	Rater 1 κ=0.80 [95% CI: 0.65-1.00] Rater 2 κ=0.85 [95% CI: 0.72-1.00] Rater 3 κ=0.75 [95% CI: 0.58-1.00]	3	Time point 1 κ=0.80 [95% CI: 0.62-0.98] Time point 2 κ=0.77 [95% CI: 0.59-0.95]

AC1 First order agreement coefficient, *CI* Confidence interval, *CL* Confidence limit, *DVJ* Drop vertical jump, *κ* kappa, *n/a* Not applicable

3.6 Tuck jump assessment

The tuck jump assessment created by Myer et al. [48] was designed to assess the movement quality associated with repeated jumping and landing and requires an individual to perform repeated tuck jumps for 10 seconds. The plyometric aspect of this exercise task is relevant since it has been reported that injury prevention interventions lacking this explosive component have demonstrated limited success in reducing knee injury [48-50]. Two studies were identified that established the intra- and inter-rater reliability of the tuck jump assessment [51, 52]. Herrington et al. [51] revealed good intra-rater reliability for two raters with kappa values ranging from 0.81-1.00. Similarly, they reported good inter-rater reliability with a kappa value of 0.88. In contrast, Dudley et al. [52] reported poor to moderate intra- and inter-rater reliability with ICC values ranging from 0.44-0.72. One possible explanation for the discrepancy between these findings – at least with regard to intra-rater reliability – may relate to the differences in sample sizes. Dudley et al. [52] viewed videos of 40 participants whereas Herrington et al. [51] only assessed 10 subjects. This may have resulted in recall bias with the raters investigated by Herrington et al. [51] potentially remembering the previous scores of the 10 participants when scoring their videos for the second time. However, since one of the creators of the tuck jump assessment was a co-author and rater within the Herrington et al. [51] article more extensive training and experience could also have contributed to the superior reliability values. The tuck jump assessment is unique amongst movement screens in that it requires the participants to perform repeated plyometric movements with the creators proposing that the increased sport-specificity of the task may aid in highlighting injury risk. However, this assessment currently only demonstrates face validity and this should be remembered when taking this assertion into consideration. The nature of this assessment means that it may be of particular interest to practitioners working in jumping and landing sports such as netball and basketball.

3.7 Athletic ability assessment

The recently developed AAA, which consists of nine subtests, is currently used within numerous high performance environments (unpublished observation) to assess athletes' movement patterns [53]. The nine subtests include a prone hold, lateral hold, overhead squat (with 10kg bar), single-leg squat off a box, walking lunge (with a 20kg bar), single-leg forward hop, lateral bound, push-up and chin-up. McKeown et al. [53] reported excellent intra- and inter-rater reliability (ICC values of 0.97 (90% CI: 0.92-0.99) and 0.96 (90% CI: 0.94-0.98) respectively). These authors also observed a strong correlation (r=0.94) between athletes' overall AAA scores when assessed live and via video recording indicating that either method is viable. The AAA maximum score is 117 and a detailed scoring system, that stipulates criteria based on body segment, is provided allowing for more precise assessment compared to some other movement screens since the continuum of possible scores is large.

3.8 Conditioning specific movement tasks

The CSMT screen was developed to aid in the assessment of young rugby union players' readiness to enter elite academies [54]. The CSMT screen comprises six subtests including an overhead squat (with a 20kg bar), Romanian deadlift (with a 20kg bar), single-leg squat, double-leg to single-leg landing, a 40 meter sprint and countermovement jump. A four-point scoring system, similar to that employed by the FMS, was used to rate the quality of the six movements. Intra-rater kappa values ranged from 0.61-1.00 indicating substantial to excellent agreement [54]. Similarly, inter-rater kappa values ranged from 0.62-1.00. The reliability values are based on the scores given by only two raters and this should be remembered when interpreting the findings. Further investigations utilizing adult populations, greater number of raters, athletes from different sports and raters of varying experience are required before definitive conclusions can be made with regard to the reliability of the CSMT screen.

3.9 Netball movement screening tool

Another sport-specific screen, the NMST, was designed to assess movement quality in patterns relevant to the sport of netball. The NMST comprises 10 subtests including a squat, lunge with a twist, a bend and pull movement, push-up, single-leg squat, vertical jump (landing on both legs), vertical jump (landing on one leg), broad jump, the SEBT and an active straight-leg raise. Reid et al. [55] reported intra- and inter-rater ICC values of 0.96 (95% CI: 0.91-0.98) and 0.84 (95% CI: 0.65-0.93) respectively. These values suggest excellent agreement within and between raters; however, the results were based on the scores given by only two examiners. The netball players assessed by Reid et al. [55] ranged between 13-17 years old and so further reliability studies conducted with adult players are needed to establish the applicability of the results to this population.

3.10 Physical performance measures

The 16-PPM is made up of 16 subtests, 10 of which are quantitative in nature e.g. measured in distance or number of repetitions completed [56]. While these 10 subtests do not meet the aforementioned definition of movement screening exercises, the 16-PPM also includes six qualitative subtests that do assess how well an athlete performs the required movement. The six qualitatively scored subtests, which are all performed without external load, include an overhead squat, downward dog, single-leg squat, shoulder mobility assessment, active straight-leg raise and Beighton hypermobility assessment. The following reliability values refer only to the six qualitative subtests. Intra-rater reliability, reported as weighted kappa values, varied between expert and novice raters according to Tarara et al. [56]. Weighted kappa values ranged from 0.32-0.81 for the expert rater representing fair to almost perfect agreement. In contrast the two novice raters' weighted kappa values ranged from -0.09-0.78 indicating poor to substantial agreement between test occasions. As such, it would appear that training is required for raters administering the 16-PPM to ensure consistent scoring. Little information was given as to the occupation or level of qualification of the expert rater so it is unclear how much training may be required to achieve an acceptable level of consistency. Inter-rater reliability varied widely with weighted kappa values ranging between 0.24-0.93 for individual subtests representing fair to almost perfect agreement. Taking all the qualitative subtests into account the 16-PPM appears to be a moderately reliable tool for assessing movement competency if administered by expert raters.

3.11 Star excursion balance test movement quality screen

The star excursion balance test (SEBT) involves the objective measurement of unilateral reach distance of the lower extremity in various directions [57]. One article was identified that applied subjective movement quality criteria to the SEBT [58]. In its original form the SEBT does not take into account *how* somebody achieves their score and reports only the objective reach distance in centimeters. Incorporating an assessment of an individual's movement quality during this test may provide additional useful information to practitioners. In the identified study, scoring criteria related to knee, pelvis and trunk position were used by three physical therapists to score 100 university students [58]. Intra-rater reliability was not assessed while inter-rater kappa values ranged from 0.18-0.60 representing slight to moderate agreement. Due to the lack of information related to within-rater variation no judgment can currently be made as to the usefulness of the movement quality version of the SEBT.

4 Injury prediction ability of movement screens

Studies that employed a prospective cohort or case-control design and investigated the association between outcome score and injury were identified for two movement screens: the FMS and LESS. Movement screening is widely used by elite sporting organizations in an attempt to detect injury risk [2]. Given this fact, it is important that the efficacy of movement screens in achieving this goal is understood. That only two of the 10 identified screens have any supporting evidence as to their association with injury risk demonstrates that much work is needed to support this practice.

4.1 Methodological quality assessment for injury prediction studies

The methodological quality of each paper that investigated the ability of a movement screen to predict injury was assessed using a previously validated checklist for retrospective and prospective studies [59]. Specifically, an amended version was used as described by McCall et al. [60] since not all of the questions included in the full checklist were relevant for cohort studies. The questions excluded were only appropriate for intervention studies. For the purposes of this review the questions included were 1, 2, 3, 5, 6, 7, 10, 11, 12, 18, 20, 21, 22, 25 as previously used [60, 61]. Following the protocol outlined by McCall et al. [60] a percentage score was awarded for each article (see Electronic Supplementary Material Appendix S2 for raw data). A 'level of evidence' was then awarded based on the procedure outlined by the Scottish Intercollegiate Guidelines Network (SIGN) [62]. Scientific levels of evidence range from one to four according to the type of study. For example, cohort and case-control studies are level two. Levels one and two can score an additional mark of '++', '+' and '-' dependent on the judged quality and risk of bias. The percentage cut-off scores to determine if a paper was either of high quality with very low risk of bias, well conducted with low risk of bias or low quality with high risk of bias were ≥75%, 50-74% and <50% respectively [60]. A graded recommendation, following the SIGN guidelines was given for each of the two movement screens that have had their injury predictive value investigated. The assignment of the graded recommendation was based on the levels of evidence of the relevant studies and the considered subjective judgment of the present authors. Graded recommendations were considered as A: strong recommendation, B: moderate recommendation, C: weak recommendation or D: insufficient evidence to make a specific recommendation [60].

4.2 Functional movement screen

Eighteen articles were included that investigated the link between FMS score and injury risk (table 7) [17, 19, 27, 63-77]. Ten [19, 27, 64, 70-74, 76, 77] of the 18 studies reported an association between the FMS composite score and injury. It should be noted that it appears one of these studies reported an incorrect odds ratio (OR) based on the data presented and the conclusions should be interpreted with caution [19]. Kiesel et al. [77] were the first to investigate the link between FMS score and injury and followed 46 American Football players over the course of a pre-season (4.5 months). All players completed the FMS at the start of pre-season and any subsequent injuries that met the defined criteria were recorded. These authors found that the greatest specificity and sensitivity were obtained when a cut-off score of 14 was used. Specificity and sensitivity are measures of the true negative and true positive rate respectively [78]. In the case of this study, the specificity value displayed the proportion of non-injured athletes with a score greater than 14 while the sensitivity value displayed the proportion of injured athletes with a score less than or equal to 14. The closer both measures are to a value of one the more robust a tool is as a predictive instrument. An OR of 11.67 for those scoring less than or equal to 14 compared to those scoring more than 14 was reported by Kiesel et al. [77] and this suggests a significant association between FMS composite score and injury risk. The specificity and sensitivity values were 0.91 and 0.54 respectively. This revealed that while the proportion of true negatives to false negatives was high the proportion of true positives to false positives was relatively even. Despite a very large OR of 11.67, only around half of the subsequent injuries were predicted by an FMS score of less than or equal to 14. Interestingly, this seminal article by Kiesel et al. [77] is often cited within the scientific literature and explains why a cut-off score of 14 is commonly used when researching the link between FMS and injury risk. Seven articles [27, 64, 70, 71, 73, 74, 76] have since replicated the finding that individuals achieving an FMS composite score of less than or equal to 14 have an increased likelihood of suffering an injury; however, the degree of the relationship varies between studies. Differences in the number of participants, length of follow-up period and sport/occupation of participants may have contributed to the inconsistencies in strength of relationship between FMS score and injury likelihood. In contrast, eight of the identified studies found no link between FMS composite score and injury risk [17, 63, 65-69, 75]. However, three of these studies [68, 69, 75] utilized very small sample sizes and this may explain the lack of association between FMS score and injury. There may simply have been too few injuries among the participants during the follow-up periods for any association to be observed. As such, the findings of these three studies [68, 69, 75] should carry minimal weight when making any judgment about the predictive value of the FMS.

Due to the inconsistency in findings the graded recommendation for the FMS is 'D'. A number of factors contribute to the ambiguity of the collective findings. Firstly, the definition of injury was not consistent among the identified articles. Indeed, this is a common issue in sports medicine at large [79]. Kiesel et al. [77] classified an injury as membership of the injured reserve group and a time-loss of three weeks – presumably meaning that only relatively serious injuries were recorded. No details of injured reserve membership criteria or details of the specific injuries suffered were provided. In contrast, O'Connor et al. [76] defined injury as any damage to the body during training that resulted in an individual seeking medical care. This broad definition could have encompassed very minor injuries. McGill et al. [75] only considered back injuries that resulted in missed game play. Such variability in the classification of injuries makes it difficult to compare the results between each study. Similarly, the length of the follow-up period varied widely between studies with the shortest reported window of observation being six weeks and the longest two years [74, 75]. In some instances the precise length of the injury-tracking period was not specified [19, 27, 67, 70]. It has been previously recommended that epidemiological sports injury studies should follow participants for at least one year as this allows sufficient time for accumulation of exposure and injury events [79, 80]. Unfortunately, most of the identified studies followed participants for less than this time period and this should be a consideration for future research.

Other relevant considerations that have been ignored by the vast majority of studies are accounting for exposure time and training load. These represent very influential confounding variables that are essential to drawing meaningful conclusions from future prospective studies. Interestingly, a number of populations were investigated by the included studies: athletes, military personnel, elite police officers and firefighters. For instance, amongst the athlete group, individuals ranged from recreationally active to elite professionals. Due to the range of occupations and performance levels of participants it is perhaps to be expected that an inconsistent relationship between FMS score and injury should be observed when all studies are viewed collectively. The injury patterns between sports and occupations differ [81-83] hence the predictive value of the FMS may not be consistent across all populations. The use of the FMS composite score has been questioned since it is not a unitary construct and may be a misleading value as a result [84, 85]. Instead, it has been proposed that using the individual sub test scores when analyzing FMS performance may be preferable. However, as is shown in table 7, of the 18 prospective studies, 10 reported an association between the composite score and injury likelihood and so it should not be disregarded entirely.

Table 7. Studies that investigated the relationship between movement screen scores and injury

Movement screen	References	Participant information: Sample size	Participant information: Occupation/sport	Participant information: Age, y [mean ± SD or range]	Association between scores and injury	Quality score (%)	Level of evidence
FMS	Bardenett et al. (2015) [63]	77 male, 90 female	Cross country, American Football, soccer, swimming, tennis and volleyball athletes (high school)	15.2	No association between composite score and injury	87	2++
	Garrison et al. (2015) [64]	160	Swimming/diving, rugby and soccer athletes (NCAA Division I)	17-22	Score ≤13 = OR of 9.52 [95% CI: 4.16-21.79]	80	2++
	Hotta et al. (2015) [17]	84 male	Middle and long distance runners (collegiate)	20.0 ± 1.1	No association between composite score and injury Runners scoring ≤3 on the deep squat and active straight leg raise components = OR of 9.7 [95% CI: 2.1-44.4]	80	2++
	McGill et al. (2015) [65]	53 male	Elite task force police officers	37.9 ± 5.0	No association between composite score and injury	80	2++
	Teyhen et al. (2015) [66]	188 male	United States army rangers	23.3 ± 3.7	No association between composite score and injury	73	2+
	Warren et al. (2015) [67]	89 male, 78 female	Basketball, cross country, American Football, golf, T&F, tennis, volleyball, soccer and swimming/diving athletes (NCAA Division I)	20.6 ± 1.6 (injured) 20.0 ± 1.4 (non-injured)	No association between composite score and injury	80	2++

Table 7. Continued

Movement screen	References	Participant information Sample size	Occupation/sport	Age, y [mean ± SD or range]	Association between scores and injury	Quality score (%)	Level of evidence
	Zalai et al. (2015) [68]	20 male	Soccer players (professional)	23.0 ± 3.0	No association between composite score and injury Players who suffered an ankle injury received a lower score for the hurdle step sub-test ($p<0.05$) Players who suffered a knee injury received a lower score for the deep squat sub-test ($p<0.05$)	73	2+
	Dossa et al. (2014) [69]	20 male	Ice-hockey players (elite youth)	16-20	No association between composite score and injury	80	2++
	Kiesel et al. (2014) [70]	238 male	American Football players (professional)	unknown	Injured vs non-injured groups' mean scores 16.9 vs 17.4 ($p<0.05$) Score ≤14 = RR of 1.87 [95% CI: 1.20-2.96] Players with at least one asymmetry had a RR of 1.80 [95% CI: 1.11-2.74]	60	2+
	Knapik et al. (2014) [71]	770 male, 275 female	Coast guard cadets	18.1 ± 0.7 (male) 17.9 ± 0.7 (female)	Males: score ≤11 = RR of 1.64 [95% CI: 1.17-2.32] Females: score ≤14 = RR of 1.93 [95% CI: 1.27-2.95]	73	2+
	Letafatkar et al. (2014) [19]	50 male, 50 female	Soccer, handball and basketball players (recreational)	22.6 ± 3.0	Score <17 = OR of 4.7	73	2+
	Shojaedin et al. (2014) [72]	50 male, 50 female	Soccer, handball and basketball players (recreational)	22.6 ± 3.0	Score ≤17 = OR of 4.7	53	2+
	Butler et al. (2013) [73]	108	Firefighters	Unknown	Score ≤14 = OR of 8.31 [95% CI: 3.2-21.6]	60	2+
	Lisman et al. (2013) [74]	874 male	Marine officer candidates	22.4 ± 2.7	Score ≤14 = OR of 2.04 [95% CI: 1.32-3.15]	60	2+

Table 7. Continued

Movement screen	References	Participant information			Association between scores and injury	Quality score (%)	Level of evidence
		Sample size	Occupation/sport	Age, y [mean ± SD or range]			
	McGill et al. (2012) [75]	14 male	Basketball players (collegiate)	20.4 ± 1.6	No association found between composite score and injury	73	2+
	O'Connor et al. (2011) [76]	874 male	Marine officer candidates	18-30	Score ≤14 = RR of 1.5 ($p<0.05$)	67	2+
	Chorba et al. (2010) [27]	38 female	Soccer, volleyball and basketball players (NCAA Division II)	19.2 ± 1.2	Score ≤14 = OR of 3.85 [95% CI: 0.98-15.13]	80	2++
	Kiesel et al. (2007) [77]	46 male	American Football players (professional)	Unknown	Injured vs non-injured groups' mean scores 14.3 vs 17.4 ($p<0.05$) Score ≤14 = OR of 11.67 [95% CI: 2.47-54.52]	53	2+
LESS	Padua et al. (2015) [86]	348 male, 481 female	Soccer players (elite youth)	13.9 ± 1.8	ACL injured vs non-injured groups' mean scores 6.2 vs 4.4 ($p<0.05$) Score ≥5 = RR of 10.7 for indirect and non-contact ACL injury	73	2+
	Smith et al. (2012) [31]	29 male, 63 female	Lacrosse, soccer, basketball, American Football, field hockey, gymnastics (high school/collegiate)	18.3 ± 2.0	No association between score and non-contact ACL injury	87	2++

ACL Anterior cruciate ligament, *CI* Confidence interval, *FMS* Functional movement screen, *LESS* Landing error scoring system, *NCAA* National Collegiate Athletic Association, *OR* Odds ratio, *RR* Risk ratio, *T&F* Track and field, *2+* Well conducted study with low risk of bias, *2++* High quality study with very low risk of bias

4.3 Landing error scoring system

Two studies investigating the link between LESS score and injury were identified through the literature search (table 7) [31, 86]. Both studies prospectively screened participants before tracking them over the course of a sporting season. Smith et al. [31] did not report any significant relationship between LESS score and subsequent injury whereas Padua et al. [86] did. Those ranked by Smith et al. [31] as 'poor' (scoring greater than 6) displayed an OR of 3.62 compared to those ranked as 'excellent' (scoring less than or equal to 4) but the 95% confidence interval crossed 1 (0.87-15.11) indicating that the groups most likely did not differ in their risk of injury. However, Smith et al. [31] only included grade III (complete tear) non-contact anterior cruciate ligament (ACL) injuries in their analysis and as such it is not clear if LESS score was associated with any other type of injury. The LESS protocol involves whole body movement and so the outcome score may potentially display an association with other injury types. The apparent lack of connection between the LESS and ACL injury reported by Smith et al. [31] is surprising since the screen assesses the degree of knee valgus and flexion during landing which are both relevant factors to both patellofemoral pain and ACL injury [87]. It was suggested that the narrow range of recorded scores (only 0-11 out of a possible 19) could have contributed to the lack of association with injury [31]. The authors also postulated that the screen may have superior predictive ability with regard to injury among less well-trained or less physically mature individuals undergoing rapid neuromuscular development. This may be due to differences in proprioceptive awareness and strength among these groups compared to more physically mature, well-trained individuals. This theory is somewhat supported by the findings of Padua et al. [86] who observed an almost 11-fold greater risk of ACL injury among individuals with scores of 5 or greater compared to those scoring less than 5. The average age of the participants followed by Padua et al. [86] was 14 years compared to 18 years for the cohort observed by Smith et al. [31]. It may be that the LESS does have some injury predictive ability but only amongst young populations and in certain sports. Further research is required among both younger and older populations before any firm conclusions can be made regarding that suggestion. Despite a theoretical link between the LESS and lower body injury, especially ACL injury, the evidence is currently ambiguous. Due to the fact that only two studies have prospectively investigated the ability of the LESS to predict injury and they reported conflicting results the graded recommendation for this movement screen is 'D'.

5 Limitations and recommendations for future research

When interpreting the results of the identified articles it is important the reader be cognizant of a number of common limitations. The majority of reliability studies were categorized as methodologically poor. While the ICC or kappa scores reported often indicated good to excellent agreement within and between raters the true value of these findings can be questioned due to the aforementioned methodological quality of the studies. In future studies investigating the reliability of movement screens rater information such as occupation, years of experience and number of tests performed should be included to allow for a more thorough interpretation of the results. In addition, larger sample sizes would help improve the methodological quality of future reliability studies. Similarly, future studies investigating the ability of movement screens to predict injury should clearly define what an 'injury' is as well as stating the length of the observation period to allow contextual appraisal of the results. In some sports like soccer and rugby established guidelines for injury reporting already exist [79, 80]. None of the studies investigating the link between movement screening score and injury reported or accounted for the exposure time of the participants. This is a crucial point that must be considered by future studies since without this information a significant confounding variable is being ignored. All else being equal, the less time a player spends training and playing then the less opportunity they have to get injured. Readers are not currently able to determine from the current research if individuals with supposed poorer movement ability are actually at increased risk of injury because of that or simply because their exposure time is greater. Another issue to consider is that if the individual responsible for recording injuries knows the movement screening scores then an element of bias may exist. Ideally, the individual recording the injury occurrence should be blinded to the outcome of the movement screen.

6 Conclusion

In conclusion, the majority of movement screens identified through the literature search lack a substantial evidence base in relation to both their reliability and ability to predict injury. However, due to its extensive research base, the FMS is the only movement screen that has consistently demonstrated good intra- and inter-rater reliability. In addition, some studies have suggested possible predictive ability with regard to injury risk for the FMS and LESS; however, this is not a unanimous finding. Based purely on the reported ICCs and kappa values all identified screens appear to have good reliability with the exception of the various single-leg squat screens and the SEBT movement quality screen. Further research is warranted to verify the initial reliability values for the identified movement screens since the evidence base is still limited and the majority of the identified reliability studies were classified as methodologically poor. None of the identified movement screens have enough supporting evidence to justify them being heralded as injury prediction tools. Overall, movement screening may be useful for practitioners to enhance their holistic knowledge of an athlete but it seems the subjectivity of scoring makes it difficult to apply these results to injury prediction with any degree of certainty.

Compliance with Ethical Standards

Funding

No funding was provided that contributed to the development of this manuscript. Robert McCunn and Hugh Fullagar are supported by 'Science and Health in Football' scholarships funded by the DAAD (German Academic Exchange Service).

Conflicts of Interest

Robert McCunn, Tim Meyer, Hugh Fullagar and Karen aus der Fünten declare that they have no conflict of interest with the content of this review. Ian McKeown declares that he is the primary developer of the athletic ability assessment (AAA) movement screen, however, does not stand to receive financial gain as a result. Additionally, the AAA is not a patented commodity nor is it protected by copyright.

References

1. Mendez-Villanueva A, Buchheit M. Football-specific fitness testing: adding value or confirming the evidence? J Sports Sci. 2013;31(13):1503-1508.

2. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. Br J Sports Med. 2014;48(18):1352-1357.

3. McGill S, Frost D, Andersen J, et al. Movement quality and links to measures of fitness in firefighters. Work. 2013;45(3):357-366.

4. Frost D, Andersen J, Lam T, et al. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics. 2013;56(4):637-649.

5. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. Sports Med. 1992;14(2):82-99.

6. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function-part 2. Int J Sports Phys Ther. 2014;9(4):549-563.

7. Giblin S, Collins D, Button C. Physical literacy: importance, assessment and future directions. Sports Med. 2014;44(9):1177-1184.

8. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function - part 1. Int J Sports Phys Ther. 2014;9(3):396-409.

9. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport. 2003;4(3):122-128.

10. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217-238.

11. Smith CA, Chimera NJ, Wright NJ, et al. Interrater and intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):982-987.

12. Parenteau-G E, Gaudreault N, Chambers S, et al. Functional movement screen test: a reliable screening test for young elite ice hockey players. Phys Ther Sport. 2014;15(3):169-175.

13. Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. J Strength Cond Res. 2012;26(2):408-415.

14. Shultz R, Anderson SC, Matheson GO, et al. Test-retest and interrater reliability of

the functional movement screen. J Athl Train. 2013;48(3):331-336.

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.

16. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651-657.

17. Hotta T, Nishiguchi S, Fukutani N, et al. Functional movement screen for predicting running injuries in 18-24 year-old competitive male runners. J Strength Cond Res. 2015. [Epub ahead of print]

18. Gulgin H, Hoogenboom B. The functional movement screening (fms)™: an inter-rater reliability study between raters of varied experience. Int J Sports Phys Ther. 2014;9(1):14-20.

19. Letafatkar A, Hadadnezhad M, Shojaedin S, et al. Relationship between functional movement screening score and history of injury. Int J Sports Phys Ther. 2014;9(1):21-27.

20. Elias JE. The Inter-rater reliability of the functional movement screen within an athletic population using untrained raters. J Strength Cond Res. 2013. [Epub ahead of print]

21. Gribble PA, Brigle J, Pietrosimone BG, et al. Intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):978-981.

22. Frohm A, Heijne A, Kowalski J, et al. A nine-test screening battery for athletes: a reliability study. Scand J Med Sci Sports. 2012;22(3):306-315.

23. Klusemann MJ, Pyne DB, Fay TS, et al. Online video-based resistance training improves the physical capacity of junior basketball athletes. J Strength Cond Res. 2012;26(10):2677-2684.

24. Teyhen DS, Shaffer SW, Lorenson CL, et al. The Functional Movement Screen: a reliability study. J Orthop Sports Phys Ther. 2012;42(6):530-540.

25. Butler RJ, Plisky PJ, Kiesel KB. Interrater reliability of videotaped performance on the functional movement screen using the 100-point scoring scale. Athl Train Sports Health Care. 2012;4(3):103-109.

26. Schneiders AG, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. Int J Sports Phys Ther. 2011;6(2):75-82.

27. Chorba RS, Chorba DJ, Bouillon LE, et al. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. N Am J Sports Phys Ther. 2010;5(2):47-54.

28. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. J Strength Cond Res. 2010;24(2):479-486.

29. 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. 2013. [Epub ahead of print]

30. McKeown I, Ball N. Current practices of long term athlete development of junior athletes in high performance sport environments. J Aust Strength Cond. 2013;21(1):4-12.

31. Smith HC, Johnson RJ, Shultz SJ, et al. A prospective evaluation of the Landing Error Scoring System (LESS) as a screening tool for anterior cruciate ligament injury risk. Am J Sports Med. 2012;40(3):521-526.

32. Padua DA, Boling MC, Distefano LJ, et al. Reliability of the landing error scoring system-real time, a clinical assessment tool of jump-landing biomechanics. J Sport Rehabil. 2011;20(2):145-156.

33. Onate J, Cortes N, Welch C, et al. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. J Sport Rehabil. 2010;19(1):41-56.

34. Padua DA, Marshall SW, Boling MC, et al. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am J Sports Med. 2009;37(10):1996-2002.

35. Junge T, Balsnes S, Runge L, et al. Single leg mini squat: an inter-tester reproducibility study of children in the age of 9-10 and 12-14 years presented by various methods of kappa calculation. BMC Musculoskelet Disord. 2012;13:203.

36. Crossley KM, Zhang WJ, Schache AG, et al. Performance on the single-leg squat task indicates hip abductor muscle function. Am J Sports Med. 2011;39(4):866-873.

37. Örtqvist M, Moström EB, Roos EM, et al. Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children. Knee Surg Sports Traumatol Arthrosc. 2011;19(12):2060-2066.

38. Ageberg E, Bennell KL, Hunt MA, et al. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskelet Disord. 2010;11:265.

39. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. J Orthop Sports Phys Ther. 2007;37(3):122-129.

40. Noyes FR, Barber-Westin SD, Fleckenstein C, et al. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. Am J Sport Med. 2005;33(2):197-207.

41. Barber-Westin SD, Galloway M, Noyes FR, et al. Assessment of lower limb neuromuscular control in prepubescent athletes. Am J Sport Med. 2005;33(12):1853-1860.

42. Dingenen B, Malfait B, Nijs S, et al. Can two-dimensional video analysis during single-leg drop vertical jumps help identify non-contact knee injury risk? A one-year prospective study. Clin Biomech. 2015. [Epub ahead of print]

43. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. J Sport Rehabil. 2012;21(1):7-11.

44. Nilstad A, Andersen TE, Kristianslund E. Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening. J Orthop Sport Phys. 2014:44(5):358-365.

45. Whatman C, Hume P, Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Phys Ther Sport. 2013;14(3):168-174.

46. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. J Orthop Sports Phys Ther. 2009;39(9):665-674.

47. Gwet KL. Handbook of inter-rater reliability; the definitive guide to measuring the extent of agreement among multiple raters. 3rd ed. Gaithersburg: Advanced Analytics, LLC; 2012.

48. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. Athl Ther Today. 2008;13(5):39-44.

49. Herman DC, Weinhold PS, Guskiewicz KM, et al. The effects of strength training on the lower extremity biomechanics of female recreational athletes during a stop-jump task. Am J Sport Med. 2008;36(4):733-740.

50. Söderman K, Werner S, Pietilä T, et al. Balance board training: prevention of traumatic injuries of the lower extremities in female soccer players? A prospective randomized intervention study. Knee Surg Sports Traumatol Arthrosc. 2000;8(6):356-363.

51. Herrington L, Myer GD, Munro A. Intra and inter-tester reliability of the tuck jump assessment. Phys Ther Sport. 2013;14(3):152-155.

52. Dudley LA, Smith CA, Olson BK, et al. Interrater and intrarater reliability of the tuck jump assessment by health professionals of varied educational backgrounds. J Sports Med (Hindawi Publ Corp). 2013;2013: Article ID 483503.

53. McKeown I, Taylor-McKeown K, Woods C, et al. Athletic ability assessment: a movement assessment protocol for athletes. Int J Sports Phys Ther. 2014;9(7):862-

873.

54. Parsonage JR, Williams RS, Rainer P, et al. Assessment of conditioning-specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. J Strength Cond Res. 2014;28(6):1497-1506.

55. Reid DA, Vanweerd RJ, Larmer PJ, et al. The inter and intra rater reliability of the Netball Movement Screening Tool. J Sci Med Sport. 2014;18(3):353-357.

56. Tarara DT, Hegedus EJ, Taylor JB. Real-time test-retest and interrater reliability of select physical performance measures in physically active college-aged students. Int J Sports Phys Ther. 2014;9(7):874-887.

57. Plisky PJ, Rauh MJ, Kaminski TW, et al. Star excusrsion balance test as a predictor of lower extremity injury in high school basketball players. J Orthop Sports Phys Ther. 2006;36(12):911-919.

58. Ness BM, Taylor AL, Haberl MD, et al. Clinical observation and analysis of movement quality during performance on the star excursion balance test. Int J Sports Phys Ther. 2015;10(2):168-177.

59. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377-384.

60. McCall A, Carling C, Davison M, et al. Injury risk factors, screening tests and preventative strategies: a systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. Br J Sports Med. 2015;49(9):583-589.

61. Freckleton G, Pizzari T. Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. Br J Sports Med. 2013;47(6):351-358.

62. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001;323(7308):334-336.

63. Bardenett SM, Micca JJ, DeNoyelles JT, et al. Functional movement screen normative values and validity in high school athletes: can the FMS be used as a predictor of injury? Int J Sports Phys Ther. 2015;10(3):303-308.

64. Garrison M, Westrick R, Johnson MR, et al. Association between the functional movement screen and injury development in collegiate athletes. Int J Sports Phys Ther. 2015;10(1):21-28.

65. McGill S, Frost D, Lam T, et al. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. Ergonomics. 2015. [Epub ahead of print]

66. Teyhen DS, Shaffer SW, Butler RJ, et al. What risk factors are associated with musculoskeletal injury in US army rangers? A prospective prognostic study. Clin

Orthop Relat Res. 2015;473(9):2948-2958.

67. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. J Sport Rehabil. 2015;24(2):163-170.

68. Zalai D, Panics G, Bobak P, et al. Quality of functional movement patterns and injury examination in elite-level male professional football players. Acta Physiol Hung. 2015;102(1):34-42.

69. Dossa K, Cashman G, Howitt S, et al. Can injury in major junior hockey players be predicted by a pre-season functional movement screen - a prospective cohort study. J Can Chiropr Assoc. 2014;58(4):421-427.

70. Kiesel KB, Butler RJ, Plisky PJ. Prediction of injury by limited and asymmetrical fundamental movement patterns in American football players. J Sport Rehabil. 2014;23(2):88-94.

71. Knapik JJ, Cosio-Lima LM, Reynolds KL, et al. Efficacy of functional movement screening for predicting injuries in coast guard cadets. J Strength Cond Res. 2014;29(5):1157-1162.

72. Shojaedin SS, Letafatkar A, Hadadnezhad M, et al. Relationship between functional movement screening score and history of injury and identifying the predictive value of the FMS for injury. Int J Inj Contr Saf Promot. 2014;21(4):355-360.

73. Butler RJ, Contreras M, Burton LC, et al. Modifiable risk factors predict injuries in firefighters during training academies. Work. 2013;46(1):11-17.

74. Lisman P, O'Connor FG, Deuster PA, et al. Functional movement screen and aerobic fitness predict injuries in military training. Med Sci Sports Exerc. 2013;45(4):636-643.

75. McGill SM, Andersen JT, Horne AD. Predicting performance and injury resilience from movement quality and fitness scores in a basketball team over 2 years. J Strength Cond Res. 2012;26(7):1731-1739.

76. O'Connor FG, Deuster PA, Davis J, et al. Functional movement screening: predicting injuries in officer candidates. Med Sci Sports Exerc. 2011;43(12):2224-2230.

77. 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;2(3):147-158.

78. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994;308(6943):1552.

79. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Br J Sports Med. 2006;40(3):193-201.

80. Fuller CW, Molloy MG, Bagate C, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. Br J Sports Med. 2007;41(5):328-331.

81. Almeida SA, Williams KM, Shaffer RA, et al. Epidemiological patterns of musculoskeletal injuries and physical training. Med Sci Sports Exerc. 1999;31(8):1176-1182.

82. Wolf BR, Ebinger AE, Lawler MP, et al. Injury patterns in Division I collegiate swimming. Am J Sports Med. 2009;37(10):2037-2042.

83. Bathgate A, Best JP, Craig G, et al. A prospective study of injuries to elite Australian rugby union players. Br J Sports Med. 2002;36(4):265-269.

84. Kazman JB, Galecki JM, Lisman P, et al. Factor structure of the functional movement screen in marine officer candidates. J Strength Cond Res. 2014;28(3):672-678.

85. Li Y, Wang X, Chen X, et al. Exploratory factor analysis of the functional movement screen in elite athletes. J Sports Sci. 2015:33(11):1166-1172.

86. Padua DA, DiStefano LJ, Beutler AI, et al. The landing error scoring system as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. J Athl Train. 2015;50(6):589-595.

87. Myer GD, Ford KR, Di Stasi SL, et al. High knee abduction moments are common risk factors for patellofemoral pain (PFP) and anterior cruciate ligament (ACL) injury in girls: is PFP itself a predictor for subsequent ACL injury? Br J Sports Med. 2015;49(2):118-122.

2. IS THE FUNCTIONAL MOVEMENT SCREEN (FMS™) ASSOCIATED WITH FOOTBALL-RELATED INJURY?

2.1 Rationale for the original investigation

The conclusions drawn from the literature review within chapter 1 of this thesis highlighted that the FMS™ is the most extensively researched movement screen of those identified. Furthermore, McCall et al.[24] reported that the majority of top-division professional clubs surveyed used the FMS™ in the applied setting. Despite the wealth of research concerning the FMS™ and its wide use in applied practice within football, very few studies have investigated the association with injury of the FMS™ among high-level, or youth, football players. The present study sought to address this issue by prospectively investigating the relationship between FMS™ scores and injury among elite youth players from an English Premier League club academy.

2.2 FMS™ score does not predict injury in English Premier League youth academy football players

The following section contains the submitted manuscript pertaining to the following publication:

Newton F, McCall A, Ryan D, Blackburne C, aus der Fünten K, Meyer T, Lewin C, McCunn R. Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players. *Science and Medicine in Football*. 2017;1(2):102-6.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of *Science and Medicine in Football*. The citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players

Florence Newton[1], Alan McCall[1,2], Desmond Ryan[1], Colin Blackburne[1], Karen aus der Fünten[3], Tim Meyer[3], Colin Lewin[1], Robert McCunn[3]

[1] Arsenal Football Club, Research & Development Department, London, UK

[2] Edinburgh Napier University, Research & Development Department, Edinburgh, UK

[3] Saarland University, Institute of Sport and Preventive Medicine, Saarbrücken, Germany

Corresponding author: Robert McCunn
Institute of Sport and Preventive Medicine
Saarland University
Campus Bldg. B8.2
66123 Saarbrücken, Germany
Phone: +49 681 302 70410
Email: bob.mccunn@me.com
Fax: +49 (0) 681 302 4296
Twitter: @RobertMcCunn

Word count: 2718

Abstract

Despite being commonly used, the interaction between Functional Movement Screen (FMS™) score and injury in any elite football population has not been studied. The aim of the present study was to investigate the relationship between FMS™ score and non-contact injury among elite youth players from a Premier League football academy. Eighty-four players were screened during the pre-season period and non-contact injuries recorded prospectively for the entirety of the 2013/14 football season. Logistic regression analysis was utilized to explore the relationships between the individual sub-tests of the FMS™ and injury. Receiver operating characteristic (ROC) curves were used to assess the predictive value of the FMS™ composite score. Logistic regression revealed no relationships between score achieved on the individual sub-tests and injury. ROC curves indicated poor predictive ability of the composite score. Players scoring below the identified cut-off values (≤14 or ≤15 depending on injury type considered) were 0.66 (95%CI: 0.40-1.10), 0.70 (95%CI: 0.32-1.57) and 1.52 (95%CI: 0.50-4.61) times as likely to suffer 'any', 'overuse' and 'severe' injuries respectively than those who scored above the identified cut-off values. There was no relationship between FMS™ score and injury. It was unable to predict any non-contact injury among English Premier League youth academy players.

Keywords: screening, soccer, risk, adolescent, elite

Introduction

Injuries in male elite youth football players have been shown to range from 2.0 to 19.4 injuries/1000h of total exposure (9.5 to 48.7 injuries/1000h of match exposure) with strains and sprains among the most common injury types mainly occurring in the upper leg, knee and ankle (Pfirrmann et al. 2016). Injuries in elite youth players are of particular concern to elite youth academy directors as time lost from training and matches has the potential to negatively affect the development of technical, tactical, physical and mental qualities of players. Indeed, a large-scale epidemiological study of elite male youth French players showed that those incurring more severe injuries were less likely to obtain a professional contract (Le Gall et al. 2009). As such, strategies aimed at reducing the risk of injury during the development period of young elite footballers should be emphasised.

While injury risk is multifactorial and complex (Bittencourt et al. 2016), one injury risk screening tool that is commonly used and deemed important by premier league football teams is the Functional Movement Screen (FMS™)(McCall et al. 2014). The purpose of this screening tool is to assess the movement quality of an individual (Cook at el. 2006a). Movement quality is not well defined but one definition put forward is the ability to "maintain correct posture and joint alignment in addition to balance while performing selected movements" (McCunn et al. 2016). The FMS™ has displayed moderate to good intra- and inter-rater reliability (Moran et al. 2016; McCunn et al. 2016). The underlying theory behind movement screening is that 'poor' movement quality may be a contributing factor to non-contact injury. To date, evidence relating to any potential relationship between non-contact injury and FMS score is conflicting (McCunn et al. 2016; McCall et al. 2015).

Despite its' widespread use within professional football clubs, only one prospective cohort study has investigated the relationship between the FMS™ and injury in top-league football players (Zalai et al. 2015). Indeed, none have been conducted with elite youth footballers. A recent systematic review that sought to outline the evidence behind the practices and perceptions of elite football clubs' injury prevention strategies concluded that insufficient evidence existed to make any recommendations in relation to the FMS™ (McCall et al. 2015). Therefore the aim of the present study was to determine whether a causative relationship existed between FMS™ score and injury among male players from an English premier league youth football academy.

Methods

Experimental Design

The present study followed a prospective cohort design. Players meeting the inclusion criteria were assessed using the FMS™ during the pre-season period. Injury surveillance was performed over the entirety of the subsequent season (2013/14) and all injury events recorded in accordance with the recommendations provided by Fuller et al. (2006).

Participants

Eighty-four male players registered with an English Premier League football club youth academy agreed to participate in the present study (age 13.0±1.3 years, height 167.0±9.4 cm, body mass 55.8±11.4 kg). Inclusion criteria required players to be registered with the club for the entirety of the observation season (2013/14), injury free at the initiation of the pre-season period (1st June) and eligible for the under-12, -13, -14, -15 or -16 squads. Participant assent and written parental consent were obtained prior to all testing procedures. The study was approved by the University College London Research Ethics Committee and conformed to the Declaration of Helsinki.

Procedures

All FMS™ testing was conducted by United Kingdom Strength and Conditioning Association accredited coaches or chartered physiotherapists. All testers had multiple years experience in conducting such assessments and undertook a re-cap of all procedures prior to testing each year. Standardised written instructions that followed the original test guidelines were provided for all raters and were delivered verbatim when instructing participants (Cook et al. 2006a; Cook et al. 2006b). Official FMS™ test kit was used. Each participant completed all 7 sub-tests sequentially in the following order: deep squat, hurdle step, in-line lunge, shoulder mobility, active straight leg raise, trunk stability push up and rotatory stability.

All injuries sustained during matches and training sessions were recorded and monitored by club physiotherapists in accordance with the recommendations provided by Fuller et al. (2006). Injury was defined as any physical complaint sustained by a player resulting from a football match or football training session that resulted in time loss. However, only non-contact injuries were included for analysis. Non-contact injuries included were also further categorised into two additional groups: overuse and severe. An overuse injury was one that was characterised by repeated microtrauma without a single, identifiable event while severe injuries were those that resulted in a time loss of more than 28 days (Fuller et al. 2006).

Statistical analyses

Data were analysed using SPSS Statistics version 22 (SPSS, Inc., Chicago, Illinois, USA) and MedCalc for Windows, version 16.4.3 (MedCalc Software, Ostend, Belgium). FMS™ composite scores were compared between injured and uninjured players (using three injury definitions: all non-contact, overuse and severe) using Mann-Whitney U tests. In addition, Cohen's *d* effect sizes (ES) were calculated and interpreted as trivial ($0 \leq ES \leq 0.2$), small ($0.2 < ES \leq 0.6$), moderate ($0.6 < ES \leq 1.2$), large ($1.2 < ES \leq 2.0$) and very large ($2.0 < ES \leq 4.0$) (Hopkins 2002; Cohen 1992). It was assumed that training and match exposure time between players in injured and non-injured groups was largely similar. Based on historical data from the academy in question we estimated that 50% of the players included in the present study would suffer a non-contact, football-related injury during the observation period. Given this estimation, a statistical power of 0.8 required a total sample size of $n = 80$ to detect a moderate effect (ES=0.65) of FMS™ composite score between injured and uninjured players (G*Power Version 3.1, Kiel, Germany). Binomial logistic regression was used to examine the relationship between injury and potential risk factors including: FMS™ composite score, each individual sub-test score, number of asymmetries displayed during the test and age group. Spearman's rank correlation coefficient test was used to detect multicollinearity between independent variables. If two variables demonstrated a strong correlation (r_s>.90) then one was selected and the other not included in any further analysis. Each risk factor was examined independently via univariable analysis and those with a p value <.10 were investigated further in a multivariable model (Engebretsen et al. 2010). Assuming the aforementioned estimation of injury incidence, a statistical power of 0.8 required a sample size of *n* = 82 to detect an odds ratio of 2 (per unit increase for each independent variable) using logistic regression (G*Power Version 3.1, Kiel, Germany). Receiver operating characteristic (ROC) curves were used to assess the predictive value of the FMS™ composite score for each injury definition and identify cut-off scores that maximized specificity and sensitivity. The identified cut-off scores were used to create 2x2 contingency tables and calculate relative risks (RR) with associated confidence intervals (CI). Additionally, positive likelihood ratios were calculated to allow contextual appraisal of injury risk after a positive test. The limit for the alpha error to be considered significant was set at p<0.05.

Results

Overall FMS™ scores for injured and uninjured players (using all definitions of injury) are presented in Table 1. Spearman's rank correlation coefficients revealed no multicollinearity existed between any independent variables included in the logistic regression analyses. The results from the univariable logistic regression analyses are presented in Table 2. None of the predictor variables met the inclusion criteria for further investigation in a multivariable model for all non-contact and severe injuries. The composite score and shoulder mobility sub-test did meet the inclusion criteria for further investigation when overuse injuries were considered. However, when included together in a multivariable analysis no statistically significant relationships were observed. ROC curves for all non-contact (area under the curve [95%CI]: 0.59 [0.47-0.72], p=0.14), overuse (area under the curve [95%CI]: 0.63 [0.50-0.77], p=0.06) and severe (area under the curve [95%CI]: 0.52 [0.34-0.70], p=0.84) injuries revealed no statistically significant results. A cut-off score of ≤15 for any non-contact and severe injuries was identified while a threshold of ≤14 maximized specificity and sensitivity when considering overuse events. The positive likelihood ratios and RR values using the identified cut-off scores for each injury definition are presented in Table 3.

Table 1. Comparison of FMS™ composite scores between injured and uninjured players for all injury definitions. Data as mean±SD.

Type of injury	Injured	Uninjured	Effect Size
All non-contact	15.8±1.8 *n*=38	15.3±2.7 *n*=46	0.22
Overuse	16.1±1.8 *n*=24	15.3±1.9 *n*=60	0.43
Severe	15.7±1.8 *n*=11	15.5±1.9 *n*=73	0.11

Table 2. Univariable logistic regression analyses for each injury definition.

Type of injury	Variable	Odds ratio	95%CI	p value
All non-contact	FMS™ composite score	1.16	0.92-1.47	0.20
	Asymmetries (number)	0.90	0.53-1.54	0.70
	Age group	1.28	0.95-1.73	0.11
	Deep squat	0.75	0.31-1.82	0.53
	Hurdle step	0.94	0.43-2.08	0.89
	In-line lunge	1.58	0.61-4.10	0.35
	Shoulder mobility	1.68	0.87-3.28	0.13
	Active straight leg raise	1.41	0.66-3.01	0.38
	Trunk stability push-up	1.34	0.69-2.61	0.38
	Rotary stability	1.13	0.32-4.02	0.85
Overuse	FMS™ composite score	1.26	0.97-1.64	0.08
	Asymmetries (number)	1.04	0.58-1.86	0.90
	Age group	1.14	0.82-1.58	0.44
	Deep squat	1.15	0.44-3.00	0.78
	Hurdle step	1.25	0.52-3.00	0.61
	In-line lunge	1.43	0.51-3.98	0.49
	Shoulder mobility	2.10	0.93-4.76	0.08
	Active straight leg raise	1.77	0.76-4.14	0.19
	Trunk stability push-up	1.38	0.65-2.92	0.40
	Rotary stability	0.68	0.16-2.96	0.61
Severe	FMS™ composite score	1.06	0.76-1.48	0.74
	Asymmetries (number)	1.25	0.60-2.61	0.55
	Age group	0.94	0.61-1.46	0.79
	Deep squat	1.20	0.34-4.32	0.78
	Hurdle step	0.42	0.13-1.40	0.16
	In-line lunge	0.72	0.16-3.14	0.66
	Shoulder mobility	1.28	0.48-3.44	0.63
	Active straight leg raise	0.68	0.22-2.06	0.49
	Trunk stability push-up	2.04	0.66-6.36	0.22
	Rotary stability	2.63	0.51-13.67	0.25

Table 3.Positive likelihood ratio and relative risk values using the identified FMS™ cut-off score for each injury definition.

	+ve LR	Chance of injury before screening (%)	Chance of injury after +ve screening (scoring ≤cut-off) (%)	RR (95%CI)
All non-contact	0.66	45	35	0.66 (0.40-1.10)
Overuse	0.71	29	22	0.70 (0.32-1.57)
Severe	1.28	13	16	1.52 (0.50-4.61)

+*ve* positive, *CI* confidence interval, *LR* likelihood ratio, *RR* relative risk

Discussion

The main findings of the present study revealed that in elite male youth football players competing in an English Premier League Academy: 1) there were no differences in FMS™ composite score observed between injured and uninjured groups regardless of the injury definition used (p>0.05 and trivial-small effect sizes) (Table 1), 2) no relationships with FMS™ score and non-contact, overuse or severe injuries existed hence the FMS™ had poor predictive ability.

FMS™ score is not related to injury in youth elite soccer players

No relationship between the FMS™ composite score and injury (all non-contact, overuse and severe) was observed. As the FMS composite score is made up from seven individual tests, some of which likely have greater relevance to football than others, (e.g. the shoulder mobility *versus* lower limb tests for outfield players) it was decided at the outset that possible relationships between injury and individual sub-tests would be investigated also. However, despite separating the FMS™ into its individual sub-sets not only were no relationships found (for any injury definitions), but no statistically significant relationships were observed between injury and any of the independent variables (including age group and FMS™ asymmetries).

What about predicting injuries?

Establishing a relationship between an attribute and injury is useful as it highlights a risk factor, which may in turn help inform the content of prevention strategies. However, predictive ability is even more appealing from a practical perspective (Bahr 2016). The most appropriate statistical measures that should be used to determine the predictive ability of a test include ROC curve analysis and likelihood ratios (Bahr 2016; Whiteley 2016; Opar et al. 2015; Pepe et al. 2004). A screening tool with excellent diagnostic accuracy would allow confident grouping of 'at-risk' players who could subsequently be targeted with specific injury prevention interventions.

In the present study ROC curve analysis revealed that the screening tool had poor predictive ability for any injury type (whether non-contact, overuse or severe). The area under the curve

(AUC) of an ROC curve provides an indication of the predictive ability of a diagnostic tool. An AUC=0.5 indicates that a diagnostic tool has no predictive value while an AUC=1 indicates a perfect test that results in no false positives or negatives (Hajian-Tilaki 2013). The ROC curves created in the current study produced AUCs between 0.52-0.63 depending on the injury definition used. These values are low and indicate that the FMS™ was likely not any better at predicting which players got injured than chance alone i.e. randomly assigning players to high/low risk groups.

Further statistical calculations providing insight to the diagnostic accuracy of screening tools include specificity and sensitivity in addition to positive and negative predictive values. However, while these values are relevant they are not as readily interpretable as a comparison of pre- and post-test odds of injury. Another relevant value for assessing the usefulness of a diagnostic tool is the positive likelihood ratio, which allows calculation of the post-test odds of injury after a positive test (an FMS™ score below the identified cut-off value) (Whiteley 2016). Likelihood ratios allow the calculation of these odds and offer practitioners clear information as to the usefulness of the screening tool in question. In the present study, positive likelihood ratios below a value of one indicated a reversal of the expected outcome and revealed a seemingly protective effect of scoring below the identified cut-off value (≤14 or ≤15 depending on injury type considered) on the FMS™ in the context of all non-contact and overuse injuries. Similarly, RR values ranged from 0.66 to 0.70 (Table 3) indicating a reduced likelihood of suffering any non-contact or overuse injury after scoring below the identified cut-off value. However, when considering this seemingly counterintuitive result it is important to note that 95%CIs for the RR values crossed one in all instances. What is clear; however, is that among the present sample of elite male youth soccer players an FMS™ composite score below the identified cut-off values was not associated with increased injury risk.

Does this mean we shouldn't use the FMS?

While the FMS™ may not be useful as a screening tool for highlighting elevated susceptibility to injury in elite male youth academy footballers it does not necessarily render the screening tool completely useless. Indeed it may provide other useful information. Its wide use among the world's top-league football clubs alludes to its appeal and perceived usefulness (McCall et al. 2014). Fuller et al. (2016) reported that young Australian rules football players were 1.5 times more likely to report pain during the FMS™ if they had suffered an injury the previous season than if they had not. The authors postulated that the FMS™ might be useful for highlighting players who have not fully recovered from previous injuries.

The National Strength and Conditioning Association's recent position statement on long-term athlete development highlights the importance of structured youth strength and conditioning programmes focusing on aspects such as fundamental movement ability (Lloyd et al. 2016). In the context of Premier League youth academies, strength and conditioning practitioners and physiotherapists are faced with the challenge of a continual turnover of players each year. The FMS™ may offer a quick, logistically viable and systematic method of quantifying movement competency and in doing so help determine readiness for introduction to formalized resistance training and progressions to more advanced exercise techniques. This may be particularly helpful in guiding the physical development plans of newly recruited players whom club support staff are not familiar with.

Limitations

While the present study represents a novel addition to the literature regarding injury risk in elite male youth footballers, there are some limitations. Firstly, exposure data was not available for the participants in the present study. This meant that no additional statistical procedures could be used (e.g. survival analysis or Cox proportional hazard modeling) which would have provided additional insight into the relationship between FMS™ score and injury (Finch & Marshall 2016; Bahr & Holme 2003). Differences in exposure time between injured and uninjured groups may have contributed to the findings. While it was assumed that exposure time between injured and uninjured groups was largely similar this could not be empirically confirmed in the present study. It may be that players who achieved better FMS™ scores generally displayed superior overall athleticism and were selected to play more frequently. It must be stressed that such a hypothesis is purely speculation; however, it is one theory that may help explain the seemingly counter-intuitive results. The greater the exposure time the greater the potential for suffering an injury. In addition, players performed strength and conditioning sessions throughout the observation season and such intervention may have mitigated the potential risk associated with scoring poorly on the FMS™. However, since all players were included in this aspect of training the protective effect should have been equally apparent in all individuals regardless of FMS™ score. Another limitation of the present study is that multiple injuries to the same player were not taken into account. This has been highlighted as an issue that needs to be addressed to advance the value of such prospective cohort studies; however, the lack of exposure data once again precluded such survival analysis (Finch & Marshall 2016). Finally, these results may only be a reflection of the present team and future work using larger samples including multiple teams is necessary.

Conclusion

The present results question the efficacy of the FMS™ for highlighting young male elite football players at increased risk of injury. The FMS™ is not recommended for this purpose. Readers should be cognizant that this conclusion relates to the FMS™ specifically and does not necessarily apply to other movement screening tools. The FMS™ may be too generic a test to highlight soccer-specific injury risk in male elite youth players. However, there may be other benefits to performing the assessment. For example, the FMS™ may help guide applied practitioners in the appropriate prescription of physical development programmes for large squads of players they are unfamiliar with. Future research should not only seek to add to this initial evidence for elite youth football players but provide further insight through incorporation of exposure data and in doing so include multiple injuries to the same player within the statistical analysis.

Disclosure statement

The authors report no conflict of interest.

References

1. Bahr R. 2016. Why screening tests to predict injury do not work – and probably never will…: A critical review. *Br J Sports Med*. 50:776-780.

2. Bahr R, Holme I. 2003. Risk factors for sports injuries – A methodological approach. *Br J Sports Med*. 37:384-392.

3. Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. 2016. Complex systems approach for sports injuries: Moving from risk factor identification to injury pattern recognition – narrative review and new concept. *Br J Sports Med* [Epub ahead of print].

4. Cohen J. 1992. A power primer. *Psychol Bull*. 112:155-159.

5. Cook G, Burton L, Hoogenboom B. 2006a. Pre-participation screening: The use of fundamental movements as an assessment of function – part 1. *N Am J Sports Phys Ther*.1:62-72.

6. Cook G, Burton L, Hoggenboom B. 2006b. Pre-participation screening: The use of functional movements as an assessment of function – part 2. *N Am J Sports Phys Ther*. 1:132-139.

7. Engebretsen AH, Myklebust G, Holme I, Engebretsen L, Bahr R. 2010. Intrinsic risk factors for hamstring injuries among male soccer players: A prospective cohort study. *Am J Sports Med*. 38:1147-1153.

8. Finch CF, Marshall SW. 2016. Let us stop throwing out the baby with the bathwater: Towards better analysis of longitudinal injury data. *Br J Sports Med*. 50:712-715.

9. Fuller CW, Ekstrand J, Junge A, Andersen TE, Bahr R, Dvorak J, Hägglund M, McCrory P, Meeuwisse WH. 2006. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Clin J Sport Med*. 16:97-106.

10. Fuller JT, Chalmers S, Debenedictis TA, Townsley S, Lynagh M, Gleeson C, Zacharia A, Thomson S, Magarey M. 2016. High prevalence of dysfunctional, asymmetrical, and painful movement in elite junior Australian football players assessed using the functional movement screen. *J Sci Med Sport*. [Epub ahead of print].

11. Hajian-Tilaki K. 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 4:627-635.

12. Hopkins, WG. 2002. A scale of magnitudes for effect statistics. *A New view of Statistics*. Retrieved from http://www.sportsci.org/resource/stats/effectmag.html. Accessed 05 September, 2016.

13. Le Gall F, Carling C, Reilly T. 2009. A comparison of injury in professional and non-professional male graduate youth soccer players. In: Reilly T, Korkusuz F, editors.

Science and Football VI: The proceedings of the sixth World Congress on Science and Football; 2007 Jan 15-20; Antalya, Turkey.

14. Lloyd RS, Cronin JB, Faigenbaum AD, Haff GG, Howard R, Kraemer WJ, Micheli LJ, Myer GD, Oliver JL. 2016. National Strength and Conditioning Association position statement on long-term athletic development. *J Strength Cond Res*. 30:1491-1509.

15. McCall A, Carling C, Davison M, Nedelec M, Le Gall F, Berthoin S, Dupont G. 2015. Injury risk factors, screening tests and preventative strategies: A systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. *Br J Sports Med*. 49:583-589.

16. McCall A, Carling C, Nedelec M, Davison M, Le Gall F, Berthoin S, Dupont G. 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med*. 48:1352-1357.

17. McCunn R, aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. 2016. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 46:763-781.

18. Moran RW, Schneiders AG, Major KM, Sullivan SJ. 2016. How reliable are functional movement screening scores? A systematic review of rater reliability. *Br J Sports Med*. 50:527-536.

19. Opar DA, Williams MD, Timmins RG, Hickey J, Duhig SJ, Shield AJ. 2015. Eccentric hamstring strength and hamstring injury risk in Australian footballers. *Med Sci Sports Exerc*. 47:857-865.

20. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 159:882-890.

21. Pfirrmann D, Herbst M, Ingelfinger P, Perikles S, Tug S. 2016. Analysis of injury incidences in male professional adult and elite youth soccer players: A systematic review. *J Athl Train*. 51:410-424.

22. Whiteley R. 2016. Screening and likelihood ratio infographic. *Br J Sports Med*. 50:837-838.

23. Zalai D, Panics G, Bobak P, Csáki I, Hamar P. 2015. Quality of functional movement patterns and injury examination in elite-level male professional football players. *Acta Physiol Hung*. 102:34-42.

3. DEVELOPMENT AND RELIABILITY OF A NEW FOOTBALL-SPECIFIC MOVEMENT SCREEN

3.1 Rationale for the original investigation

Given the popularity of movement screening yet the lack of association with injury of the FMS™ among footballers (chapter 2), the lack of a movement screen developed specifically for football was apparent. The FMS™ was designed as a tool to measure overall movement quality among both general and athletic populations; however, without a specific sport in mind.[7, 8] A screen consisting of movements targeting the most common sites of injury observed in football was deemed necessary. This would help determine if the lack of association with football injury of the FMS™ was reflective of the relationship between movement screening and football injury in general or only that particular assessment. However, in order to justify use of the new football-specific movement screen in further research or applied practice, the intra- and inter-rater reliability needed to be established. This is because test reliability is a prerequisite for test validity.[4] In addition, it is important for applied practitioners to be aware of what constitutes a true change in test performance. This allows meaningful feedback to be provided to the player and coaches after assessment. Establishing the reliability properties of a test allow such judgments to be made.[32]

3.2 The development and reliability of the Soccer Injury Movement Screen (SIMS)

The following section contains the submitted manuscript pertaining to the following publication:

McCunn R, aus der Fünten K, Govus A, Julian R, Schimpchen J, Meyer T. The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS). *Int J Sports Phys Ther*. 2017;12(1):53-66.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of the *International Journal of Sports Physical Therapy*. The numerical citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS)

Robert McCunn[1], Karen aus der Fünten[1], Andrew Govus[2], Ross Julian[1], Jan Schimpchen[1] & Tim Meyer[1]

[1] Saarland University, Institute of Sport and Preventive Medicine, Saarbrücken, Germany

[2] University of Bedfordshire, Institute of Sport Science and Physical Activity, Bedford, United Kingdom

Corresponding author: Robert McCunn
Institute of Sport and Preventive Medicine
Saarland University
Campus Bldg. B8.2
66123 Saarbrücken, Germany
Phone: +49 681 302 70410
Email: bob.mccunn@me.com
Fax: +49 (0) 681 302 4296

Word count: 4042

Abstract

Background/purpose: The growing volume of movement screening research reveals a belief among practitioners and researchers alike that movement quality may have an association with injury risk. However, existing movement screening tools have not considered the sport-specific movement and injury patterns relevant to soccer. The present study introduces the Soccer Injury Movement Screen (SIMS), which has been designed specifically for use within soccer. Furthermore, the aim of the present study was to assess the intra- and inter-rater reliability of the SIMS and determine its suitability for use in further research.

Methods: The study comprised a test-retest design. Twenty-five (11 males, 14 females) university students (age 25.5±4.0 years, height 171±9 cm, weight 64.7±12.6 kg) agreed to participate. The SIMS contains five sub-tests: the anterior reach, single-leg deadlift, in-line lunge, single-leg hop for distance and tuck jump. Three raters conducted the SIMS with each participant on three occasions separated by an average of 3.5 days (minimum 1 day, maximum 7 days). Rater 1 re-scored the filmed movements for all participants on all occasions six months later to establish the 'pure' intra-rater (intra-occasion) reliability for those movements.

Results: Intraclass correlation coefficient (ICC) values for intra- and inter-rater composite score reliability ranged from 0.66-0.72 and 0.79-0.86 respectively. Weighted kappa values representing the intra- and inter-rater reliability of the individual sub-tests ranged from 0.35-0.91 indicating fair to almost perfect agreement.

Conclusions: Establishing the reliability of the SIMS is a prerequisite for further research seeking to investigate the relationship between test score and subsequent injury. The present results indicate acceptable reliability for this purpose; however, scope exists to improve the intra-rater reliability of some of the individual sub-tests further.

Keywords: Assessment, football, kinematic, screening

Level of evidence: 2b

What is known about the subject: The use of movement screening is widespread within soccer and this is usually done in an effort to separate players into high- and low-risk groups with regard to injury likelihood. The underlying theory is that players who demonstrate 'poor' movement quality are more likely to suffer an injury than those who display 'good' movement quality. The most researched and well-known test of this type is the Functional Movement Screen™, however, it was not designed with any particular sport in mind and as such only incorporates a range of generic movements.

What this study adds to existing knowledge: This study introduces a new movement screening tool called the Soccer Injury Movement Screen (SIMS). As the name suggests, this test has been designed specifically for use in soccer and incorporates movements targeting the most common sites of soccer related injury. The fundamental premise behind the development of this new tool is that the SIMS may eventually prove to be more useful than the FMS™ at identifying soccer players with elevated risk of injury. This study establishes the intra- and inter-rater reliability of the SIMS. This is vital and reveals that the SIMS is suitable for use in future research/applied practice.

Introduction

The proliferation of movement screening research and its widespread use in professional soccer reveals a belief among practitioners and researchers alike that movement quality may have an association with injury risk.[1, 2] Movement quality is ill defined but relates to the ability of an individual to perform a given movement in a controlled manner while demonstrating good or acceptable technique. Exactly what constitutes good technique is a topic of debate. While it is arguable that no 'correct' movement pattern exists for any given exercise there are certain characteristics that may be undesirable, such as restricted range of motion and an inability to control coordinated movements. The rationale behind movement screening is that such limitations may result in acute injuries or contribute to insidious overuse complaints.[3-5]

Numerous screens exist; however, the supporting evidence with regard to both their reliability and association with injury varies widely in both volume and methodological quality.[1] The majority of such research has focused on the Functional Movement Screen (FMS™), which has demonstrated good reliability but conflicting relationships with injury likelihood.[1, 6] The FMS™ was designed as a 'general' movement assessment tool and has been used within a wide range of sports and professional domains including the military and emergency services.[7-9] In contrast, some screens such as the Landing Error Scoring System (LESS) have been designed with the intention of identifying those at an increased risk of a particular type of injury, for example, anterior cruciate ligament rupture.[10] In addition, some have been designed for use within particular sports, for example, netball and rugby union.[3, 11] Despite movement screening's popularity within professional soccer no soccer-specific tool currently exists.[2] The present study introduces the Soccer Injury Movement Screen (SIMS), which has been designed specifically for use within soccer. The movements contained within the assessment were selected to reflect the most common sites (lower extremities) and types (strains and sprains) of soccer-related injury and hence they primarily tax the mobility and stability of the ankle, knee and hip joints in addition to the strength and flexibility of the surrounding musculature.[12]

The efficacy of screening tests that seek to identify or predict which players will get injured has recently been questioned.[13] In the context of sports-related injuries the idea that a single attribute such as movement quality for example, could be predictive is unlikely.[14] As a result, the ultimate objective of the SIMS will be to investigate whether a *causative relationship* exists between movement quality and injury. Any potential relationship between movement quality and injury is unlikely to be substantial enough to justify the SIMS being considered 'predictive' but it may help inform the content of injury prevention programmes nonetheless by highlighting risk factors.[15]

There is reason to expect such a causative relationship between movement quality and injury may exist since some studies have reported poor FMS™ scores preceding subsequent injury.[8, 16] However, numerous studies utilizing the same movement screening tool have not observed any link.[17] The SIMS may eventually demonstrate a stronger association to injury risk than the FMS™ due to its more explicit scoring criteria (Appendix 2) focusing on specific aspects of each movement. Furthermore the FMS™ includes movements targeting the upper limbs, which have limited relevance for outfield soccer players whereas the SIMS concentrates on the lower limbs only.

Before any prospective cohort studies can be conducted using the SIMS its reliability must first be established. The reliability of an assessment tool is of critical importance since it is a pre-requisite for test validity.[18] Therefore, the aim of the present study was to test the intra- and inter-rater reliability of the SIMS and determine its suitability for use in further research.

Methods

Participants

Twenty-five (11 males, 14 females) university students (age 25.5±4.0 years, height 171±9 cm, weight 64.7±12.6 kg) agreed to participate in the present study. Inclusion criteria required participants to be aged between 18-40 years of age, free of injury (any physical condition that precluded them from completing the assessment) and recreationally active. Information pertaining to the study protocol and requirements were provided for each participant before written informed consent was collected. The study was approved by the local ethics committee (ref number: 270/15, Ärztekammer des Saarlandes, Saarbrücken, Germany) and conformed to the Declaration of Helsinki.

Raters

Three raters carried out the SIMS in the present study; all possessed postgraduate sport science qualifications and had previous professional experience delivering movement assessments. In addition, Rater 1 was an accredited strength and conditioning coach with both the United Kingdom Strength and Conditioning Association (UKSCA) and the National Strength and Conditioning Association (NSCA). Prior to the present study all raters conducted pilot testing using the SIMS with 10 participants. The pilot testing incorporated two 2-hour sessions where raters reviewed the test instructions (Appendix 1), the scoring criteria (Appendix 2) and familiarized themselves with the camera positioning (Figure 1). In addition, three more 2-hour sessions were conducted where raters practiced scoring video footage and discussed the interpretation of the scoring criteria. In total, rater training amounted to ~12 hours (10 classroom-based and two field-based).

Design

The present study comprised a test-retest design. Participants performed the SIMS on three occasions separated by an average of 3.5 days (minimum 1 day, maximum 7 days). The SIMS contains five sub-tests: the anterior reach (AR), single-leg deadlift (SLDL), in-line lunge (ILL), single-leg hop for distance (SLHD) and tuck jump (TJ) (Figure 2). Raters 1 and 2 scored all participants whereas Rater 3 only scored 15 of the 25 (for reasons unrelated to the study). Raters scored two of the five movements (AR and SLHD) included in the SIMS in real-time on each occasion. The remaining three movements (SLDL, ILL and TJ) were filmed from both the frontal and sagittal planes using iPhone 4S devices (Apple Inc., California, USA) and scored retrospectively. These sub-tests were scored from video footage, as opposed to in real-time; to allow raters to view the movements in slow motion and increase the likelihood of identifying errors. A minimum of one week separated the scoring of participants' filmed movements for occasions one, two and three respectively in an attempt to reduce the risk of rater bias (i.e. remembering the previous scores given). Scores for occasions one, two and three were compared within each rater to investigate 'real-world' intra-rater (inter-occasion) reliability. Scores were also compared between raters for each occasion to assess inter-rater reliability. Rater 1 re-scored the filmed movements for all participants on all occasions six months later to establish the 'pure' intra-rater (intra-occasion) reliability for those movements.

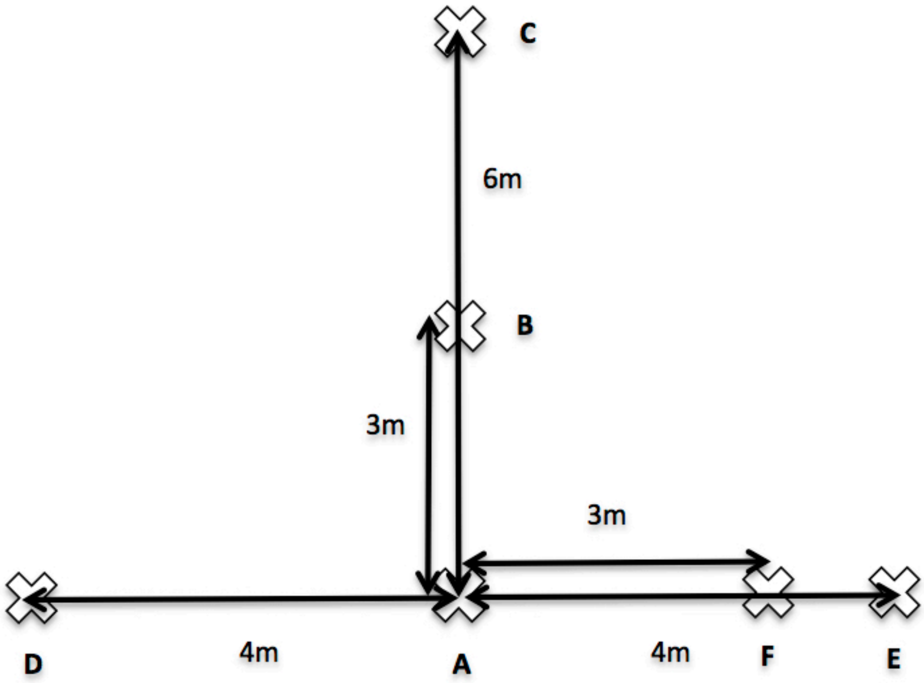


Figure 1. Schematic showing the equipment setup for the SIMS.

For all movements the participants start at A. Anterior reach: measuring tape is fixed to the floor between A and B; Single-leg deadlift: camera at B (portrait) and E (landscape) when standing on right leg, camera at B (portrait) and D (landscape) when standing on left leg; In-line lunge: camera at B (portrait) and E (landscape) when right leg forward, camera at B (portrait) and D (landscape) when left leg forward; Single-leg hop for distance: measuring tape is fixed to the floor between A and C; Tuck jump: taped cross on floor at A (60x60cm), camera at B (portrait) and F (portrait).

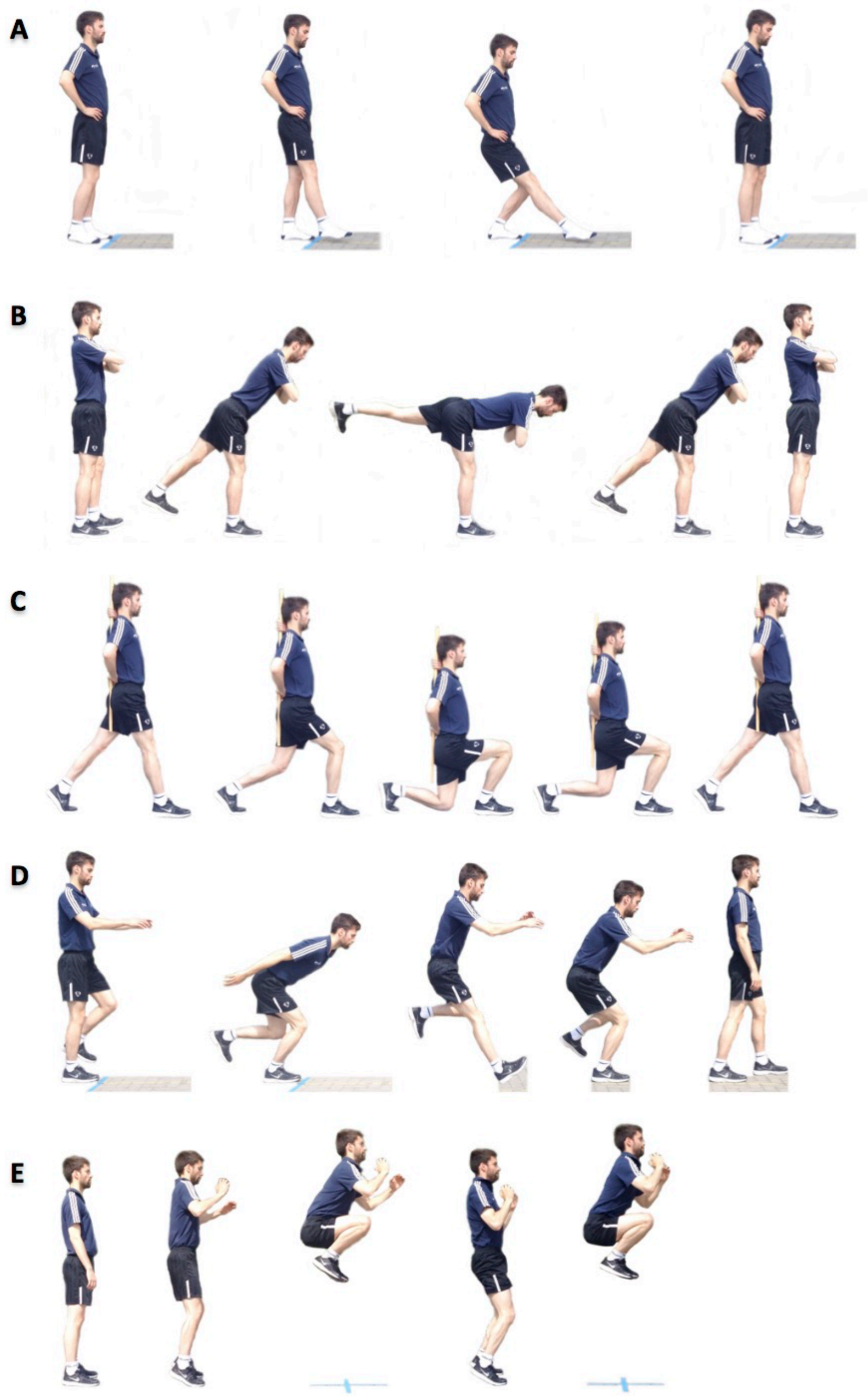


Figure 2. Demonstration cards that were shown to participants along with verbal instructions prior to test execution.

A: anterior reach; B: single-leg deadlift; C: in-line lunge; D: single-leg hop for distance; E: tuck jump.

Soccer Injury Movement Screen (SIMS)

Detailed descriptions of each movement contained within the SIMS and associated scoring criteria are outlined in Appendices 1 and 2. The ILL is the same in its setup as when performed as part of the FMS™ albeit it is scored differently while the tuck jump is performed and scored exactly as described by Myer et al.[5, 19] A standardized five minute warm up was completed before each occasion and included dynamic bodyweight exercises (e.g. squats, walking lunges, hamstring walkouts, diagonal hop and holds). The assessments were performed outdoors on a hard, rubberized sports court during summertime in dry temperate weather conditions. Participants were instructed to wear tight fitting sports clothing and the same training shoes on each occasion. The five component movements were performed in sequential order starting with the AR followed by the SLDL, ILL, SLHD and TJ. Prior to each sub-test participants were read the test instructions (Appendix 1) verbatim and shown demonstration cards (Figure 2). Participants were then allowed three practice attempts for each sub-test where any obvious miscommunication or misunderstandings relating to how to execute the movements were clarified. Time to complete the assessment was 10-15 minutes per participant.

Each component movement was scored out of 10 resulting in a theoretical maximum composite score of 50. A higher score indicated poorer performance; hence, zero was the theoretical ‘best’ score while 50 was the ‘worst’. The AR and SLHD scoring criteria were objective in nature and were based on reach and jump distance respectively. In contrast, the SLDL, ILL and TJ relied on subjective assessment of movement quality. Raters were allowed to watch the clips of the filmed movements, both in real-time speed and slow motion, as many times as they deemed necessary to make an accurate judgment when scoring.

Statistical analyses

Descriptive data are presented as means ± standard deviation. Reliability statistics are accompanied with 95% confidence intervals (CI). Data were analysed using R statistics programme (R Core Development Team 2014) and MedCalc for Windows, version 16.4.3 (MedCalc Software, Ostend, Belgium). Comparison of composite and individual sub-test scores between male and female participants was performed using the Mann-Whitney U statistic. Cohen's *d* effect size (ES) was also calculated and interpreted as follows: ≤0.2, trivial; 0.21-0.60, small; 0.61-1.2, moderate; 1.21-2.0, large; 2.1-4.0, very large.[20, 21] Two way mixed model intraclass correlation coefficients ($ICC_{3,1}$), weighted kappas (quadratic) and minimal detectable change (MDC) were used to determine the intra- and inter-rater reliability of the composite score. MDC values were calculated at both a 95% and 80% level of confidence in order to provide applied practitioners with the means to identify 'true' changes in test performance. Typically, MDC values are calculated to reflect a 95% confidence interval; however, this results in very conservative estimates of how much a test score has to change to be considered real and may be of limited usefulness in the applied setting where small improvements/decrements in test performance can be meaningful.[22] MDC values at lower levels of confidence (e.g. 80%) can be calculated and are useful to applied practitioners who may be willing to rely on more liberal estimates of test score changes. In addition, weighted kappas (quadratic) were used to determine intra- and inter-rater reliability of each individual subtest. ICC values were interpreted according to the following criteria: <0.40, poor; 0.40-0.59, fair; 0.60-0.74, good; ≥0.75, excellent.[23] Similarly, weighted kappa values were interpreted according to the guidelines outlined by Landis and Koch[24]: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect. Alpha was set at $p \leq 0.05$.

Results

Composite scores were not significantly different between males (18.3) and females (15.3) (Table 1). Only the SLDL scores differed between genders (males=4.3, females=1.8) (Table 1).

$ICC_{3,1}$, weighted kappa and MDC values for intra-rater (inter-occasion) reliability are presented in Table 2. Weighted kappa values for the individual subtests ranged from fair to substantial (0.35-0.77). With regard to the composite score, weighted kappa values were interpreted as substantial (0.63-0.68) while the ICCs were classified as good (0.66-0.72) for each rater.

$ICC_{3,1}$ and weighted kappa values for inter-rater reliability are presented in Table 3. Weighted kappa values for the individual subtests ranged from moderate to almost perfect (0.43-0.91). With regard to the composite score weighted kappa values ranged from substantial to almost perfect (0.78-0.81) while the ICCs were classified as excellent (0.79-0.86) for all three occasions.

Weighted kappa scores for ‘pure’ intra-rater (intra-occasion) reliability are presented in Table 4. The kappa values were evaluated as almost perfect for the SLDL (0.90) and ILL (0.85) while the TJ value was interpreted as substantial (0.73).

Table 1. Comparison of test scores between males and females

	Overall (n=25)	Males (n=11)	Females (n=14)	*P*	Male vs female effect size (qualitative inference)
Composite score (mean ± SD)	16.6 ± 4.9	18.3 ± 3.0	15.3 ± 5.8	0.080	0.6 (Small)
AR (mean ± SD)	1.7 ± 1.8	2.1 ± 2.3	1.4 ± 1.3	0.648	0.4 (Small)
SLDL (mean ± SD)	2.9 ± 2.1	4.3 ± 2.0	1.8 ± 1.5	<0.01	1.4 (Large)
ILL (mean ± SD)	2.6 ± 1.5	2.5 ± 1.5	2.6 ± 1.6	0.825	0.1 (Trivial)
SLHD (mean ± SD)	4.1 ± 2.3	4.2 ± 1.9	4.0 ± 2.7	0.718	0.1 (Trivial)
TJ (mean ± SD)	5.4 ± 1.3	5.2 ± 1.0	5.5 ± 1.6	0.534	0.2 (Trivial)

Test scores drawn from Rater 1 on the third testing occasion. AR anterior reach, ILL in-line lunge, SLDL single-leg deadlift, SLHD single-leg hop for distance, TJ tuck jump

Table 2. Summary of intra-rater (inter-occasion) reliability values

	Weighted kappa						$ICC_{3,1}$	MDC @ 95% confidence	MDC @ 80% confidence
	AR	SLDL	ILL	SLHD	TJ	Composite score	Composite score		
Rater 1	0.47 (0.17-0.77)	0.77 (0.67-0.87)	0.64 (0.52-0.77)	0.44 (0.26-0.61)	0.58 (0.43-0.73)	0.68 (0.54-0.81)	0.71 (0.52-0.85)	7.0	4.5
Rater 2	0.46 (0.22-0.69)	0.68 (0.55-0.81)	0.48 (0.30-0.66)	0.35 (0.15-0.55)	0.58 (0.44-0.72)	0.64 (0.49-0.80)	0.72 (0.54-0.85)	7.5	4.9
Rater 3	0.39 (0.02-0.77)	0.68 (0.55-0.81)	0.63 (0.49-0.77)	0.36 (0.11-0.61)	0.45 (0.26-0.65)	0.63 (0.45-0.80)	0.66 (0.38-0.86)	6.7	4.4

AR anterior reach, ICC intra-class correlation coefficient, ILL in-line lunge, MDC minimum detectable change, SLDL single-leg deadlift, SLHD single-leg hop for distance, TJ tuck jump

Table 3. Summary of inter-rater reliability values

	Weighted kappa						$ICC_{3,1}$
	AR	SLDL	ILL	SLHD	TJ	Composite score	Composite score
Occasion 1	0.83 (0.72-0.95)	0.51 (0.35-0.66)	0.71 (0.58-0.85)	0.84 (0.69-1.00)	0.60 (0.40-0.81)	0.78 (0.68-0.88)	0.79 (0.58-0.92)
Occasion 2	0.76 (0.62-0.90)	0.48 (0.29-0.66)	0.70 (0.56-0.84)	0.91 (0.85-0.97)	0.43 (0.18-0.68)	0.81 (0.71-0.90)	0.86 (0.70-0.95)
Occasion 3	0.59 (0.33-0.84)	0.64 (0.50-0.79)	0.58 (0.41-0.75)	0.91 (0.86-0.97)	0.50 (0.35-0.65)	0.79 (0.70-0.87)	0.79 (0.58-0.92)

AR anterior reach, ICC intra-class correlation coefficient, ILL in-line lunge, SLDL single-leg deadlift, SLHD single-leg hop for distance, TJ tuck jump

Table 4. Summary of intra-rater (intra-occasion) reliability values for filmed movements

	Weighted kappa		
	SLDL	ILL	TJ
Rater 1	0.90 (0.86-0.95)	0.85 (0.80-0.91)	0.73 (0.62-0.83)

ILL in-line lunge, SLDL single-leg deadlift, TJ tuck jump

Discussion

Overall, the present results indicate sufficient reliability for the SIMS to be considered useful for further research and applied practitioners alike. The intra-rater reliability of the SIMS composite score was classed as substantial and good for all raters based upon the weighted kappa and ICC scores respectively (Table 2). The MDC vales calculated at an 80% level of confidence demonstrate that if a one-point increase or decrease in each sub-test were observed a 'real' change in composite score would have likely occurred. The inter-rater reliability was classified as substantial to almost perfect when considering the weighted kappa values and excellent according to the ICCs (Table 3). The SLDL sub-test was the only movement where a discrepancy in scores between males and females was apparent (Table 1). Male participants regularly cited hamstring inflexibility as a limiting factor during this task whereas female participants rarely mentioned this. Females generally display superior hamstring flexibility as compared to men.[25] This difference in hamstring flexibility between males and females may potentially explain the gender difference in SLDL score observed in the present study.

The AR portion of the Y-balance test has previously been investigated as a risk factor with limb asymmetry >4 cm equating to a 2.3 – 2.7 times greater likelihood of non-contact injury among basketball and track and field athletes.[26, 27] Our scoring criteria (Appendix 2) required the rater to assign a score (0 – 10) based on the difference in reach distance between limbs. The reason for limiting our scoring range to a maximum of 10 points (a reach asymmetry of ≥10 cm) was to maintain equal weighting between all five sub-tests (each of which was scored out of 10). The scoring criteria were clearly objective for this sub-test and therefore did not directly assess movement quality. However, it was decided that the AR warranted inclusion in the SIMS regardless of not directly assessing movement quality, due to the promising evidence surrounding its relationship to injury.[26, 27] The test reflects a number of physical qualities including neuromuscular control, strength and ankle stability: all of which are likely contributors to movement quality.[1, 26, 27] Therefore, while this sub-test did not assess movement quality directly the variable that we did measure (difference in reach distance) is likely a reasonable surrogate marker. Ankle injuries occur frequently within soccer therefore the anterior reach may be a promising tool for highlighting increased risk of such events.[28] The intra-rater weighted kappa values for the AR ranged from fair to moderate (Table 2). In contrast, the inter-rater values ranged from moderate to almost perfect (Table 3).

The difference between the intra- and inter-rater weighted kappa values suggests that the scoring criteria were clear but that a large proportion of the variation in the test scores stemmed from the participants and/or the influence of time between testing occasions. As such, additional participant familiarization with the test may help improve the intra-rater reliability.

While the SLDL is multifaceted in its demands, eccentric strength and flexibility of the hamstrings are clearly primary aspects of the movement due to the flexion of the hip with an extended knee on the standing leg. Both eccentric strength and flexibility of the hamstrings have been proffered as injury risk factors within soccer players.[29, 30] Hence, the ability to perform the SLDL with a high degree of movement quality may indicate proficiency in these important attributes (hamstring flexibility and eccentric strength). The intra-rater SLDL weighted kappa values for each rater represented substantial agreement (Table 2) while the inter-rater reliability values ranged from moderate to substantial (Table 3). These findings suggest that while raters were very consistent in their scoring of the SLDL within themselves there is scope to improve the between-rater agreement. Such a scenario is somewhat inevitable when considering subjective scoring criteria; however, more detailed guidelines on what constitutes a movement 'error' may help improve consensus between raters in the future.

The ILL, or split squat, is a widely used exercise within soccer both during warm-up routines and resistance training sessions.[31, 32] According to Cook et al.[34] the ILL focuses on the "stresses simulated during rotational, decelerating and lateral type movements". All of these movement patterns are frequently observed during soccer match play.[34] The ability to perform this exercise correctly is important to ensure players do not use compensatory movements that potentially cause or exacerbate acute and overuse injuries. When performing the ILL the same test setup was used as with the FMS™; however, our scoring criteria (Appendix 2) differed.[33] The alternative scoring criteria were employed with the intention of explicitly outlining the potential movement flaws and hence enhancing clinical usefulness of the results. Both intra- and inter-rater reliability of the ILL ranged from moderate to substantial (Tables 2 & 3). The weighted kappa values reported in the present investigation are in keeping with those observed in studies of the FMS™ version of the ILL.[35-37] The more detailed scoring criteria adopted by the SIMS as compared with the FMS™ did not appear to adversely affect the reliability yet will provide practitioners with a clearer indication of where any potential movement dysfunction originates from.

It is important for soccer-specific movement assessments to incorporate explosive actions such as jumping and landing since they occur frequently during match play and often precede serious injury.[34, 38] While bilateral, vertical drop jumps have long been used for injury risk stratification[39, 40] many explosive soccer-specific actions are unilateral in nature and involve horizontal as well as vertical displacement (for example: kicking, changing direction and landing after a header).[34] The scoring criteria for the SLHD were objective and incorporated both the jump distance and the between limb difference in jump distance (Appendix 2) with each of these aspects weighted equally. The precise distances that characterized the different scoring ranges were based on pilot testing conducted with recreationally active university students and therefore may not be applicable to professional soccer players. Revised criteria may need to be established for higher-level athletes. We opted for objective, as opposed to subjective, scoring in this instance due to recent evidence suggesting jump distance as a risk factor for non-contact hamstring injury.[41] While the intra-rater weighted kappa values ranged from fair to moderate the inter-rater values indicated almost perfect agreement between raters (Tables 2 & 3). The discrepancy between the intra- and inter-rater weighted kappa values suggests that a large proportion of the variation in the test scores stemmed from the participants and/or the influence of time between testing occasions rather than the application of the scoring criteria per se.

Allowing more jump attempts may increase the likelihood of maximum jump distance being reached and a plateau in performance occurring, which may in turn help improve reliability. On 32 of the 75 SLHD tests scored by Rater 1, (25 participants on three occasions) participants recorded their best jump distance (for that occasion) on their last attempt. Similarly, 15 of the 25 participants recorded their best jump distances overall on testing occasion 3. In addition, 12 of the 25 participants scored by Rater 1 recorded their best between limb difference score on their third testing occasion. This demonstrates that incorporating a number of familiarization sessions on multiple days prior to testing may improve reliability for the same reasons highlighted previously (plateauing of performance). However, it should be remembered that the more attempts allowed and the more familiarization sessions performed the less practically feasible the assessment may become. There may be a trade-off between improved reliability and the feasibility of using the SIMS as a screening tool in the applied environment. A recent systematic review by Hegedus et al.[42] assessed the methodological quality of studies exploring the reliability and validity of commonly used field-expedient screening tests such as the SLHD. They found no studies of satisfactory methodological quality reporting the reliability of the SLHD precluding comparison of our results to previous findings.

The TJ assessment has been proposed as a field-expedient assessment of lower limb neuromuscular control.[19] It is unique as an assessment of movement quality since it requires the participant to continuously perform plyometric vertical jumps for 10 seconds.[19] While it is unlikely a player would replicate this precise activity during match-play the taxing nature of the test means it is likely to expose potentially injurious lower-limb movement patterns (particularly those associated with the onset of fatigue) that other, typically lower intensity assessments may not highlight. It has been suggested as a particularly useful tool for highlighting knee valgus movement during landing, which has been proposed as a risk factor for anterior cruciate ligament (ACL) injury.[19, 43] Considering the long-term sequelae associated with ACL injury we adjudged the TJ worthy of inclusion in the SIMS.[44, 45] Both the intra- and inter-rater weighted kappa values represented moderate agreement within and between raters (Tables 2 & 3). While this indicates acceptable reliability the weighted kappa values calculated are lower than previously reported by Myer et al.[19] However, Myer et al.[19] only assessed 10 participants and so raters may have remembered the previous scores given, leading to recall bias. In addition, they scored the same video footage twice as opposed to scoring participants on two separate occasions. The scoring criteria (Appendix 2) are inherently subjective but reliability may be improved by adding some objective guidelines to certain scoring items. For example, one of the scoring items asks: “was there a pause between jumps”? This could potentially be changed to: “was there a pause, lasting longer than one second (or another defined time period), between jumps”? Such amendments may improve consistency of scoring within and between raters. However, future research is needed to assess the difference in reliability when objective instructions are given compared with when they are not.

In an effort to separate some of the sources of variation within the test-retest design, one rater scored all the filmed movements (SLDL, ILL and TJ) from each testing occasion twice. This removed the influence of variation in test performance stemming from the participants and revealed the ‘pure’ intra-rater, or intra-occasion, reliability. The weighted kappa values for the SLDL and ILL represented almost perfect agreement while the score for the TJ indicated substantial reliability (Table 4). These higher weighted kappa values (as compared to those reported in Table 2) are not surprising since they reflect only the variation in scoring associated with the rater. These results suggest that improvements in the ‘real world’ intra-rater reliability are more likely to arise from aspects related to the participants rather than the raters. Bearing this in mind, future strategies aimed at improving the intra-rater reliability of the SIMS further may include extended participant familiarization with the test and allowing them to read the scoring criteria. Explicitly explaining the scoring criteria for the FMS™ to participants elicited improved scores.[46] This suggests that ambiguity related to what is being asked of participants during movement screening may influence their test execution and potentially contribute to variation in performance.

A number of limitations should be considered when interpreting the results of the present study. Perhaps most importantly, the pilot testing conducted to establish the scoring ranges for the SLHD (Appendix 2) were based on recreationally active university students’ scores. As such, it may be necessary to revise this aspect of the scoring criteria in the future if the SIMS is used with professional soccer players. Similarly, if the SIMS were to be utilized with youth soccer players then amendments to the scoring criteria may be necessary. In addition, the results presented here are from only 25 participants, which, is a relatively modest sample size for assessing reliability according to Terwee et al.[47]; however, we included the scores from three trials rather than the usual two in an effort to improve the credibility of our conclusions. Furthermore, our raters represented a homogenous group. All were PhD students with postgraduate degrees in sport science. Further research may be needed to assess the reliability of the SIMS when conducted by other groups of raters, for example, undergraduate students or sports coaches.

Conclusions

Until now, no movement screen has been developed specifically for use among soccer players. The SIMS composite score demonstrated good to excellent intra- and inter-rater reliability. However, the intra-rater reliability of the individual sub-tests ranged from fair to substantial indicating scope for further improvement. Establishing the reliability of the SIMS is a prerequisite for further research seeking to investigate the relationship between test score and subsequent injury. The present results indicate at least acceptable reliability for this purpose.

References

1. McCunn R, aus der Fünten K, Fullagar HHK, et al. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 2016;46(6):763-781.

2. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med*. 2014;48(18):1352-1357.

3. Reid DA, Vanweerd RJ, Larmer PJ, et al. The inter and intra rater reliability of the Netball Movement Screening Tool. *J Sci Med Sport*. 2015;18(3):353-357.

4. Padua DA, Marshall SW, Boling MC, et al. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. *Am J Sports Med*. 2009;37(10):1996-2002.

5. Cook G, Burton L, Hoogenboom B. Pre-participation screening: The use of fundamental movements as an assessment of function – part 1. *N Am J Sports Phys Ther*. 2006;1(2):62-72.

6. Moran RW, Schneiders AG, Major KM, et al. How reliable are Functional Movement Screening scores? A systematic review of rater reliability. *Br J Sports Med*. 2016;50(9):527-536.

7. McGill S, Frost D, Lam T, et al. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. *Ergonomics*. 2015;58(10):1682-1689.

8. Lisman P, O'Connor FG, Deuster PA, et al. Functional movement screen and aerobic fitness predict injuries in military training. *Med Sci Sports Exerc*. 2013;45(4):636-643.

9. 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;2(3):147-158.

10. Padua DA, DiStefano LJ, Beutler AI, et al. The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. *J Athl Train*. 2015;50(6):589-595.

11. Parsonage JR, Williams RS, Rainer P, et al. Assessment of conditioning-specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. *J Strength Cond Res*. 2014;28(6):1497-1506.

12. Ekstrand J, Hägglund, Waldén M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med*. 2011;45(7):553-558.

13. Bahr R. Why screening tests to predict injury do not work – and probably never will…: A critical review. *Br J Sports Med*. 2016;50(13):776-780.

14. Cook C. Predicting future physical injury in sports: It's a complicated dynamic system. *Br J Sports Med*. 2016 [Epub ahead of print].

15. McCunn R, Meyer T. Screening for risk factors: If you liked it then you should have put a number on it. *Br J Sports Med*. 2016 [Epub ahead of print].

16. Garrison M, Westrick R, Johnson MR, et al. Association between the functional movement screen and injury development in college athletes. *Int J Sports Phys Ther*. 2015;10(1):21-28.

17. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. *J Sport Rehabil*. 2015;24(2):163-170.

18. Batterham AM, George KP. Reliability in evidence-based clinical practice: A primer for allied health professionals. *Phys Ther Sport*. 2003;4(3):122-128.

19. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. *Athl Ther Today*. 2008;13(5):39-44.

20. A new view of statistics: Effect magnitudes. http://www.sportsci.org/resource/stats/effectmag.html Published 2002. Updated August 7, 2006. Accessed August 1, 2016.

21. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159.

22. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.

23. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessment*. 1994;6(4):284-290.

24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

25. Marshall PW, Siegler JC. Lower hamstring extensibility in men compared to women is explained by differences in stretch tolerance. *BMC Musculoskelet Disord*. 2014;15:223.

26. Smith CA, Chimera NJ, Warren M. Association of y balance test reach asymmetry and injury in division I athletes. *Med Sci Sports Exerc*. 2015;47(1):136-141.

27. Plisky PJ, Rauh MJ, Kaminski TW, et al. Star excursion balance test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther*. 2006;36(12):911-919.

28. Walls RJ, Ross KA, Fraser EJ, et al. Football injuries of the ankle: A review of injury mechanisms, diagnosis and management. *World J Orthop*. 2016;7(1):8-19.

29. Timmins RG, Bourne MN, Shield AJ, et al. Short biceps femoris fascicles and eccentric knee flexor weakness increase the risk of hamstring injury in elite football (soccer): A prospective cohort study. *Br J Sports Med*. 2015 [Epub ahead of print].

30. van Beijsterveldt AM, van de Port IG, Vereijken AJ, et al. Risk factors for hamstring injuries in male soccer players: A systematic review of prospective studies. *Scand J Med Sci Sports*. 2013;23(3):253-262.

31. Bizzini M, Dvorak J. FIFA 11+: An effective programme to prevent football injuries in various player groups worldwide – a narrative review. *Br J Sports Med*. 2015;49(9):577-579.

32. Owen AL, Wong DP, Dellal A, et al. Effect of an injury prevention program on muscle injuries in elite professional soccer. *J Strength Cond Res*. 2013;27(12):3275-3285.

33. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: The use of fundamental movements as an assessment of function – part 1. *Int J Sports Phys Ther*. 2014;9(3):396-409.

34. Stølen T, Chamari K, Castagna C, et al. Physiology of soccer: An update. *Sports Med*. 2005;35(6):501-536.

35. Teyhen DS, Shaffer SW, Lorenson CL, et al. The Functional Movement Screen: A reliability study. *J Orthop Sports Phys Ther*. 2012;42(6):530-540.

36. Schneiders AG, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther*. 2011;6(2):75-82.

37. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. *J Strength Cond Res*. 2010;24(2):479-486.

38. Waldén M, Krosshaug T, Bjørneboe J, et al. Three distinct mechanisms predominate in non-contact anterior cruciate ligament injuries in male professional football players: A systematic video analysis of 39 cases. *Br J Sports Med*. 2015;49(22):1452-1460.

39. Nilstad A, Andersen TE, Kristianslund E, et al. Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening. *J Orthop Sports Phys Ther*. 2014;44(5):358-365.

40. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J Orthop Sports Phys Ther*. 2009;39(9):665-674.

41. Goossens L, Witvrouw E, Vanden Bossche L, et al. Lower eccentric hamstring strength and single leg hop for distance predict hamstring injury in PETE students. *Eur J Sport Sci*. 2015;15(5):436-442.

42. 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;49(10):642-648.

43. Ford KR, Myer GD, Hewett TE. Valgus knee motion during landing in high school female and male basketball players. *Med Sci Sports Exerc*. 2003;35(10):1745-1750.

44. Waldén M, Hägglund M, Magnusson H, et al. ACL injuries in men's professional football: A 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after ACL rupture. *Br J Sports Med*. 2016;50(12):744-750.

45. Øiestad BE, Engebretsen L, Storheim K, et al. Knee osteoarthritis after anterior cruciate ligament injury: A systematic review. *Am J Sports Med*. 2009;37(7):1434-1443.

46. 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;29(11):3037-3044.

47. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.

4. THE ASSOCIATION WITH INJURY OF THE SOCCER INJURY MOVEMENT SCREEN (SIMS)

4.1 Rationale for the original investigation

Having observed acceptable reliability (chapter 3) for use in future research and applied practice, an important question remained: is the SIMS associated with football injury? A prospective cohort study was designed and conducted to address this question.

4.2 Soccer Injury Movement Screen (SIMS) composite score is not associated with injury among semi-professional football players

The following section contains the submitted manuscript pertaining to an original investigation conducted as part of this course of study.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of the journal it is currently under peer-review with (as of November 2017). The numerical citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

Title: Soccer Injury Movement Screen (SIMS) composite score is not associated with injury among semi-professional soccer players

Submission type: Research report

Authors:

Robert McCunn, MSc[1, 2]
Karen aus der Fünten, MD[1]
Matthew Whalan, MPthy[3, 4, 5]
John A Sampson, PhD[3]
Tim Meyer, PhD, MD[1]

Institutions:

[1]Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany
[2]Oriam: Scotland's Sports Performance Centre, Heriot-Watt University, Edinburgh, UK
[3]Centre for Human and Applied Physiology, School of Medicine, University of Wollongong, Wollongong, New South Wales, Australia
[4]Wollongong Wolves Football Club, New South Wales, Australia
[5]Player Welfare Department, Football South Coast, New South Wales, Australia

Corresponding author:

Robert McCunn
Institute of Sport and Preventive Medicine
Campus, Geb. B8.2
Saarland University
Saarbrücken
Saarland, 66123
Germany
Tel: +49 (0) 681 302 70410
Fax: +49 (0) 681 302 4296
E-mail: bob.mccunn@me.com

Running head: Relationship between SIMS and injury

Abstract word count: 250

Word count: 3320

Statement of IRB approval: The University of Wollongong's Human Research Ethics Committee (ref number: HE15/340) approved this prospective cohort study

Soccer Injury Movement Screen (SIMS) composite score is not associated with injury among semi-professional soccer players

ABSTRACT

Study Design: Cohort study.

Background: The association between movement quality and injury is equivocal. No soccer-specific movement assessment has been prospectively investigated in relation to injury risk.

Objectives: To investigate the association between a soccer-specific movement quality assessment and injury risk among semi-professional soccer players.

Methods: Semi-professional soccer players (n=306) from 12 clubs completed the Soccer Injury Movement Screen (SIMS) during the pre-season period. Individual training/match exposure and non-contact time loss injuries were recorded prospectively for the entirety of the 2016 season. Relative risks (RR) were calculated, and presented with 90% confidence intervals (CI), for the SIMS composite and individual sub-test scores from generalized linear models with Poisson distribution offset for exposure.

Results: When considering non-contact time loss lower extremity injuries (primary level of analysis), there was a *most likely trivial* association with the SIMS composite score. Similarly, SIMS composite score demonstrated *most likely* to *likely trivial* associations to all injury categories included in the secondary level of analysis (non-contact time loss hip/groin, thigh, knee and ankle injuries). When considering hamstring strains and ankle sprains specifically (tertiary level of analysis) the SIMS composite score, again, demonstrated *very likely trivial* associations. A total of 262 non-contact time loss injuries were recorded. The overall (training and match exposure combined) incidence of non-contact time loss injury was 12/1000 hours.

Conclusion: The SIMS composite score demonstrated no association to any of the investigated categories of soccer-related injury. The SIMS composite score should not be used to group players into 'high' or 'low' risk groups.

Level of evidence: 3

Key Terms: Association football, predict, epidemiology, screening

INTRODUCTION

Reducing sports injury incidence is a worthwhile endeavor for both applied practitioners and researchers alike. From a competitive point of view, lower injury burden and greater player availability have been linked to superior league ranking in professional soccer in addition to reduced financial and psychological costs.[11, 15, 25] However, it should be acknowledged that the financial costs associated with injury are not limited to professional players; for example, healthcare system and broader economic consequences due to missed days work may ensue following injury in recreational players. The 'sequence of prevention' model proferred by van Mechelen et al.[38] posits that the first and second steps to reducing injury incidence are establishing the extent of the problem (i.e. incidence) and subsequently the etiology of injury (i.e. risk factors).

Numerous risk factors have been highlighted in relation to soccer-specific injury including (but not limited to): previous injury, age, running load and eccentric knee flexor strength.[1, 24, 35] Movement quality has recently been investigated as a potential injury risk factor within soccer; however, evidence is equivocal.[4, 31, 33] While firm consensus on what constitutes movement quality is lacking; one definition offered, at least in the context of movement screening, is that it encapsulates "the maintenance of correct posture and joint alignment in addition to balance while performing the selected movements".[28] One of the underlying principles behind movement screening as a practice is that poor movement quality increases ones likelihood of injury.[28]

Many movement screens exist; however, the majority have been designed for general athletic populations and not soccer players specifically.[28] To date, no soccer-specific movement screen has been prospectively investigated in relation to injury risk; despite the use of movement screens being widespread within professional soccer.[27] The Soccer Injury Movement Screen (SIMS) is one such sport-specific tool and has been shown to be a reliable means to assess movement quality.[29] The SIMS comprises five movements; chosen to reflect the most common sites (lower extremities) and types (strains and sprains) of soccer-related injury. Hence these sub-tests primarily tax the mobility and stability of the hip, knee and ankle joints in addition to the strength and flexibility of the surrounding musculature.[29]

Therefore, the aims of the present study were two-fold: 1) to investigate the relationship between SIMS composite score and injury risk, and 2) to investigate the relationship between the individual sub-tests comprising the SIMS and injury risk.

METHODS

Participants

The University of Wollongong's Human Research Ethics Committee (ref number: HE15/340) approved this prospective cohort study. The study was conducted in accordance with the Declaration of Helsinki. In total, 306 male soccer players (mean age, 22 ± 4 years; mean height, 179 ± 7 cm; mean body mass, 75 ± 10 kg) from two National Premier Leagues New South Wales division 1 (NPL NSW) clubs and 10 Illawarra Premier League (IPL) clubs provided written informed consent to participate. If players were under the age of 18 then their legal guardian provided written informed consent and the player provided informed verbal assent. The NPL NSW represents one of eight regional divisions that collectively make up the second tier of the Australian soccer system whereas the IPL represents a smaller regional league within the geographical boundary of the NPL NSW constituency. Due to the organizational structure of Australian soccer it is unclear exactly where the IPL would rank in a comprehensive pyramid structure like those more commonly observed in European soccer. Based purely on the number of divisions between the IPL and the NPL NSW 1 league an approximate estimation would be that it represents the sixth tier. However, a caveat to viewing the structure of soccer from a national perspective within Australia is that the popularity and participation within the sport varies widely from state to state. For example, the sixth tier within New South Wales may hypothetically represent a higher standard of play compared with the second tier in the state of South Australia where soccer is less popular. Playing standard aside, all participants were semi-professional, trained two to three times per week and each club played at least one competitive game per week.

Procedures

Soccer Injury Movement Screen (SIMS)

Each participant completed the SIMS exactly as described by McCunn et al.[29] during the preseason period (March 2016). The SIMS has previously demonstrated good to excellent intra- and inter-rater reliability.[29] The SIMS is primarily a movement quality assessment comprising five sub-tests: anterior reach (AR), single-leg deadlift (SLDL), in-line lunge (ILL), single-leg hop for distance (SLHD) and the tuck jump assessment (TJ). Each sub-test is scored out of 10 points resulting in a theoretical maximum composite score of 50 when the score from each sub-test is summed. A higher score indicates poorer performance; hence, zero is the theoretical 'best' score while 50 is the 'worst'. The AR and SLHD scoring criteria are objective in nature and are based on reach and jump distance respectively. Conversely, the SLDL, ILL and TJ rely on subjective assessment of movement quality from video footage. The lead researcher was present at every testing session and acted as the test rater; scoring all video footage. Video footage was recorded using iPad 3 devices (Apple Inc, California, USA). The rater possessed undergraduate and postgraduate sport science qualifications, was an accredited strength and conditioning coach with both the United Kingdom Strength and Conditioning Association (UKSCA) and the National Strength and Conditioning Association (NSCA); and had extensive previous experience conducting/scoring the SIMS (>100 previous tests). In addition to the lead researcher, undergraduate Exercise Science students assisted in the collection of the SIMS test data. All student helpers were required to attend two training sessions (4 hours in total), covering how to set up the testing equipment and instruct participants correctly (see McCunn et al.[29]), with the lead researcher prior to assisting with any testing. All testing was conducted either in a university biomechanics laboratory or at the training ground of the respective club when suitable facilities were available. All testing was conducted on hard, non-slip surfaces. Height, weight and date of birth were also collected for each participant during testing sessions.

Injury Data Collection

Undergraduate Exercise Science students with additional training (Sports Medicine Australia Sports Trainer Level 1 award) were recruited to act as injury and exposure data collectors for the present study. In Australia, sports trainers are employed by clubs to deliver on-site first aid and acute injury management, hence, they are also well placed to record injury data.[10] In this study, the sports trainers attended every training session and match for the entirety of the 2016 season for each club. An electronic version of the injury data recording form presented by Fuller et al.[14] was used to record all physical complaints (both time loss and non-time loss). Completed electronic injury forms were sent to the lead researcher every week for review. For each recorded injury a detailed event description was also requested from the sports trainer. The descriptions included the circumstances that immediately preceded the injury event, weather/pitch conditions, the players' own explanation of how the injury occurred and any other information the sports trainer considered relevant. Each completed injury form was blinded by the lead researcher and then reviewed in conjunction with the injury description by both a chartered physiotherapist and an orthopedic doctor separately and assigned a diagnosis based on the Orchard Sports Injury Classification System version 10.1.[32] If the diagnoses provided by the physiotherapist and the orthopedic doctor differed then the lead researcher flagged the injury and all parties reconsidered the case together until consensus on the most likely diagnosis was achieved. This method of retrospective injury diagnosis has recently been advocated for and also used in previous research.[16, 30] Only non-contact injuries were included within the analyses since contact injuries are dependent on interaction with other individuals and were adjudged by the authors not inherently related to movement quality. Sports trainers also recorded training and match exposure time (in minutes) for each individual participant and included this data in the weekly submission to the lead researcher.

Statistical Analysis

All estimations were made using SPSS Statistics version 24 (SPSS, Inc., Chicago, Illinois, USA). Data are presented as means ± standard deviations and absolute or relative frequencies. The effects of the SIMS composite and individual sub-test scores on injury risk were analyzed using a generalized linear model (GLM) with a Poisson distribution, log-linear link function and offset for minutes of combined training and match exposure. Relative risks (RR) and 90% confidence intervals (90%CI) were calculated to express the effect on injury risk per one-point increase in SIMS composite or individual sub-test scores. Several injury categories were analyzed using the GLM. These injury categories were incorporated into three levels of analyses (primary, secondary and tertiary). The primary level included one category: all non-contact time loss lower extremity injuries. The secondary level included four separate injury categories: all non-contact time loss hip/groin, thigh, knee and ankle injuries, which were selected since they represent the most frequently injured body locations within soccer.[12] The tertiary level included two categories: all non-contact time loss hamstring muscle strains and ankle sprains, which were selected since these are two very commonly investigated specific injury types within soccer.[13, 39] In addition, the observed frequencies of both these injury types exceeded 20 cases. According to Bahr & Holme[3] 20—50 injury cases are required to detect moderate to strong associations between risk factors and injury likelihood. Bonferroni correction was applied to the P-values for all secondary and tertiary level injury categories to counteract the issue of multiple comparisons. Injury rates are reported as the number of injuries per 1000 hours of training, match and combined (both training and match) exposure.

Inferences regarding the effects of SIMS composite and individual sub-test scores were assessed against a pre-defined smallest worthwhile effect on injury risk, using a spreadsheet for deriving a confidence interval and clinical inference from a P-value.[18] The smallest worthwhile beneficial effect was given by an RR of 0.90 (i.e., a 10% lower injury rate), and conversely the smallest worthwhile harmful effect was given as an RR of 1.11 (i.e., an 11% higher injury rate) as previously established.[19] Effects were classified as clear if the percentage likelihood that the true effect was beneficial (i.e., reduced injury risk: RR≤0.90) was greater than 25% and the odds ratio between benefit and harm was greater than 66, otherwise the effect was deemed unclear. Effects (risk changes) were qualified against pre-defined probabilistic terms from the following scale: <0.5%, most unlikely; 0.5-5%, very unlikely; 5-25%, unlikely; 25-75%, possibly; 75-95%, likely; 95-99.5%, very likely and >99.5%, most likely.[5]

RESULTS

When considering all non-contact time loss lower extremity injuries (primary level of analysis), there was a *most likely trivial* association with the SIMS composite score (Table 1). Similarly, SIMS composite score demonstrated *most likely* to *likely trivial* associations to all injury categories included in the secondary level of analysis (time loss, non-contact hip/groin, thigh, knee and ankle injuries)(Table 1). When considering hamstring strains and ankle sprains specifically (tertiary level of analysis) the SIMS composite score, again, demonstrated *very likely trivial* associations (Table 1).

The majority of SIMS individual sub-test scores demonstrated *trivial* to *unclear* associations with hamstring strain and ankle sprain injuries (Table 2). However, a greater (worse) SLHD score *possibly increased* the risk of suffering an ankle sprain. In contrast, a greater (worse) SLDL score *possibly decreased* the risk of suffering a hamstring strain.

The frequencies and relative distributions of non-contact time loss injuries categorized by location and severity are displayed in Table 3. A total of 262 non-contact time loss injuries were recorded. The average exposure time experienced during training and match play per player was 55±26 and 18±11 hours respectively. The overall (training and match exposure combined) incidence of non-contact time loss injury was 12/1000 hours. The incidence of non-contact time loss injuries sustained during training and matches was 6/1000 hours and 28/1000 hours respectively. Injuries originating from trauma versus overuse equated to 48% (n=125) and 52% (n=137) respectively.

TABLE 1. Association between SIMS composite score and injury risk

	Relative risk	Lower 90%CI	Upper 90%CI	P-value	Bonferroni corrected P-value	Qualitative inference
Primary analysis						
Lower extremity injuries (n=244)	0.98	0.96	1.00	0.07	n/a	Most likely trivial
Secondary analysis						
Hip/groin injuries (n=48)	1.01	0.96	1.07	0.76	3.04	Most likely trivial
Thigh injuries (n=81)	0.95	0.91	0.99	0.03	0.14	Very likely trivial
Knee injuries (n=41)	0.94	0.89	0.99	0.07	0.26	Likely trivial
Ankle injuries (n=48)	1.02	0.97	1.07	0.49	1.96	Most likely trivial
Tertiary analysis						
Hamstring muscle strains (n=64)	0.94	0.90	0.98	0.01	0.02*	Very likely trivial
Ankle sprains (n=41)	1.04	0.99	1.09	0.21	0.42	Very likely trivial

* Bonferroni corrected P-value <0.05. CI; confidence interval, n/a; not applicable.

TABLE 2. Association between SIMS sub-test scores and hamstring muscle strain/ankle sprain injury risk

	Relative risk	Lower 90%CI	Upper 90%CI	P-value	Qualitative inference
Hamstring muscle strains (n=64)					
Anterior reach	0.91	0.81	1.02	0.16	Possibly trivial
Single-leg deadlift	0.90	0.80	1.02	0.15	Possibly ↓
In-line lunge	0.93	0.78	1.11	0.49	Possibly trivial
Single-leg hop for distance	0.96	0.88	1.05	0.43	Likely trivial
Tuck jump	0.97	0.85	1.11	0.71	Likely trivial
Ankle sprains (n=41)					
Anterior reach	1.06	0.94	1.20	0.43	Possibly trivial
Single-leg deadlift	1.10	0.95	1.28	0.29	Possibly trivial
In-line lunge	0.90	0.73	1.11	0.41	Unclear
Single-leg hop for distance	1.11	1.00	1.23	0.10	Possibly ↑
Tuck jump	0.97	0.83	1.14	0.75	Unclear

CI; confidence interval

TABLE 3. Non-contact time loss injury pattern by location and severity of injuries

	Total	1-3 Days (minimal)	4-7 Days (mild)	8-28 Days (moderate)	>28 days (severe)
Injury location					
Head/face	0	0	0	0	0
Neck/cervical spine	3 (1)	1	0	2 (2)	0
Shoulder/clavicula	0	0	0	0	0
Sternum/ribs/upper back	1	0	1 (2)	0	0
Abdomen	2	0	0	1 (1)	1 (5)
Low back/sacrum/pelvis	11 (4)	7 (6)	2 (4)	2 (2)	0
Upper arm	0	0	0	0	0
Elbow	1	0	0	0	1 (5)
Forearm	0	0	0	0	0
Wrist	0	0	0	0	0
Hand/finger/thumb	0	0	0	0	0
Hip/groin	48 (18)	21 (19)	11 (22)	15 (18)	1 (5)
Thigh	81 (31)	30 (27)	12 (25)	32 (39)	7 (33)
Knee	41 (16)	17 (16)	10 (20)	11 (13)	3 (14)
Lower leg/Achilles tendon	23 (9)	13 (12)	5 (10)	4 (5)	1 (5)
Ankle	48 (18)	19 (17)	8 (16)	14 (17)	7 (33)
Foot/toe	3 (1)	2 (2)	0	1 (1)	0
Total injuries	262	110	49	82	21

Values within brackets show percentage of total (values below 1% not shown).

DISCUSSION

The SIMS composite score was not meaningfully related to any of the injury categories investigated (Table 1). Similarly, the individual sub-test scores were not associated with injury, with the exceptions of the SLDL and the SLHD in relation to hamstring strains and ankle sprains respectively (Table 2). While a greater (worse) SLHD score was *possibly* associated with higher risk of ankle sprain injury it should be noted the observed association between SLDL score and hamstring strains was counter-intuitive, with a theoretically better score equating to increased risk of injury.

SIMS Composite Score

The present study suggests the SIMS does not display an association (or any predictive relationship) with injury risk. When discussing risk factors an important distinction should be made between the terms 'association' and 'prediction'.[26] Bahr[2] demonstrated the difference and explained that while an association can exist between risk factors and injury likelihood this does not necessarily equate to predictive ability. Outcome statistics related to prediction include (although are not limited to) Area Under the Curve (AUC), sensitivity, specificity and positive/negative predictive value; however, no clear guidelines exist to determine at what point these values distinguish a test as 'predictive'.[26] To date, no injury screening test has demonstrated satisfactory predictive ability, yet several have shown an association.[2]

The association with injury for the SIMS composite score was *trivial* for all categories investigated (Table 1). Despite a Bonferroni corrected P-value of <0.05 being observed with regard to hamstring strains the clinical inference was nonetheless *trivial*, indicating that no meaningful relationship existed between the SIMS composite score and injury likelihood.[21] This lack of association is consistent with previous research that has explored the relationship between movement quality and injury risk. Krosshaug et al.[22] reported that the commonly used vertical drop jump test was unable to predict anterior cruciate ligament injury in a large cohort of soccer and handball players. Similarly, the Functional Movement Screen (FMS) developed by Cook et al.[8, 9] is widely used within soccer yet its association with injury in this population is limited.[4, 27, 31] The potential contributors to sports injury are numerous and while intuitively appealing it seems movement quality is not strongly associated with injury risk. While movement quality may potentially contribute to injury likelihood in combination with other risk factors, individually it does not appear to be a significant risk factor. The etiology of injury are multifactorial and investigating individual risk factors in isolation, while scientifically sound, may not adequately address the real-world issues of injury prediction and prevention.[6, 7] However, the lack of association with injury risk does not necessarily render movement screening useless.[37] Other benefits of continuing the practice include establishing return to play test values, highlighting current musculoskeletal conditions and establishing trust/rapport between the practitioner and the athlete.[37] Furthermore, movement screening offers a systematic way for applied practitioners to identify fundamental movement patterns relevant to safe strength training and potentially performance enhancement. Some evidence suggests that movement quality may be related to physical attributes such as sprinting and jumping; ergo, the application of movement screening may relate more so to performance enhancement than injury prediction.[23, 40]

Individual Sub-test Scores

The associations with injury for the individual sub-tests mirrored the results for the composite score for the most part, with *trivial* and *unclear* relationships observed (Table 2). Two exceptions were the SLHD and SLDL when considering ankle sprains and hamstring strains respectively. A higher (worse) SLHD was *possibly* associated with a greater risk of suffering an ankle sprain. This potential relationship between the SLHD and ankle sprain risk makes intuitive sense since there is moderate evidence linking ankle instability and poor performance on this test.[17] However, a higher (worse) SLDL score was *possibly* associated with a reduced risk of suffering a hamstring strain. The observed relationship between SLDL score and hamstring strain injury is counterintuitive. It is unclear why better performance on this test should potentially result in greater risk. Although not quantified directly by the SLDL test; flexibility, eccentric strength and neuromuscular control all contribute to successful test performance. These attributes are generally believed to contribute to lower risk of injury; hence, the observed association is surprising.[34-36]

Methodological Considerations

A number of limitations should be considered when interpreting the results of the present study. Collecting injury data in a non-professional environment is fraught with challenges. The injury data collection method may have influenced the observed injury incidence.

McCunn et al.[30] highlighted various challenges associated with applying the recommendations presented in the current consensus statement on soccer injury data collection within non-professional soccer.[14] Using time loss to define injury severity has significant limitations when applied within an environment where players are not required to report for training/matches on a daily basis (such as in the present study). Consider the following example: a player suffers a suspected hamstring strain during a match (Saturday) and is removed from play; however, the next scheduled training session is not until the following Wednesday (four days later). Based on the time loss definition of injury, if the player returned to training on Wednesday and participated fully, a time loss injury of three days should be recorded. However, we cannot be sure whether the player could have participated fully if the next scheduled training session had been the day immediately following the match (which would have equated to a time loss of zero days). Such scenarios could potentially have inflated the number of minimal (one to three days time loss) injuries recorded. In addition, the reality of conducting injury research within non-professional soccer dictated that access to advanced medical technology was not always possible. As a result, when deciding upon the most appropriate injury diagnosis, objective indicators such as X-ray and magnetic resonance imaging scans were not always available. In addition, the results of the present study are only generalizable to semi-professional male players and further research may seek to investigate full professional, female or youth populations.

A number of methodological strengths should also be acknowledged. The number of injuries observed in the present study allowed for multiple categories to be investigated while still satisfying the suggestion by Bahr & Holme[3] that a minimum of 20-50 cases be included for meaningful analysis. In addition, the individuals responsible for collecting the injury data and determining the diagnoses were blinded to the SIMS score of the participants, reducing the likelihood of bias. The statistical approach utilized accounted for multiple injuries to the same player and the exposure time of each individual. This is rare within research that has investigated the association with injury of other movement screening tests. Furthermore, the use of magnitude-based inferences provided an estimation of the strength of relationship between SIMS score and injury risk, rather than simply relying on null hypothesis significance (P-values) testing.[20]

CONCLUSION

The SIMS composite score was not associated with any of the injury categories investigated. Similarly, the individual sub-test scores were not associated with injury, with the exceptions of the SLDL and the SLHD in relation to hamstring strains and ankle sprains respectively. It should be noted the observed association between SLDL score and hamstring strains was counter-intuitive, with a theoretically better score equating to increased risk of injury. Therefore, the SIMS should not be used to categorize players as 'high' or 'low' risk. However, the SIMS may be useful in other ways. It may help practitioners identify physical qualities – for example, limb asymmetries related to strength and or flexibility – that warrant development from a performance enhancement perspective.

FINDINGS: The SIMS composite score was not associated with any of the injury categories investigated. Similarly, the individual sub-test scores were not associated with injury, with the exceptions of the SLDL and the SLHD in relation to hamstring strains and ankle sprains respectively.

IMPLICATIONS: The SIMS should not be used to categorize players as 'high' or 'low' risk.

CAUTION: Using time loss to define injury severity has significant limitations when applied within an environment where players are not required to report for training/matches on a daily basis (such as in the present study).

REFERENCES

1. Arnason A, Sigurdsson SB, Gudmundsson A, Holme I, Engebretsen L, Bahr R. Risk factors for injuries in football. *Am J Sports Med*. 2004;32(suppl 1):5S-16S. https://doi.org/10.1177/0363546503258912

2. Bahr R. Why screening tests to predict injury do not work – and probably never will…: a critical review. *Br J Sports Med*. 2016;50(13):776-780. http://dx.doi.org/10.1136/bjsports-2016-096256

3. Bahr R, Holme I. Risk factors for sports injuries—a methodological approach. *Br J Sports Med*. 2003;37(5):384-392. http://dx.doi.org/10.1136/bjsm.37.5.384

4. Bakken A, Targett S, Bere T et al. The functional movement test 9+ is a poor screening test for lower extremity injuries in professional male football players: a 2-year prospective cohort study [published online May 16 2017]. *Br J Sports Med*. 2017. http://dx.doi.org/10.1136/bjsports-2016-097307

5. Batterham A, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1(1):50-57. https://doi.org/10.1123/ijspp.1.1.50

6. Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition – narrative review and new concept. *Br J Sports Med*. 2016;50(21):1309-1314. http://dx.doi.org/10.1136/bjsports-2015-095850

7. Cook C. Predicting future physical injury in sports: it's a complicated dynamic system. *Br J Sports Med*. 2016;50(22):1356-1357. http://dx.doi.org/10.1136/bjsports-2016-096445

8. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function – part 1. *Int J Sports Phys Ther*. 2014;9(3):396-409.

9. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function – part 2. *Int J Sports Phys Ther*. 2014;9(4):549-563.

10. Ekegren CL, Gabbe BJ, Finch CF. Injury surveillance in community sport: Can we obtain valid data from sports trainers? *Scand J Med Sci Sports*. 2015;25(3):315-322. http://dx.doi.org/10.1111/sms.12216

11. Ekstrand J. Keeping your top players on the pitch: the key to football medicine at a professional level. *Br J Sports Med*. 2013;47(12):723-724. http://dx.doi.org/10.1136/bjsports-2013-092771

12. Ekstrand J, Hägglund M, Waldén M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med*. 2011;45(7):553-558. http://dx.doi.org/10.1136/bjsm.2009.060582

13. Ekstrand J, Waldén M, Hägglund M. Hamstring injuries have increased by 4% annually in men's professional football, since 2001: a 13-year longitudinal analysis of the UEFA Elite Club injury study. *Br J Sports Med*. 2016;50(12):731-737. http://dx.doi.org/10.1136/bjsports-2015-095359

14. Fuller CW, Ekstrand J, Junge A et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Scand J Med Sci Sports*. 2006;16(2):83-92. http://dx.doi.org/10.1111/j.1600-0838.2006.00528.x

15. Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *Br J Sports Med*. 2013;47(12):738-742. http://dx.doi.org/10.1136/bjsports-2013-092215

16. Hammes D, aus der Fünten K, Bizzini M, Meyer T. Injury prediction in veteran football players using the Functional Movement Screen™. *J Sports Sci*. 2016;34(14):1371-1379. http://dx.doi.org/10.1080/02640414.2016.1152390

17. Hegedus EJ, McDonough SM, Bleakley C, Baxter D, Cook CE. 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;49(10):649-656. http://dx.doi.org/10.1136/bjsports-2014-094341

18. Hopkins WG. A spreadsheet for deriving a confidence interval, mechanistic inference and confidence inference from a P-value. *SportScience*. 2007;11:16-20. http://sportsci.org/2007/wghinf.htm Accessed May 10, 2017.

19. Hopkins WG. Linear models and effect magnitudes for research, clinical and practical applications. *SportScience*. 2010;14:49-58. http://sportsci.org/2010/wghlinmod.htm Accessed May 10, 2017.

20. Hopkins WG, Batterham AM. Error rates, decisive outcomes and publication bias with several inferential methods. *Sports Med*. 2016;46(10):1563-1573. http://dx.doi.org/10.1007/s40279-016-0517-x

21. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 2009;41(1):3-13. https://doi.org/10.1249/MSS.0b013e31818cb278

22. Krosshaug T, Steffen K, Kristianslund E et al. The vertical drop jump is a poor screening test for ACL injuries in female elite soccer and handball players: A prospective cohort study of 710 athletes. *Am J Sports Med*. 2016;44(4):874-883. https://doi.org/10.1177/0363546515625048

23. Lloyd RS, Oliver JL, Radnor JM, Rhodes BC, Faigenbaum AD, Myer GD. Relationships between functional movement screen scores, maturation and physical performance in young soccer players. *J Sports Sci*. 2015;33(1):11-19. http://dx.doi.org/10.1080/02640414.2014.918642

24. Malone S, Owen A, Mendes B, Hughes B, Collins K, Gabbett TJ. High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? [published online May 25 2017]. *J Sci Med Sport*. 2017. http://dx.doi.org/10.1016/j.jsams.2017.05.016

25. McArdle S. Psychological rehabilitation from anterior cruciate ligament-medial collateral ligament reconstructive surgery: a case study. *Sports Health*. 2010;2(1):73-77. https://doi.org/10.1177/1941738109357173

26. McCall A, Fanchini M, Coutts AJ. Prediction: The modern day sports science/medicine 'Quest for the holy grail' [published online May 10 2017]. *Int J Sports Physiol Perform*. 2017. https://doi.org/10.1123/ijspp.2017-0137

27. McCall A, Carling C, Nedelec M et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med*. 2014; 48(18):1352-1357. http://dx.doi.org/10.1136/bjsports-2014-093439

28. McCunn R, aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 2016;46(6):763-781. doi:10.1007/s40279-015-0453-1.

29. McCunn R, aus der Fünten K, Govus A, Julian R, Schimpchen J, Meyer T. The intra- and inter-rater reliability of the Soccer Injury Movement Screen (SIMS). *Int J Sports Phys Ther*. 2017;12(1):53-66.

30. McCunn R, Sampson JA, Whalan M, Meyer T. Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement? *Science and Medicine in Football*. 2017;1(1):86-88. http://dx.doi.org/10.1080/24733938.2016.1256554

31. Newton F, McCall A, Ryan D et al. Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players. *Science and Medicine in Football*. 2017;1(2):102-106. http://dx.doi.org/10.1080/24733938.2017.1283436

32. Orchard J, Rae K, Brooks J et al. Revision, uptake and coding issues related to the open access Orchard Sports Injury Classification System (OSICS) versions 8, 9 and 10.1. *Open Access J Sports Med*. 2010;11(1):207-214. https://doi.org/10.2147/OAJSM.S7715

33. Padua DA, DiStefano LJ, Beutler AI, de la Motte SJ, DiStefano MJ, Marshall SW. The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. *J Athl Train*. 2015;50(6):589-595. https://doi.org/10.4085/1062-6050-50.1.10

34. Thorborg K, Krommes KK, Esteve E, Clausen MB, Bartels EM, Rathleff MS. Effect of specific exercise-based football injury prevention programmes on the overall injury rate in football: a systematic review and meta-analysis of the FIFA 11 and 11+

programmes. *Br J Sports Med*. 2017;51(7):562-571. http://dx.doi.org/10.1136/bjsports-2016-097066

35. Timmins RG, Bourne MN, Shield AJ, Williams MD, Lorenzen C, Opar DA. Short biceps femoris fascicles and eccentric knee flexor weakness increase the risk of hamstring injury in elite football (soccer): a prospective cohort study. *Br J Sports Med*. 2016;50(24):1524-1535. http://dx.doi.org/10.1136/bjsports-2015-095362

36. van der Horst N, van de Hoef S, Reurink G, Huisstede B, Backx F. Return to play after hamstring injuries: A qualitative systematic review of definitions and criteria. *Sports Med*. 2016;46(6):899-912. http://dx.doi.org/10.1007/s40279-015-0468-7

37. van Dyk N, Clarsen B. Prevention forecast: cloudy with a chance of injury [Epub ahead of print]. *Br J Sports Med*. 2017. http://dx.doi.org/10.1136/bjsports-2017-097861

38. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Med*. 1992;14(2):82-99. https://doi.org/10.2165/00007256-199214020-00002

39. Waldén M, Hägglund M, Ekstrand J. Time-trends and circumstances surrounding ankle injuries in men's professional football: an 11-year follow-up of the UEFA Champions League injury study. *Br J Sports Med*. 2013;47(12):748-753. http://dx.doi.org/10.1136/bjsports-2013-092223

40. Woods CT, McKeown I, Keogh J, Robertson S. The association between fundamental athletic movements and physical fitness in elite junior Australian footballers [Epub ahead of print]. *J Sports Sci*. 2017. http://dx.doi.org/10.1080/02640414.2017.1313996

5. INJURY DATA COLLECTION PROCEDURES WITHIN LOWER LEAGUES

5.1 Rationale for the publication

The large-scale prospective cohort study conducted as part of this research degree (chapter 4) presented numerous logistical and methodological challenges. While designing this study a significant issue concerning how best to collect the required injury data was encountered. With over 300 players from 12 different clubs participating in the study and considering the part-time nature of semi-professional football it was impossible for a medical professional to see each player every day (or even on a regular basis). However, this is precisely the degree of contact required to strictly follow the football injury research consensus guidelines.[16] While the current consensus statement serves full-time professional football well in terms of suggested best practice for collecting and recording injuries it does not translate to non-professional environments in the same way. The numerous challenges along with suggested solutions related to injury data collection in non-professional football environments were discussed in the following point-counterpoint article.

5.2 Issues related to conducting a prospective cohort study within a semi-professional environment

The following section contains the submitted manuscript pertaining to the following publication:

McCunn R, Sampson JA, Whalan M, Meyer T. Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement? *Science and Medicine in Football*. 2017;1(1):86-88.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of *Science and Medicine in Football*. The citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement?

Robert McCunn[a], John A Sampson[b], Matthew Whalan[b, c, d], Tim Meyer[a]

[a] *Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany*

[b] *Centre for Human and Applied Physiology, School of Medicine, University of Wollongong, Wollongong, New South Wales, Australia*

[c] *Player Welfare Department, Football South Coast, New South Wales, Australia*

[d] *Wollongong Wolves FC, Wollongong, New South Wales, Australia*

Corresponding author: Robert McCunn

Institute of Sport and Preventive Medicine
Saarland University
GEB. B8.2
66123 Saarbrücken
Germany
Email: bob.mccunn@me.com
Phone: +49681-302 70400 Fax: +49681-302 4296

Running head: Lower league injury data collection

Abstract

In 2006 a consensus statement recommending how football injury data should be collected and reported was published. These recommendations have provided a useful framework for research over the last 10 years. However, many questions related to the underlying methodology of studies concerned with injury epidemiology and prevention in football still exist. This is particularly true for research conducted in non-professional environments. The present point-counterpoint article highlights some of these issues and asks the question: are we in need of an updated consensus statement?

Key words: epidemiology, injury, semi-professional, amateur, soccer

Football participation carries an inherent risk of injury. Consistent and accurate records of injury are imperative for planning, evaluating and delivering injury prevention programs such as the FIFA 11+ (van Mechelen, Hlobil & Kemper 1992). In this regard, Fuller et al.'s (2006) consensus statement serves research within full-time professional football well by streamlining research methods and providing definitions/recommendations for injury data collection. Furthermore, Fuller et al.'s (2006) definition of injury (any physical complaint sustained as a result of football activities) and the proposed injury incident data recording sheets are appropriate for both full- and part-time players. However, problems are encountered when applying some of the guidelines to "part-timers", who cumulatively compose the largest playing group (e. g. youth academy, semi-professional and recreational/community level players). In our view there are two main issues at the forefront of injury research conducted within part-time groups that require clarification; (1) who should record the data and (2) how should injury severity be quantified? A number of solutions to each question may exist. This point-counterpoint article presents various options for discussion in an attempt to ensure future injury research conducted within part-time playing groups is of high methodological quality.

Data collection

Collecting accurate and reliable data is essential if one hopes to present valid results that allow consistent comparisons across the literature. However, within existing studies of part-time playing groups there are discrepancies. For example, the data has been collected by a variety of individuals including: 1) coaches (Ekstrand & Hilding 1999; Froholdt, Olsen & Bahr 2010), 2) parents (Emery, Meeuwisse & Hartmann 2005; Emery & Meeuwisse 2006), 3) medical staff (aus der Fünten, Faude, Lensch & Meyer 2014; Brito et al. 2012; Silvers et al. 2015; Herrero, Salinero & Del Coso 2014) and 4) academic researchers (Hammes, aus der Fünten, Bizzini & Meyer 2016; McNoe & Chalmers 2010; Schmikli, de Vries, Inklaar & Backx 2011). The varying depth of medical knowledge and skills among these parties may consequently lead to inconsistencies in data collection and present difficulties when comparing outcomes. For example, all parties would most likely correctly record certain types of 'dramatic' injuries such as bone breaks; however, discrepancies may arise when more ambiguous cases occur e.g. minor muscle strains. Indeed, certain groups may over- or under-report the incidence of injuries depending on their relationship with the players and their specific role within the club. For example, players may withhold physical complaints from their coach for fear of not being selected to play in matches or student research assistants may not be assertive enough when attempting to obtain relevant injury information from players they are unfamiliar with. Medical practitioners (e.g. doctors and physiotherapists) are undoubtedly the most qualified to diagnose and record injuries. However, the majority of part-time teams do not have immediate access to medical practitioners. An injury recording method that relies on individuals with this level of expertise would therefore, in many cases, make large-scale research projects untenable. Alternatively, player self-reporting of exposure time and injury incidence (e. g. via online resources) removes the requirement for any third party to record the information. However, self-reporting arguably raises questions over compliance and data accuracy. In our view, the most accurate data can be collected from a third party present at training and on match days. Hence, we suggest each team must appoint a primary data collector who would record basic details such as: exposure, injury location, incident description and symptoms. Much of the research highlighted above has applied this method; however, we propose that this method should be improved by stipulating that the nominated individual must posses a minimum standard of medical knowledge and undergo training with respect to injury data collection procedures. Such an approach would ensure the recording of detailed injury descriptions, facilitating retrospective injury diagnosis by trained medical professionals (Hammes, aus der Fünten, Bizzini & Meyer 2016). Allowing non-experts such as parents, coaches and university students to contribute in this manner offers a viable strategy to ensure high quality data is collected, and for research teams to conduct meaningful large-scale projects. The next step in this process would be for the scientific community to agree upon the extent and content of the required training. Some variability in the methodology between studies is inevitable. Hence, the importance of meticulously describing data collection protocols cannot be overstated, since it is vital in allowing readers to judge the quality of the results presented.

Injury quantification

Reporting injury incidence per 1000 hours of exposure as recommended by Fuller et al. (2006) seems appropriate for part-time playing groups. However, Fuller et al.'s (2006) guidelines for recording injury severity (days lost from full participation in training/match play) are problematic in a part-time environment where, in contrast to full-time athletes, players are not seen by medical staff on a daily basis. Indeed several days between training/match days may pass when the part-time player is not seen by anyone associated with the club and the opportunity to observe a "return to full participation" may only arise once or twice per week. One potential method of determining injury severity is via the player self-reporting recovery by providing the date they believe they were fit to return to full participation. Similarly, players could report perceived injury severity e.g. using a Likert-type scale via an online questionnaire to indicate how serious they believe any given case is (Clarsen, Rønsen, Myklebust, Flørenes & Bahr 2014). Alternatively, follow up appointments with injured players by medical practitioners (either via phone or in-person) on non-training/playing days could be used. However, each of these options pose compliance and logistical implementation issues. We suggest that the definition proposed by Fuller et al. (2006) should apply even though adopting such an approach would almost certainly over-estimate injury severity. The greatest impact of over-estimation would be observed at the less serious end of the spectrum. For example, in the event of a player suffering an injury on a Saturday match-day, missing their only scheduled training session the following Tuesday and returning to play the subsequent Saturday match-day, a six-day lay-off would be recorded. However, the player may have been fit to train/play by Wednesday meaning the injury should have been classified as a three-day time-loss injury. As such, some 'minimal' injuries may be erroneously recorded as 'mild' (Fuller et al. 2006). However, upholding the current definition consistently across the literature would, at least, provide comparable data sets.

Conclusion

In summary, while a number of solutions to the highlighted questions may exist, it is important to acknowledge that each is somewhat flawed. Agreement surrounding the recommendations for injury data collection procedures among part-time playing groups is thus crucial. Such consensus will allow comparisons between studies and ensure that the conclusions drawn from future research are meaningful. Any updated consensus statement should carefully consider the logistical implications for researchers when making recommendations. Herein, the arguments have been presented with a focus on football. However, the issues presented apply across sports and as such epidemiologists and applied practitioners from different domains would add value to this discussion.

References

aus der Fünten, K., Faude, O., Lensch, J., & Meyer, T. (2014) Injury characteristics in the German professional male soccer leagues after a shortened winter break. *Journal of Athletic Training, 49*(6), 786-793. doi:10.4085/1062-6050-49.3.51

Brito, J., Malina, R. M., Seabra, A., Massada, J. L., Soares, J. M., Krustrup, P., & Rebelo, A. (2012) Injuries in Portuguese youth soccer players during training and match play. *Journal of Athletic Training, 47*(2), 191-197.

Clarsen, B., Rønsen, O., Myklebust, G., Flørenes, T. W., & Bahr, R. (2014) The Oslo Sports Trauma Research Center questionnaire on health problems: a new approach to prospective monitoring of illness and injury in elite athletes. *British Journal of Sports Medicine, 48*(9), 754-760. doi:10.1136/bjsports-2012-092087

Ekstrand, J., & Hilding, J. (1999) The incidence and differential diagnosis of acute groin injuries in male soccer players. *Scandinavian Journal of Medicine & Science in Sports, 9*, 98-103. doi:10.1111/j.1600-0838.1999.tb00216.x

Emery, C. A., & Meeuwisse, W. H. (2006) Risk factors for injury in indoor compared with outdoor adolescent soccer. *The American Journal of Sports Medicine, 34*(10), 1636-1642. doi:10.1177/0363546506288018

Emery, C. A., Meeuwisse, W. H., & Hartmann, S. E. (2005) Evaluation of risk factors for injury in adolescent soccer: implementation and validation of an injury surveillance system. *The American Journal of Sports Medicine, 33*(12), 1882-1891. doi:10.1177/0363546505279576

Froholdt, A., Olsen, O. E., & Bahr, R. (2010) Low risk of injuries among children playing organized soccer. *The American Journal of Sports Medicine, 37*(6), 1155-1160. doi:10.1177/0363546508330132

Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., … Meeuwisse, W. H. (2006) Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *British Journal of Sports Medicine, 40*, 193-201. doi:10.1136/bjsm.2005.025270

Hammes, D., aus der Fünten, K., Bizzini, M., & Meyer, T. (2016) Injury prediction in veteran football players using the Functional Movement Screen™. *Journal of Sports Sciences*. doi:10.1080/02640414.2016.1152390 [E-pub ahead of print]

Herrero, H., Salinero, J. J., & Del Coso, J. (2014) Injuries among Spanish male amateur soccer players: a retrospective population study. *The American Journal of Sports Medicine, 42*(1), 78-85. doi:10.1177/0363546513507767

McNoe, B. M., & Chalmers, D. J. (2010) Injury in community-level soccer: development of an injury surveillance system. *The American Journal of Sports Medicine, 38*(12), 2542-2551. doi:10.1177/0363546510377415

Schmikli, S. L., de Vries, W. R., Inklaar, H., & Backx, F. (2011) Injury prevention target groups in soccer: injury characteristics and incidence rates in male junior and senior players. *Journal of Science and Medicine in Sport, 14*(3), 199-203. doi:10.1016/j.jsams.2010.10.688

Silvers-Granelli, H., Mandelbaum, B., Adenji, O., Insler, S., Bizzini, M., Pohlig, R., … Dvorak, J. (2015) Efficacy of the FIFA 11+ injury prevention program in the collegiate male soccer player. *The American Journal of Sports Medicine*. doi:10.1177/0363546515602009 [E-pub ahead of print]

van Mechelen, W., Hlobil, H., & Kemper, H. C. (1992) Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Medicine, 14*(2), 82-99. doi:10.2165/00007256-199214020-00002

6. GENERAL DISCUSSION

6.1 Summary of findings

The work included within the present thesis revealed a number of findings relevant for both applied practitioners and academic researchers. Firstly, numerous movement screens were identified within the scientific literature and collectively summarized (chapter 1); providing applied practitioners with a resource to assist with decision-making related to which, if any, screen to use. The reliability of the identified movement screens varied but was acceptable for the most part. However, the majority of the identified movement screens had no supporting evidence related to their association with injury. Studies that investigated the association with injury of the FMS™ and Landing Error Scoring System (LESS) were identified; however, the results were equivocal with some studies reporting an association and some not. It was concluded that none of the movement screens identified had enough supporting evidence to warrant the moniker of "injury prediction tool".

The literature review conducted as part of chapter 1 revealed that the FMS™ was the most heavily researched movement screen of those identified. Despite this canon of research, a paucity of studies directly addressed the issue of whether FMS™ scores were associated with injury among football players specifically. As a result, the study outlined in chapter 2 sought to investigate the association with injury of the FMS™ among high-level youth football players. No association with injury was observed.

The SIMS was developed as a football-specific movement screen with the rationale that an assessment targeting the most common sites of football-related injury may provide greater insight into any potential relationship between movement quality and injury risk among footballers. Initially, the intra- and inter-rater reliability of the SIMS was assessed since test reliability is a prerequisite for validity.[4] The SIMS demonstrated acceptable reliability for use in further research and applied practice (chapter 3).

Following on from the reliability study described in chapter 3, a prospective cohort study was subsequently designed and conducted to investigate the association with injury of the SIMS within a football population. Over 300 semi-professional football players performed the SIMS during the pre-season period and injury incidence was recorded during the in-season period. However, no association with injury was observed for the SIMS composite score (chapter 4). These findings do not support the creation of high and low risk of injury groups based on SIMS score.

Finally, during the design phase of the prospective cohort study described in chapter 4, a number of logistical and methodological issues related to injury data collection in a non-professional environment were encountered. The challenges and potential solutions to conducting this type of research are presented and discussed in chapter 5. While the football injury research consensus guidelines[16] serve professional football well, they do not translate to the non-professional environment neatly. Within the professional setting, club medical staff can assess players' injury status on a daily basis. Yet, this is not possible when players train and compete on a part-time basis. Alternative solutions to recording and diagnosing football related injuries in non-professional environments include the use of non-medically trained primary data collectors and retrospective diagnosis via detailed injury event descriptions.

In summary, the present findings suggest that movement quality, specifically when assessed using the FMS™ and the SIMS, is not associated with injury among football players. Furthermore, when conducting injury research within non-professional football environments, the use of non-medically trained primary data collectors and retrospective diagnosing via detailed injury event descriptions represent viable methodological options.

6.2 If movement screens are not predictive of injury, do they still have value?

The usefulness of screening in general, not just limited to movement screening, has been questioned recently.[2] The main tenet from the review by Bahr[2] was that screening tests used within the scope of sports medicine do not demonstrate diagnostic values sufficient to call them predictive. However, exactly how good a diagnostic tool needs to be before it can be deemed 'predictive' remains unclear. In addition, while movement screening does not appear to be a useful tool for creating high and low risk of injury groups within football, it may be helpful to researchers and applied practitioners in other ways. In response to Bahr's[2] review article on the subject of screening and injury risk stratification a correspondence manuscript was submitted to the same journal to highlight this point.

The following section contains the submitted manuscript pertaining to the following publication:

McCunn R, Meyer T. Screening for risk factors: If you liked it then you should have put a number on it. *Br J Sports Med*. 2016;50:1354.

The citations and references contained herein apply to this manuscript only and are formatted to the requirements of the *British Journal of Sports Medicine*. The numerical citations relate to the reference list within this section only and not to the reference list included at the end of this thesis.

Screening for risk factors: if you liked it then you should have put a number on it

Robert McCunn[1] & Tim Meyer[1]

[1] Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany

Corresponding Author: Robert McCunn
Institute of Sport and Preventive Medicine
Saarland University
Campus Bldg. B8.2
66123 Saarbrücken, Germany
Phone: +49 681 302 70410
Email: robert.mccunn@uni-saarland.de
Fax: +49 (0) 681 302 4296

Submission type: Correspondence
In relation to Bahr (2016) Why screening tests to predict injury do not work – and probably never will…: a critical review.

Word count: 171

The usefulness of screening tests to predict injury has recently been questioned.[1, 2] However, from our point of view it is important that screening not be completely demonised. A screening test may not accurately identify who will get injured (albeit providing a likelihood) but this should not result in obsolescence. Usefulness is not solely dependent on predictive ability. As Bahr [1] highlights: causative relationships (i.e. injury risk factors) can be identified using screening tests e.g. eccentric hamstring weakness.[3] Should we therefore completely admonish screening tests because we cannot definitively say that poor eccentric hamstring strength will 100% result in a strain injury? Screening puts a number on an attribute that allows us to quantify injury risk and in-turn modify the design of injury prevention strategies. Clearly the syntax surrounding screening tests needs adapted. Bahr[1] is absolutely correct that the practice of applied practitioners using screening results to categorise their athletes into intervention and 'control' groups is not supported by the evidence but let's not throw the baby out with the bathwater.

References

1. Bahr R. Why screening tests to predict injury do not work – and probably never will…: a critical review. *Br J Sports Med* Published Online First: 19 Apr 2016 doi:10.1136/bjsports-2016-096256

2. McCunn R, aus der Fünten K, Fullagar HHK *et al*. Reliability and association with injury of movement screens: a critical review. *Sports Med* Published Online First: 31 Dec 2015 doi:10.1007/s40279-015-0453-1

3. Timmins RG, Bourne MN, Shield AJ, *et al*. Short biceps femoris fascicles and eccentric knee flexor weakness increase the risk of hamstring injury in elite football (soccer): a prospective cohort study. *Br J Sports Med* Published Online First: 16 Dec 2015 doi:10.1136/bjsports-2015-095362

6.3 How might movement screening still be useful from an injury perspective?

Bittencourt et al.[5] challenged the concept of simply identifying risk factors for injury via prospective cohort studies, at least in the context of prediction. The aetiology of football-related injury is multifaceted.[1, 3] Furthermore, the interaction between risk factors is not linearly cumulative.[5] Indeed, different combinations of risk factors and circumstances may result in the same risk profile.[5] When viewed from this perspective it is easy to appreciate how difficult the task of predicting injury truly is. However, reducing the incidence of injury within football is not necessarily dependent on the ability to precisely predict who will get injured and when. The widely accepted model of sports injury prevention presented by van Mechelen et al.[30] identifies four key stages: 1) establish the extent of the problem (epidemiology), 2) establish aetiology (risk factors and mechanisms), 3) introduce preventive measures, and 4) assess their effectiveness. In order to create effective preventive measures (as described in stage 3), researchers must identify the risk factors they can realistically address and subsequently create interventions that influence them. As a result, while prospective cohort studies that seek to identify individual risk factors (for example: chapter 4) may not be suitable to address the question of prediction, they can contribute towards injury reduction strategies by highlighting attributes that should be incorporated into preventive interventions. While the appeal of individually targeted injury reduction interventions is clear, current evidence does not support such practice. Rather, universal interventions – whose design should be based on available evidence – that apply to all players likely offer the optimal approach to reducing injury incidence within football. Indeed, universal interventions such as the FIFA 11+ warm-up programme and generic Nordic hamstring curl protocols have demonstrated effectiveness in terms of reducing injury incidence.[26, 28]

6.4 How might movement screening be useful to applied practitioners in other ways?

While not the primary focus of the present research it is important to acknowledge that movement screening is potentially useful to applied practitioners in ways unrelated to injury reduction. Screening allows the establishment of a baseline value and quantification of a player's movement quality when in a 'healthy' state.[29] This may be useful when making return to play decisions after the rehabilitation from any potential injury. Movement screening also offers applied practitioners the opportunity to build rapport with unfamiliar players.[29] In the context of professional youth academies, strength and conditioning practitioners and physiotherapists are faced with the challenge of a continual turnover of players each year. Movement screening may offer a quick, logistically viable and systematic method of quantifying movement quality and in doing so help determine readiness for introduction to formalized resistance training and progressions to more advanced exercise techniques. This may be particularly helpful in guiding the physical development plans of newly recruited players whom club support staff are not familiar with.

6.5 General limitations

A number of limitations within the current body of work should be considered when interpreting the findings. Firstly, the type of movement screening discussed herein relies on predominantly subjective judgment by the test rater. While the scoring guidelines and criteria help to guide the rater, ultimately their opinion determines the assigned score. This is a notable consideration since it relates to test validity. A relevant question is: do subjectively assessed movement screen scores align with scores based on objectively measured movement quality criteria, for example: joint angles and alignment quantified using advanced camera equipment? However, the accessibility and simplicity of subjectively assessed movement screens make them an attractive type of assessment for applied practitioners within football. Furthermore, when prospectively investigating potential risk factors for injury it is important to consider not only the risk factor in question and injury incidence but also the exposure time of each individual. Without incorporating the amount of time players are exposed to training and playing into the statistical analysis, a significant contribution to injury risk is being omitted. This perhaps represented the greatest limitation of the prospective cohort study described in chapter 2. However, this limitation was addressed and overcome in the prospective cohort study described in chapter 4. In addition, investigating potential risk factors in isolation, as in this body of work, provides limited insight into the multifactorial nature of sports injury. It is acknowledged that many factors, both intrinsic and extrinsic, contribute to injury risk and understanding the complex interactions between these factors would represent an advancement of knowledge. However, the necessary large sample sizes and resultant logistical challenges associated with such potential studies make this next leap very challenging. Realistically, multi-centre collaboration will be required to conduct studies capable of investigating these complex interactions between risk factors and injury.[5, 12, 19]

6.6 Future research directions

Considering the widespread use of movement screening within professional football the paucity of related research within this population is perhaps surprising. The present body of work encompassed studies conducted with elite youth footballers representing a professional club academy and adult semi-professional players; however, scope exists to conduct similar investigations with senior professional players. Furthermore, increasing the scoring objectivity of movement screens using accessible and affordable filming equipment represents an avenue for future research. Of the 10 screens identified within the present literature review (chapter 1) only two (FMS™ and LESS) had been prospectively investigated in relation to their association with injury. Krosshaug et al.[20] subsequently reported that the drop vertical jump was a poor predictor of anterior cruciate ligament injury. However, the majority of the identified screens have not had their association with injury investigated and this offers another potential avenue of research. The interaction and combination of potential risk factors in terms of improving injury prediction ability provides a complex yet hitherto relatively unexplored topic worthy of future study. Finally, as Bahr[2] has raised previously: conducting randomized controlled trials to investigate the effectiveness of screening and targeting intervention programmes versus universal implementation of injury prevention strategies represents an important step in clarifying the efficacy of using movement screening to risk stratify individuals.

6.7 Practical recommendations and conclusions

The outcomes arising from the present body of work highlight that numerous movement screens demonstrating acceptable reliability exist; however, such assessments do not appear to be strongly associated with injury. Furthermore, the newly developed SIMS presented herein was not associated with injury among a semi-professional cohort of football players. The present results challenge current convention that advocates risk stratification and targeted intervention based on screening score.[18] Rather, prospective cohort studies identifying risk factors should be used to inform the content of universal injury prevention programmes that are performed by all players. It is important to acknowledge that the lack of association with injury for the FMS™ and the SIMS does not preclude their use in the applied field. Movement screening may prove useful for applied practitioners in a number of other ways including: establishing healthy baseline values useful during potential return to play decision-making, building rapport with players, identifying current injuries and informing appropriate resistance training programme content. Both applied practitioners and researchers should be able to clearly articulate their rationale for using movement screening with football players to ensure that all parties involved have appropriate expectations regarding the usefulness of the resultant data. Informing players and coaches that movement screening can predict who will or will not get injured and targeting preventive interventions only to so-called high-risk groups may ultimately harm credibility if and when supposedly low-risk players get injured.

7. REFERENCES

1. Arnason A, Sigurdsson SB, Gudmundsson A, Holme I, Engebretsen L, Bahr R. Risk factors for injuries in football. *Am J Sports Med*. 2004;32(1 Suppl):5S-16S.

2. Bahr R. Why screening tests to predict injury do not work – and probably never will…: a critical review. *Br J Sports Med*. 2016;50(13):776-780.

3. Bahr R, Holme I. Risk factors for sports injuries – a methodological approach. *Br J Sports Med*. 2003;37(5):384-392.

4. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. *Phys Ther Sport*. 2003;4(3):122-128.

5. Bittencourt NF, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition – narrative review and new concept. *Br J Sports Med*. 2016;50(21):1309-1314.

6. Carling C, Le Gall F, McCall A, Nédélec M, Dupont G. Squad management, injury and match performance in a professional soccer team over a championship-winning season. *Eur J Sport Sci*. 2015;15(7):573-582.

7. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function – part 1. *Int J Sports Phys Ther*. 2014;9(3):396-409.

8. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function – part 2. *Int J Sports Phys Ther*. 2014;9(4):549-563.

9. D'Hooghe M. Why is UEFA carrying out injury studies? *Br J Sports Med*. 2016;50(12):707.

10. Eirale C, Tol JL, Farooq A, Smiley F, Chalabi H. Low injury rate strongly correlates with team success in Qatari professional football. *Br J Sports Med*. 2013;47(12):807-808.

11. Ekstrand J. Keeping your top players on the pitch: the key to football medicine at a professional level. *Br J Sports Med*. 2013;47(12):723-724.

12. Ekstrand J. Preventing injuries in professional football: thinking bigger and working together. *Br J Sports Med*. 2016;50(12):709-710.

13. Ekstrand J, Hägglund M, Waldén M. Epidemiology of muscle injuries in professional football (soccer). *Am J Sports Med*. 2011;39(6):1226-1232.

14. Ekstrand J, Hägglund M, Waldén M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med*. 2011;45(7):553-558.

15. Finch CF, Cook J, Kunstler BE, Akram M, Orchard J. Subsequent injuries are more common than injury recurrences: An analysis of 1 season of prospectively collected injuries in professional Australian football. *Am J Sports Med*. 2017;45(8):1921-1927.

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;16(2):97-106.

17. Hägglund M, Waldén M, Magnusson H, Kristenson K, Bengtsson H, Ekstrand J. Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *Br J Sports Med*. 2013;47(12):738-742.

18. Hewett TE. Response to: 'Why screening tests to predict injury do not work – and probably never will…: a critical review'. *Br J Sports Med*. 2016;50(21):1353.

19. Impellizzeri FM. Together we are stronger: Multicenter studies. *Int J Sports Physiol Perform*. 2017;12(2):141.

20. Krosshaug T, Steffen K, Kristianslund E, et al. The vertical drop jump is a poor screening test for ACL injuries in female elite soccer and handball players: A prospective cohort study of 710 athletes. *Am J Sports Med*. 2016;44(4):874-883.

21. Leventer L, Eek F, Hofstetter S, Lames M. Injury patterns among elite football players: A media-based analysis over 6 seasons with emphasis on playing position. *Int J Sports Med*. 2016;37(11):898-908.

22. 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;19(5):347-365.

23. McArdle S. Psychological rehabilitation from anterior cruciate ligament-medial collateral ligament reconstructive surgery: A case study. *Sports Health*. 2010;2(1):73-77

24. McCall A, Carling C, Nédélec M, et al. Risk factors, testing and preventive strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med*. 2014;48(18):1352-1357.

25. Øiestad BE, Engebretsen L, Storheim K, Risberg MA. Knee osteoarthritis after anterior cruciate ligament injury: a systematic review. *Am J Sports Med*. 2009;37(7):1434-1443.

26. Thorborg K, Krommes KK, Esteve E, Clausen MB, Bartels EM, Rathleff MS. Effect of specific exercise-based football injury prevention programmes on the overall injury rate in football: a systematic review and meta-analysis of the FIFA 11 and 11+ programmes. *Br J Sports Med*. 2017;51(7):562-571.

27. Toohey LA, Drew MK, Cook JL, Finch CF, Gaida JE. Is subsequent lower limb injury associated with previous injury? A systematic review and meta-analysis. *Br J Sports Med*. 2017 [Epub ahead of print].

28. van der Horst N, Smits DW, Petersen J, Goedhart EA, Backx FJ. The preventive effect of the Nordic hamstring exercise on hamstring injuries in amateur soccer players: a randomized controlled trial. *Am J Sports Med*. 2015;43(6):1316-1323.

29. van Dyk N, Clarsen B. Prevention forecast: cloudy with a chance of injury. *Br J Sports Med*. 2017 [Epub ahead of print].

30. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Med*. 1992;14(2):82-99.

31. Waldén M, Hägglund M, Magnusson H, Ekstrand J. ACL injuries in men's professional football: a 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after ACL rupture. *Br J Sports Med*. 2016;50(12):744-750.

32. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.

33. Windholz E. Professional sport, work health and safety law and reluctant regulators. *Sports Law eJournal*. 2015. Available from: http://epublications.bond.edu.au/slej/28/ [Accessed September 1, 2017].

8. APPENDICES

8.1 Assessment of methodological quality for reliability studies raw data (chapter 1)

The methodological quality of the identified reliability studies was assessed using the COSMIN checklist described by Terwee et al. (2011). Only the reliability section (box B) was completed since this was the most relevant aspect. Box B contained 14 questions related to study design. According to the checklist scoring system a 'worst score counts' approach should be taken, meaning that if one of the 14 questions is answered as 'poor' the study is rated as such. The scoring options range from 'excellent' to 'poor' on a four-point scale, however, not all questions in box B use the entirety of this scale. For example, question one can only be answered as either 'excellent' or 'good' whereas question three can be answered as either 'excellent', 'good', 'fair', or 'poor'. Due to the 'worst score counts' approach the questions where 'poor' was a potential answer (Q3, 4, 5, 7, 8, 9, 10, 11, 12, 13) were addressed first for each study (in numerical order) and if any of these questions were scored as 'poor' then no further questions were answered in accordance with the scoring protocol. The first question (in numerical order) recorded as 'poor' was noted below. If no questions were answered as 'poor' then the first question (in numerical order) recorded as 'fair' was noted below. If no questions were answered as 'poor' or 'fair' then the first question (in numerical order) answered as 'good' was noted below. If no questions were answered as 'poor', 'fair' or 'good' then the first question (in numerical order) to be answered as 'excellent' was noted below. The worst score for each study is documented in this appendix. In some cases a very brief explanation may be given for the decision where ambiguity existed.

FMS

Hotta et al. (2015)	Q3 = poor
Gulgin & Hoogenboom (2014)	Q3 = poor
Letafatkar et al. (2014)	Q3 = poor
Parenteau-G et al. (2014)	Q3 = poor
Elias (2013)	Q3 = poor
Gribble et al. (2013)	Q3 = poor
Shultz et al. (2013)	Q3 = fair
Smith et al. (2013)	Q3 = poor

Frohm et al. (2012)	Q3 = poor
Klusemann et al. (2012)	Q3 = poor
Onate et al. (2012)	Q3 = poor
Teyhen et al. (2012)	Intra-rater reliability study Q3 = poor *
	Inter-rater reliability study Q1 = good *
Butler et al. (2012)	Q3 = fair
Schneiders et al. (2011)	Q3 = poor
Chorba et al. (2010)	Q3 = poor
Minick et al. (2010)	Q3 = fair

* Intra-rater reliability sample size was small (<30 participants) resulting in 'poor' classification. The number of participants (64) listed in Table 3 refers to the overall sample size used for the calculation of inter-rater reliability. Eighteen participants were selected from the overall pool of 64 and performed the FMS a second time to allow calculation of intra-rater reliability.

LESS

Smith et al. (2012)	Q3 = poor
Padua et al. (2011)	Q3 = fair
Onate et al. (2010)	Q3 = poor
Padua et al. (2009)	Q1 = good*

* It can be safely assumed that there were no missing items e.g. none of the 50 participants were removed from the analysis and each was assigned a score by all raters so all items on the LESS scoring sheet must have been addressed. However, this was not explicitly stated so it was decided that question one must be answered as 'good'.

Single-leg squat screens

Junge et al. (2012)	Q1 = good*
Crossley et al. (2011)	Q3 = poor
Örtqvist et al. (2011)	Intra-rater reliability study Q3 = fair
	Inter-rater reliability study Q3 = poor**
Ageberg et al. (2010)	Q3 = poor
Chmielewski et al. (2007)	Q3 = poor

* The number of missing items was partially described e.g. two participants were removed from the sample due to pre-existing medical conditions. Of the 72 participants who did take part it can be assumed that no missing items (in terms of test scores) existed since all individuals' scores were included in the statistical analysis, however, this was not explicitly stated so it was decided that question one must be answered as 'good'.

** Inter-rater reliability sample size was small (<30 participants) resulting in 'poor' classification. The number of participants (33) listed in Table 4 refers to the sample size used for the calculation of intra-rater reliability.

DVJ screens

Nilstad et al. (2014)	Q13 = fair
Whatman et al. (2013)	Q3 = poor
Ekegren et al. (2009)	Q9 = poor*

* The test conditions were not similar for the two intra-reliability conditions. One was performed collectively and the other performed in isolation at the raters' home. Similarly, the raters were allowed to rate the videos at home for one of the inter-rater testing sessions meaning that it is unclear whether they all abided by the rating guidelines e.g. no pausing of videos.

Tuck jump assessment

Herrington et al. (2013)	Q3 = poor
Dudley et al. (2013)	Q3 = fair

AAA

McKeown et al. (2014)	Q3 = poor

CSMT

Parsonage et al. (2014)	Q3 = fair

NMST

Reid et al. (2014)	Q3 = poor

16-PPM

Tarara et al. (2014)	Q3 = poor

SEBT

Ness et al. (2015)	Q9 = fair*

* It is unclear if the test conditions were similar for each rater. It is unclear how strictly the rater guidelines were followed e.g. how many times the raters watched the clips, whether or not they paused any clips or where the ratings took place.

8.2 Assessment of methodological quality for injury studies raw data (chapter 1)

Reference	Question number														Quality score (%)	Level of evidence
	1	2	3	5	6	7	10	11	12	18	20	21	22	25		
Bardenett et al. (2015)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	87	2++
Garrison et al. (2015)	1	1	1	1	1	1	0	1	1	1	1	1	1	0	80	2++
Hotta et al. (2015)	1	1	1	1	1	1	0	1	1	1	1	1	0	1	80	2++
McGill et al. (2015)	1	1	0	1	1	1	1	1	1	1	1	1	1	0	80	2++
Teyhen et al. (2015)	1	1	1	0	1	1	1	1	1	1	1	1	0	0	73	2+
Warren et al. (2015)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	80	2++
Zalai et al. (2015)	1	1	1	0	1	1	0	1	1	1	1	1	1	0	73	2+
Dossa et al. (2014)	1	1	1	0	1	1	1	1	1	1	1	1	1	0	80	2++
Kiesel et al. (2014)	1	1	0	0	1	1	0	1	1	1	1	1	0	0	60	2+
Knapik et al. (2014)	1	1	1	1	1	1	0	1	1	1	1	1	0	0	73	2+
Letafatkar et al. (2014)	1	1	1	0	1	1	1	1	1	1	1	1	0	0	73	2+
Shojaedin et al. (2014)	1	0	1	0	1	0	1	1	1	1	0	1	0	0	53	2+
Butler et al. (2013)	1	1	0	0	1	1	0	1	1	1	1	1	0	0	60	2+
Lisman et al. (2013)	1	1	0	0	1	1	1	1	1	1	1	0	0	0	60	2+
McGill et al. (2012)	1	0	0	1	1	1	1	1	1	1	0	1	1	1	73	2+
O'Connor et al. (2011)	1	1	0	0	1	1	0	1	1	1	1	1	1	0	67	2+
Chorba et al. (2010)	1	1	1	0	1	1	1	1	1	1	1	1	1	0	80	2++
Kiesel et al. (2007)	1	0	0	0	1	1	0	1	1	1	0	1	1	0	53	2+
Padua et al. (2015)	1	1	1	1	1	1	0	1	1	1	1	1	0	0	73	2+
Smith et al. (2012)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	87	2++

8.3 Description of the Soccer Injury Movement Screen (SIMS) (chapter 3)

Movement name	Rationale/perceived usefulness	Instructions
Pre-assessment	N/A	"For each exercise you have three practice attempts and three scored attempts on each leg. In the case of the tuck jump you have three practice jumps followed by the scored 10 second effort."
Anterior reach	- Provides an indication of ankle mobility (dorsi flexion) - Highlights limb asymmetry (ankle mobility and/or leg strength) - Provides an indication of single-leg control	"Remove your shoes. Place the big toe of your standing leg so it is touching the back of the taped line. Place hands on your hips. Reach the toes of the other leg as far along the measuring tape as possible – hovering around 5 centimeters off the ground. You must keep your standing foot in contact with the floor throughout, e.g. you cannot rise up on to your toes. Try to hover at the point of maximal reach for a couple of seconds to allow scoring. You must return to the start position for the attempt to be counted. Likewise, you must maintain balance throughout each attempt for the score to be recorded."*
Single-leg deadlift	- Provides an indication of ability to simultaneously flex and extend at the hip with extended knees while maintaining neutral spinal alignment - Provides an indication of hamstring flexibility - Provides an indication of single-leg control	"Put your shoes back on. Tuck your t-shirt into your shorts. Stand on the middle of the cross, taped on the floor, and cross arms over your chest. Imagine a straight line between your head and your right heel. Try to hinge at the hip while keeping that line straight until parallel to the floor. Try to keep your standing leg (left) extended. Return to the start position with both feet touching the floor between each repetition." Switch the words 'right' and 'left' when instructing the participant when testing the other side.

* If available, a slider device (e.g. Y Balance Test Kit™) can be used to perform the anterior reach.

Movement name	Rationale/perceived usefulness	Instructions
In-line lunge	- Provides an indication of ability to simultaneously flex and extend at the hip with flexed knees while maintaining neutral spinal alignment	As per instructions from Functional Movement Screen (Cook et al. 2006a) (see reference list for full article details). "Place your left toes so they are touching the back of the taped line. Place the heel of your right foot xx centimeters (as marked by instructor)** directly in front of your left foot. Hold the dowel behind your back gripping it with your left hand at your neck and your right hand at your lower back. Make sure the dowel is touching your head, upper back and buttocks. While maintaining an upright posture, descend into a lunge touching your left knee to the floor. Maintain contact with the dowel at the head, upper back and bum throughout. Return to the start position with knees fully extended between each repetition." Switch the words 'right' and 'left' when instructing the participant when testing the other side.
Single-leg hop for distance	- Provides an indication of lower-limb unilateral power - Highlights limb asymmetry (lower-limb power and/or ankle stability and/or lower-limb eccentric strength) - Provides an indication of single-leg control	"Place the toes of the jumping leg so they are touching the back of the taped line. Jump as far as you can while still able to stick the landing on the same leg and hold your position to allow measurement. You must record three successful scored jumps on each leg and you will receive as many attempts as necessary to achieve this."
Tuck jump	- Allows quick assessment of bilateral knee control during plyometric activity - Highlights limb asymmetry (lower-limb power and/or hip mobility)	As per instructions from Myer et al. (2008) (see reference list for full article details). "Stand on the middle of the cross taped on the floor with feet shoulder width apart. Upon signal from the tester, perform continuous vertical jumps on the spot for 10 seconds making sure to lift your knees towards your chest so that your upper thighs are parallel with the floor each time. Try to perform as many jumps as possible."

** Foot placement is determined by measuring the distance from the floor to the tibial tuberosity (shin length).

8.4 SIMS scoring criteria (chapter 3)

General rater instructions

Record each participant's height, weight and tibial tuberosity height (distance from the floor to their tibial tuberosity). If a participant cannot physically perform any test due to pain then they should be considered injured, this should be reported to the relevant club staff members and the test should be postponed.

Scoring guidelines for the anterior reach and single-leg hop for distance (objective assessments)

Anterior reach

Measure the distance (in centimeters) from the start line to the most distal part of the foot of the reaching leg. Round to the nearest centimeter. Three repetitions are performed on each leg and reach distance should be recorded for each attempt. The maximum reach distances achieved by each leg should be used to calculate the difference between left and right. The maximum theoretical score achievable is 10 and this would represent a 'poor' score. In contrast, the theoretical minimum score is zero and this would represent a 'good' score.

Difference in reach distance (cm) between legs	Test score
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
≥10	10

Single-leg hop for distance

Measure the distance (in centimeters) from the start line to the heel of the jumping/landing leg. Round to the nearest centimeter. Three repetitions are performed on each leg and jump distance should be recorded for each attempt. Both jump distance and limb symmetry are taken into account when assigning a test score. The maximum jump distance achieved on each leg should be used to calculate the score. Combine the scores for jump distance and jump symmetry to produce the final score out of 10.

Sum of right and left best jump distances (cm)		Test score
Males:	Females:	
<320	<220	5
321-340	221-240	4
341-360	241-260	3
361-380	261-280	2
381-400	281-300	1
>400	>300	0

Difference between best right and left jumps (cm)	Test score
>20	5
17-20	4
13-16	3
9-12	2
4-8	1
<4	0

<u>Scoring guidelines for the single-leg deadlift, in-line lunge and tuck jump (subjective assessments)</u>

- If an error occurs once and the rater judges it to be egregious then it should be scored as an error.
- If an error (but only to a minor extent) is observed once then it should not be scored.

- If the same error (but only to a minor extent) is observed twice then it should be scored as an error.

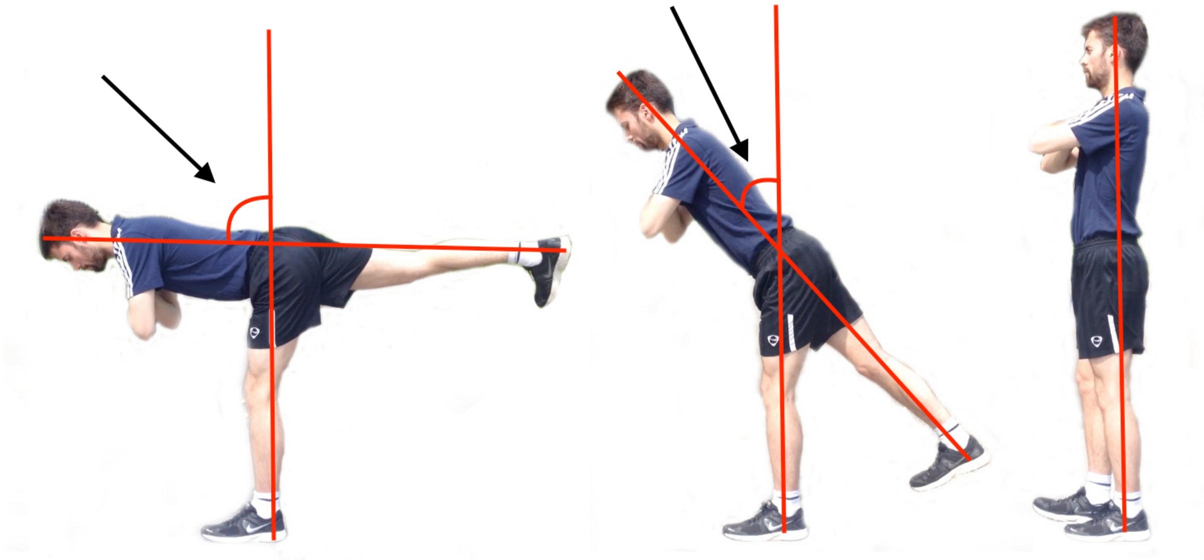
Defining specifically what constitutes "minor extent" or "egregious" is not possible. These judgments are left to the discretion of each individual rater. An important consideration is that raters are consistent in their judgments within themselves.

Single-leg deadlift

The score for this test is based on the 'movement quality' criteria outlined below. Three repetitions are performed on each leg. The maximum theoretical score achievable is 10 and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Both legs are scored and the average of both right and left scores is assigned to the individual.

Item #		
1	Is external hip rotation (standing leg) visible?	Yes=1 No=0
2	Does lumbar spine remain neutral?	Yes=0 No=1
3	Does thoracic spine remain neutral?	Yes=0 No=1
4	Does knee of raised leg remain extended throughout?	Yes=0 No=1
5	Is upper and lower body movement synchronized?	Yes=0 No=1
6	Is footprint maintained?	Yes=0 No=1
7	Is hip abduction (standing leg) present?	Yes=1 No=0
8	Does the standing leg knee remain extended throughout?	Yes=0 No=1
9	Parallel to floor position achieved?	Parallel (90°)=0, 89°-45°=1, <45°=2 (all relative to the stance leg hip flexion angle)

In relation to item #9 – the angle being assessed is displayed in the following diagram:



In-line lunge

The score for this test is based on the 'movement quality' criteria outlined below. Three repetitions are performed on each side. The maximum theoretical score achievable is eight and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Both legs are scored and the average of both right and left scores is assigned to the individual. To generate a score out of 10 multiply the fractional score out of eight by 10 e.g. if an individual displays four out of eight possible errors then the score out of 10 is: (4/8)x10 = 5. The reason for generating a score out of 10 is to maintain the same weighting between the five sub-tests.

Item #		
1	Does dowel remain vertical in frontal plane throughout?	Yes=0 No=1
2	Does torso rotation (transverse plane) occur?	Yes=1 No=0
3	Does dowel remain vertical in sagittal plane throughout?	Yes=0 No=1
4	Does back knee touch the floor?	Yes=0 No=1
5	Does heel of front foot lift off the floor?	Yes=1 No=0
6	Is footprint maintained throughout?	Yes=0 No=1

7 Are the three dowel contact points with body maintained? Yes=0 No=1

8 Does knee valgus occur during the movement? Yes=1 No=0

Tuck jump

Mark a cross on the floor using tape (two 60cm strips that intersect). The score for this test is based on the 'movement quality' criteria outlined below. The maximum theoretical score achievable is 10 and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Myer et al. (2008) created the tuck jump assessment and any further clarification on scoring procedures can be sought from their original article (see reference list for full article details).

Item #		
1	Was there knee valgus at landing?	Yes=1 No=0
2	Do thighs reach parallel (peak of jump)?	Yes=0 No=1
3	Were thighs equal side-to-side (during flight)?	Yes=0 No=1
4	Was foot placement shoulder width apart?	Yes=0 No=1
5	Was foot placement parallel (front to back)?	Yes=0 No=1
6	Was foot contact timing equal?	Yes=0 No=1
7	Was there excessive contact landing noise?	Yes=1 No=0
8	Was there a pause between jumps?	Yes=1 No=0
9	Did technique decline prior to 10 seconds?	Yes=1 No=0
10	Were landings in same footprint (within taped cross)?	Yes=0 No=1

8.5 Comparison of SIMS scores between injured and non-injured players (chapter 4)

	Injured player scores						Non-injured player scores					
	Composite	AR	SLDL	ILL	SLHD	TJ	Composite	AR	SLDL	ILL	SLHD	TJ
Primary analysis												
Lower extremity injuries (injured/non-injured n=123/183)	17.0 ± 5.0	2.2 ± 1.9	3.6 ± 1.7	2.5 ± 1.2	4.5 ± 2.4	4.1 ± 1.8	17.9 ± 4.4	2.6 ± 2.2	3.6 ± 1.8	2.7 ± 1.2	4.7 ± 2.4	4.3 ± 1.7
Secondary analysis												
Hip/groin injuries (injured/non-injured n=34/272)	17.2 ± 5.3	2.8 ± 2.5	3.5 ± 1.8	2.5 ± 1.4	4.6 ± 2.5	3.7 ± 1.7	17.6 ± 4.6	2.4 ± 2.0	3.6 ± 1.8	2.6 ± 1.2	4.6 ± 2.4	4.3 ± 1.7
Thigh injuries (injured/non-injured n=55/251)	16.6 ± 4.7	2.2 ± 1.7	3.7 ± 1.7	2.5 ± 1.3	4.3 ± 2.3	3.9 ± 1.8	17.7 ± 4.7	2.5 ± 2.2	3.6 ± 1.8	2.6 ± 1.2	4.7 ± 2.4	4.3 ± 1.7
Knee injuries (injured/non-injured n=29/277)	15.6 ± 5.6	2.1 ± 1.9	2.7 ± 1.7	2.5 ± 1.5	4.2 ± 2.3	4.0 ± 1.7	17.7 ± 4.5	2.5 ± 2.1	3.7 ± 1.7	2.6 ± 1.2	4.7 ± 2.4	4.2 ± 1.7
Ankle injuries (injured/non-injured n=37/269)	18.1 ± 5.5	2.4 ± 1.8	3.9 ± 1.6	2.6 ± 1.1	4.9 ± 2.7	4.3 ± 2.1	17.5 ± 4.6	2.5 ± 2.1	3.6 ± 1.8	2.6 ± 1.2	4.6 ± 2.3	4.2 ± 1.7
Tertiary analysis												
Hamstring muscle strains (injured/non-injured n=41/265)	16.2 ± 5.0	2.0 ± 1.6	3.4 ± 1.6	2.5 ± 1.3	4.4 ± 2.2	3.9 ± 1.7	17.7 ± 4.6	2.5 ± 2.2	3.6 ± 1.8	2.6 ± 1.2	4.6 ± 2.4	4.3 ± 1.7
Ankle sprains (injured/non-injured n=34/272)	18.1 ± 5.4	2.4 ± 1.8	4.0 ± 1.6	2.6 ± 1.1	4.9 ± 2.7	4.3 ± 2.1	17.5 ± 4.6	2.5 ± 2.1	3.6 ± 1.8	2.6 ± 1.2	4.6 ± 2.3	4.2 ± 1.7

AR, anterior reach; ILL, in-line lunge; SLDL, single-leg deadlift; SLHD, single-leg hop for distance; TJ, tuck jump

8.6 Full published texts relevant to this thesis

The following appendices represent the fully published versions of the texts relevant to this thesis and are presented in chronological order of publication date.

Sports Med
DOI 10.1007/s40279-015-0453-1

REVIEW ARTICLE

Reliability and Association with Injury of Movement Screens: A Critical Review

Robert McCunn[1] · Karen aus der Fünten[1] · Hugh H. K. Fullagar[1,2] · Ian McKeown[3] · Tim Meyer[1]

Abstract Subjective assessment of athletes' movement quality is widely used by physiotherapists and other applied practitioners within many sports. One of the beliefs driving this practice is that individuals who display 'poor' movement patterns are more likely to suffer an injury than those who do not. The aim of this review was to summarize the reliability of the movement screens currently documented within the scientific literature and explore the evidence surrounding their association with injury risk. Ten assessments with accompanying reliability data were identified through the literature search. Only two of these ten had any evidence directly related to injury risk. A number of methodological issues were present throughout the identified studies, including small sample sizes, lack of descriptive rater or participant information, ambiguous injury definitions, lack of exposure time reporting and risk of bias. These factors, combined with the paucity of research on this topic, make drawing conclusions as to the reliability and predictive ability of movement screens difficult. None of the movement screens that appear within the scientific literature currently have enough evidence to justify the tag of 'injury prediction tool'.

Electronic supplementary material The online version of this article (doi:10.1007/s40279-015-0453-1) contains supplementary material, which is available to authorized users.

✉ Robert McCunn
bob.mccunn@me.com

[1] Institute of Sport and Preventive Medicine, Saarland University, Geb. B8.2, 66123 Saarbrücken, Germany

[2] Sport and Exercise Discipline Group, UTS: Health, University of Technology, Sydney, NSW, Australia

[3] Port Adelaide Football Club, Adelaide, SA, Australia

Published online: 31 December 2015

Key Points

- Subjective assessment of athletes' movement quality is commonplace within professional sport, often in an attempt to predict injury risk.
- Of the ten movement screens identified within the scientific literature, only two have had their injury predictive ability investigated via prospective cohort studies.
- None of the movement screens present within the scientific literature currently have enough supporting evidence to justify being heralded as 'injury prediction tools'; however, they may well provide practitioners with greater holistic understanding of their athletes' physical capabilities.

1 Introduction

The use of fitness assessments to profile and categorize athletes' physical capabilities is commonplace and a central aspect of many applied practitioners' jobs [1]. The data collected from traditional fitness tests are typically objective in nature, i.e., can be measured in units such as seconds, centimeters, or grams. Movement screening is a type of assessment frequently used within professional soccer as well as other sports and is predominantly a subjective process that aims to measure the 'quality' of a movement pattern [2]. However, for various reasons, including the subjective nature of such assessments and its relatively recent adoption by practitioners, this practice has received limited attention within the scientific literature. A consensus on what defines movement quality is not available; however, the concept encapsulates the maintenance of correct posture and joint alignment in addition to balance while performing the selected movements. While some sporting institutions measure intuitively related parameters such as strength and joint range of motion, movement quality has been identified as an independent attribute [3, 4]. Therefore, a fitness testing battery that seeks to build a comprehensive profile of an athlete should incorporate an assessment of movement quality. This highlights the need for reliable and valid movement screening tools.

The foundation of a comprehensive injury prevention program is identifying individuals with a high risk of injury [5], and this is one of the key concepts underpinning the practice of movement screening. If athletes who display 'poor' movement patterns have a greater risk of injury than those who display 'good' movement patterns, then screening protocols may be an important component of injury prevention strategies. However, the purpose of movement screens is not to diagnose why a poor movement pattern exists but simply to highlight it [6]. It is up to the judgment of the practitioner as to the course of action, if any, taken in response to the outcome. Furthermore, elite sport is not the only environment in which movement ability is important. Giblin et al. [7] stated that fundamental movement ability (core stability, balance, coordination) is related to perceived competence and confidence associated with physical activity. As such, movement quality is linked to general health as well as sports performance.

Despite movement quality being an important skill for the general population, in addition to athletes, measuring it is problematic due to its subjective nature. A variety of movement screens exist, the most well-known being the functional movement screen (FMS) [6, 8]. The FMS has received attention from researchers, and different aspects of this protocol—such as its reliability and association with injury—have been investigated. However, other screens do exist, with some—but not all—appearing within the scientific literature. No collective critique of the movement screens detailed in the scientific literature, necessary to raise awareness of the available options, currently exists. This would allow practitioners to make informed decisions about which, if any, movement screen is most appropriate for them. Accordingly, the aim of the present review was to summarize the intra- and inter-rater reliability of the available movement screens and discuss the evidence surrounding their ability to determine injury risk.

2 Literature Search

We performed a computerized literature search (Fig. 1) in PubMed, Web of Science, and ScienceDirect for articles published up until 1 July 2015 using the search terms 'movement', 'screen', 'screening', 'reliability', 'injury', 'prediction', 'predicts', 'landing error scoring system', 'tuck jump assessment', 'functional movement screen', 'functional movement screening', 'single leg squat test', 'squat', 'test', 'drop jump', 'drop vertical jump', and 'movement quality' in various combinations. In addition, articles were identified manually from the reference lists of original manuscripts; a total of 51 relevant articles were identified. For the purpose of this review, a movement screen was defined as a protocol designed for use with apparently healthy, uninjured individuals to primarily assess the 'quality' of a movement(s) rather than objective outcomes such as number of repetitions, distance, or time achieved. The movement(s) included should rely on multiple physical qualities to execute correctly, e.g., strength, balance, and flexibility. It is not used to identify specific clinical conditions and does not require interpretation by a medical professional.

3 Reliability of Identified Screens

Ten movement screens that met the definition outlined above and with accompanying reliability data were identified through the literature search (Fig. 2). These screens consisted of the FMS, the Landing Error Scoring System (LESS), single-leg squat screen variations, drop vertical jump screen variations, tuck jump assessment, athletic ability assessment (AAA), conditioning specific movement tasks (CSMT), the netball movement screening tool (NMST), the physical performance measures screen (16-PPM), and the star excursion balance test (SEBT) movement quality screen. A description of the exercises involved in each screen is provided in Table 1.

The reliability of an assessment tool is paramount because it is a pre-requisite for test validity [9]. As such, before any given movement screen can be investigated with respect to injury prediction, it must first be demonstrated that the test is reliable. Throughout the 51 articles identified by the literature search, intra-class correlation coefficients (ICC) were commonly reported. Atkinson and Nevill [10] stated that various qualitative interpretations of ICC values exist, yet none were related to "analytical goals for research" and so it is difficult to say exactly what value constitutes 'good' or 'excellent' reliability. Some of the identified studies classified an ICC value of ≥0.75 as good [11], whereas others [12, 13] classified scores of ≥0.80 and

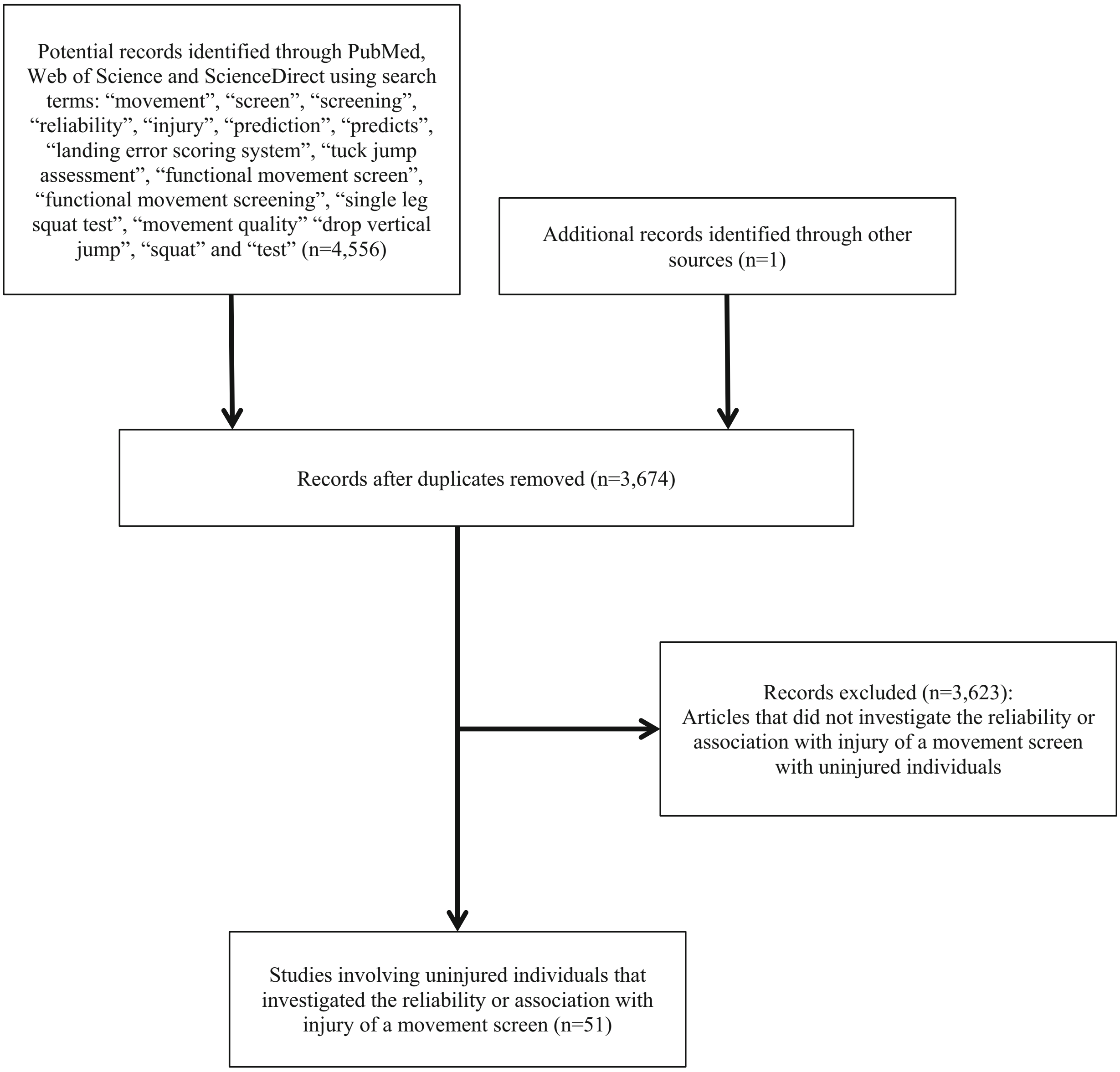


Fig. 1 Flow diagram showing the identification and selection of movement screen studies in the scientific literature for the current review

≥0.90 as good and excellent, respectively. Shultz et al. [14] reported ICC values of 0.40–0.75 as fair to good and >0.75 as excellent. A reasonable consensus as to what can be considered good reliability appears to be an ICC ≥0.75; thus, this classification is used throughout this review. In addition to ICCs, kappa values were also often reported. Guidelines presented by Landis and Koch [15] are used to classify these scores, with kappa score indicating strength of agreement as follows: poor <0.00, 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.

3.1 Methodological Quality Assessment for Reliability Studies

The methodological quality of each paper that reported the reliability of a movement screen was assessed using the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist [16]. This checklist utilizes a four-point scoring system (poor, fair, good, excellent) and contains sub-sections relating to numerous aspects of study design. For the purposes of this critique, only the reliability section (box B), which contains

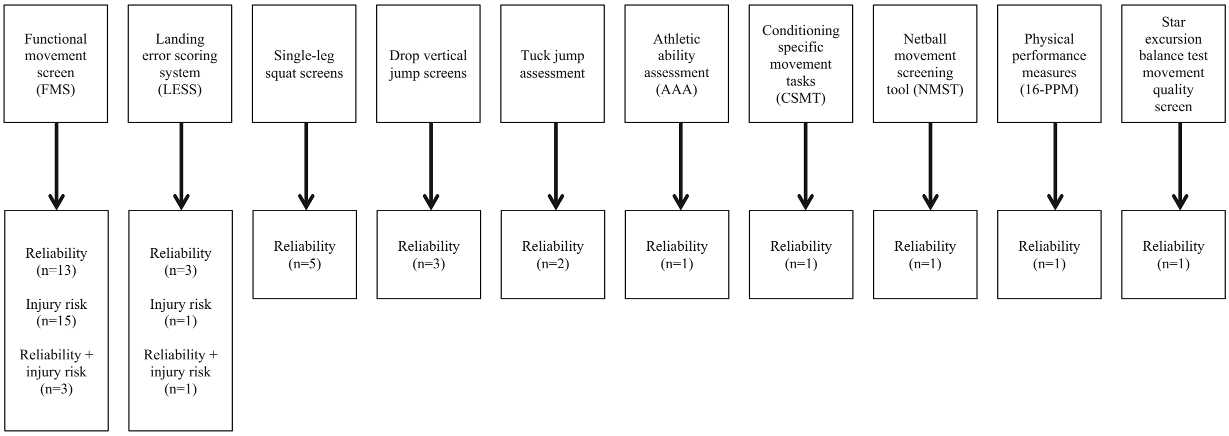


Fig. 2 Breakdown of the individual movement screens identified from the literature search and the number of the articles that investigated their reliability or association with injury

14 questions, was addressed. As per the checklist instructions, if the answer to any of the 14 questions was 'poor' then, based on the 'worst score counts' principle, the study was classified as such. The outcomes of this methodological quality assessment are presented in Table 2 (see the Electronic Supplementary Material [ESM] Appendix S1 for raw data).

3.2 Functional Movement Screen

The FMS comprises seven subtests, including an overhead squat, hurdle step, in-line lunge, shoulder mobility assessment, active straight-leg raise, trunk stability push-up, and a prone two-point rotary stability movement, all performed without external load [6, 8]. As can be seen from Table 3, a total of 16 studies were identified that reported either the intra- or inter-rater reliability of the FMS or related variations [11–14, 17–28]. Seven of the eight studies that investigated the intra-rater reliability reported ICC values $\geq$0.75. In addition, Teyhen et al. [24] reported an ICC of 0.74 (95 % confidence interval [CI] 0.60–0.83). As such, it would appear that the FMS has consistently demonstrated good intra-rater reliability. Of the 16 identified studies, 14 reported inter-rater reliability either in the form of ICC, weighted kappa, or Krippendorff's α. Twelve of these observed an ICC $\geq$0.75, representing good inter-rater reliability. In addition, Minick et al. [28] used the weighted kappa statistic to measure inter-rater reliability; however, rather than use the FMS composite score, they compared raters using the individual subtests. The weighted kappa values ranged from 0.79 to 1.0, with the authors stating that this represented substantial to excellent agreement. In contrast, Shultz et al. [14] used Krippendorff's α statistic to quantify reliability and classified a value $\geq$0.80 as acceptable; however, they reported a score of 0.38 (95 % CI 0.35–0.41). These authors concluded that the FMS demonstrated poor inter-rater reliability and suggested that improved rater training may have resulted in an improved reliability score. The authors also highlighted the difference between years of experience and the number of tests a rater has administered, stating that the latter is likely of greater relevance to improving reliability.

Five studies [11, 18, 20, 21, 29] included information as to the raters' experience (years of clinical practice, number of FMS tests performed, or level of certification), which allowed comparison of test reliability based on these variables. Four studies [11, 18, 20, 29] suggested the experience of the rater did not influence the inter-rater reliability. Additionally, Smith et al. [11] observed good intra-rater reliability (ICCs >0.80) for all raters, regardless of experience. In contrast, Gribble et al. [21] showed that intra-rater reliability did vary depending on the experience of raters. The raters were divided into three groups: athletic training students, athletic trainers who had not previously used the FMS, and athletic trainers who had at least 1 year of experience administering FMS tests. The greater number of raters included in the study by Gribble et al. [21] suggests a stronger experimental foundation, and this may have contributed to their contrasting findings. A clear trend highlighting the importance of rater experience was apparent, with the following ICC values: students, 0.37 (95 % CI −0.79 to 0.78), athletic trainers, 0.76 (95 % CI 0.32–0.92), and experienced athletic trainers, 0.95 (95 % CI 0.68–0.99). However, while Gribble et al. [21] employed a greater number of raters than the other four studies, the ICC values reported were based on the ratings of only three participants. Therefore, the weight of the currently available evidence suggests that the experience of the rater is not a significant factor influencing scoring.

Table 1 Content of the identified movement screens

Screen name	Exercises (*N*)	Name of exercises	Protocol description
Functional movement screen	7	Deep squat	Cook et al. [8]
		Hurdle step	Cook et al. [6]
		In-line lunge	
		Shoulder mobility	
		Active straight-leg raise	
		Trunk stability push up	
		Rotary stability	
Landing error scoring system	1	Drop jump	Padua et al. [34]
Single-leg squat screens			
Single-leg squat task	1	Single-leg half squat	Crossley et al. [36]
Single-leg mini squat	1	Single-leg half squat	Ageberg et al. [38]
Unilateral lower extremity functional tasks	2	Single-leg half squat	Chmielewski et al. [39]
		Lateral step down	
Drop vertical jump screens	1	Drop vertical jump	Nilstad et al. [44]
			Whatman et al. [45]
			Ekegren et al. [46]
Tuck jump assessment	1	Tuck jump	Myer et al. [48]
Athletic ability assessment	9	Prone hold on hands	McKeown et al. [53]
		Lateral hold on hands	
		Overhead squat	
		Single-leg squat off box	
		Walking lunge	
		Single-leg forward hop	
		Lateral bound	
		Push ups	
		Chin ups	
Conditioning specific movement tasks	6	Overhead squat	Parsonage et al. [54]
		Romanian deadlift	
		Single-leg squat	
		Double-leg to single-leg landing	
		Sprint (40 m)	
		Countermovement jump	
Netball movement screening tool	10	Squat	Reid et al. [55]
		Lunge and twist	
		Bend and pull	
		Push up	
		Single-leg squat	
		Vertical jump (land on both legs)	
		Vertical jump (land on one leg)	
		Broad jump	
		Star excursion balance test	
		Active straight leg raise	

Table 1 continued

Screen name	Exercises (*N*)	Name of exercises	Protocol description
Physical performance measures	16	Broad jump	Tarara et al. [56]
		Closed kinetic chain upper extremity stability test	
		Y-balance test	
		In-line lunge for distance	
		Lateral lunge for distance	
		Lumbar endurance	
		Side plank hip abduction	
		Side plank hip adduction	
		Triple hop for distance	
		Nordic hamstring	
		Full squat	
		Downward dog	
		Single-leg squat	
		Shoulder mobility test	
		Active straight leg raise	
		Beighton hypermobility	
Star excursion balance test movement quality screen	1	Anterior reach	Ness et al. (2015) [58]

Another aspect of movement screening pertinent to the reliability of such assessments is whether the rater scores participants in real time or after the test via video recording. One study sought to address this issue [14] and found that the intra-rater reliability was superior when using recorded footage to score participants. One rater assessed individuals live while they were being filmed and retrospectively assessed the footage. They also assessed the participants again 1 week later and scored the tests in real time. The ICCs for the live–live re-test and live–recorded re-test were 0.60 (95 % CI 0.35–0.77) and 0.92 (95 % CI 0.85–0.96), respectively. However, only one rater's scoring was investigated in this manner, hence it is difficult to draw firm conclusions on the respective merits of live and recorded FMS scoring. Nonetheless, the outcome of this study suggests that using recorded footage to assess participants may elicit greater intra- and inter-rater reliability than assessing them in real time.

Recently, it has been shown that participant knowledge of the scoring criteria can influence FMS total score [29]. Participants were assessed prior to and after having the criteria for a perfect score explained to them. Significant improvements in scores were observed simply by providing this information. This finding demonstrates that test reliability may be affected by how the test is delivered by the assessor. If the extent of task instruction and explanation differs between assessors or test occasions it is likely that intra- and inter-rater reliability will be impacted. To ensure that any changes in score are not simply due to familiarization, it is recommended that participants have the scoring criteria clearly explained to them and are allowed practice attempts before being scored. This is not to say that individuals should be coached through the movements, rather, it is imperative they know what is being asked of them without being told how to do it.

Overall, the majority of the identified studies reveal the FMS possesses good intra- and inter-rater reliability, although it should be noted that this conclusion is not unanimous throughout the literature. It should also be noted that the majority of the identified studies were classified as demonstrating poor methodological quality (Table 2). The influence of rater experience on reliability appears negligible. Furthermore, the practice of scoring tests via video footage may aid reliability, yet evidence on this issue is limited, so only tentative conclusions can be drawn. Test reliability is likely influenced by the performer's knowledge of the scoring criteria; as such, it is advisable to provide clear instructions to participants and allow practice attempts to reduce the influence of any learning effect. The depth of research investigating the FMS is much greater than for any other movement screen, yet despite this some organizations choose to use alternative tools [2, 30]. While the specific reasons are not clear, many professional football clubs seemingly do not feel the FMS meets their screening needs. However, a number of other screens

Table 2 Methodological quality (according to COSMIN checklist) of reliability studies[a]

Movement screen	Methodological quality rating		
	Poor	Fair	Good
FMS	Teyhen et al. intra-rater [24]	Shultz et al. [14]	Teyhen et al. inter-rater [24]
	Hotta et al. [17]	Butler et al. [25]	
	Gulgin and Hoogenboom [18]	Minick et al. [28]	
	Letafatkar et al. [19]		
	Parenteau-G et al. [12]		
	Elias [20]		
	Gribble et al. [21]		
	Smith et al. [11]		
	Frohm et al. [22]		
	Klusemann et al. [23]		
	Onate et al. [13]		
	Schneiders et al. [26]		
	Chorba et al. [27]		
LESS	Smith et al. [31]	Padua et al. [32]	Padua et al. [34]
	Onate et al. [33]		
Single-leg squat screens	Örtqvist et al. inter-rater [37]	Örtqvist et al. intra-rater [37]	Junge et al. [35]
	Crossley et al. [36]		
	Ageberg et al. [38]		
	Chmielewski et al. [39]		
DVJ screens	Whatman et al. [45]	Nilstad et al. [44]	
	Ekegren et al. [46]		
Tuck jump assessment	Herrington et al. [51]	Dudley et al. [52]	
AAA	McKeown et al. [53]		
CSMT		Parsonage et al. [54]	
NMST	Reid et al. [55]		
16-PPM	Tarara et al. [56]		
SEBT movement quality screen		Ness et al. [58]	

AAA athletic ability assessment, *COSMIN* consensus-based standards for the selection of health measurement instruments, *CSMT* conditioning specific movement tasks, *DVJ* drop vertical jump, *FMS* functional movement screen, *LESS* Landing Error Scoring System, *NMST* Netball Movement Screening Tool, *PPM* physical performance measures, *SEBT* Star Excursion Balance Test

[a] A quality rating of 'Excellent' is possible but was not achieved by any of the listed studies

appear within the scientific literature, albeit with much less supporting evidence.

3.3 Landing Error Scoring System

The LESS comprises one movement: drop vertical jumps. The accompanying scoring criteria relate to observed errors in technique and result in a potential minimum and maximum score of zero or 19, respectively, with a higher score indicating poorer performance. The four studies reporting either the intra- or inter-rater reliability of the LESS are presented in Table 4 [31–34]. Both Smith et al. [31] and Padua et al. [34] reported excellent intra-rater reliability (ICCs >0.90). When interpreting these findings, it should be noted that, in total, data from only three raters were used to calculate these ICC values. Further research is needed to establish the robustness of such findings. All four of the identified studies measured the inter-rater reliability of the LESS, with the ICC values ranging from 0.72 to 0.92, indicating good repeatability. Again, caution should be employed when analyzing these results, since this conclusion is based on the data from only nine raters. One study investigated the influence of rater experience on LESS scoring and found that novice (<1 year of experience as a certified athletic trainer) and expert (15 years' experience as a certified athletic trainer) raters displayed moderate to perfect agreement on all items [33]. The detailed scoring criteria employed by the LESS likely explain this high level of agreement between raters. The drawback to such a thorough scoring system is inevitably the time it takes to

Table 3 Studies that reported the intra- and/or inter-rater reliability of the functional movement screen

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/ sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Hotta et al. [17]	10 M	Middle- and long-distance runners (collegiate)	2 Physical therapists	NA	NA	2	ICC 0.98 (0.93–1.00)
Gulgin and Hoogenboom [18]	10 M, 10 F	University students	1 Expert rater, 3 physical therapy students	NA	NA	4	ICC 0.88 (0.77–0.95)
Letafatkar et al. [19]	20	Unknown	2 Physical therapists	NA	NA	2	ICC 0.92
Parenteau-G et al. [12]	28 M	Ice-hockey players (elite youth)	1 Physiotherapist, 3 physiotherapy students	2	Rater 1 ICC 0.96 (0.92–0.98) Rater 2 ICC 0.96 (0.92–0.98)	2	ICC 0.96 (0.92–0.98)
Elias [20]	3 M, 2 F	Squash players (elite)	20 Physiotherapists	NA	NA	20	ICC 0.91
Gribble et al. [21]	2 M, 1 F	University students	16 Students, 15 ATs, 7 expATs	38	All raters ICC 0.75 (0.53–0.87) ExpATs ICC 0.95 (0.68–0.99) ATs ICC 0.76 (0.32–0.92) Student ICC 0.37 (–0.79 to 0.78)	NA	NA
Shultz et al. [14]	18 M, 21 F	NCAA Division 1 varsity athletes	1 Student, 1 physical therapist, 2 ATs, 2 S&C coaches	1	Live test–retest ICC 0.60 (0.35–0.77) Live-recorded test–retest ICC 0.92 (0.85–0.96)	6	Krippendorff's $\alpha = 0.38$ (0.35–0.41)
Smith et al. [11]	10 M, 10 F	University students	2 Students, 1 faculty member, 1 FMS certified instructor	4	Rater 1 ICC 0.90 (0.76–0.96) Rater 2 ICC 0.81 (0.57–0.92) Rater 3 ICC 0.91 (0.78–0.96) Rater 4 ICC 0.88 (0.72–0.95)	4	Occasion 1 ICC 0.89 (0.80–0.95) Occasion 2 ICC 0.87 (0.76–0.94)

Table 3 continued

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/ sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Frohm et al. [22]	26 M	Soccer players (elite)	8 Physiotherapists	8	Rater 1 ICC 0.87 Rater 2 ICC 0.77 Rater 3 ICC 0.83 Rater 4 ICC 0.77 Rater 5 ICC 0.79 Rater 6 ICC 0.45 Rater 7 ICC 0.79 Rater 8 ICC 0.75	8	Occasion 1 ICC 0.80 Occasion 2 ICC 0.81
Klusemann et al. [23]	10	Basketball players (elite youth)	8 (unspecified combination of S&C coaches/physiotherapists)	8	ICC 0.82	NA	NA
Onate et al. [13]	12 M, 7 F	University students	1 AT, 1 S&C coach	1	ICC 0.92	2	ICC 0.98
Teyhen et al. [24]	53 M, 11 F	Military personnel	8 Physical therapy students	4	ICC 0.74 (0.60–0.83)	8	ICC 0.76 (0.63–0.85)
Butler et al. [25]	30	Middle school students	1 FMS creator, 1 FMS certified instructor	NA	NA	2	ICC 0.99
Schneiders et al. [26]	10	Recreationally active individuals	2 Academic researchers	NA	NA	2	ICC 0.97
Chorba et al. [27]	3 M, 5 F	University students	2 Physical therapists	NA	NA	2	ICC 0.98
Minick et al. [28]	17 M, 23 F	University students	2 FMS creators, 2 FMS certified instructors	NA	NA	4	Weighted κ values for each test ranged from 0.79 to 1.0 when comparing novice and experienced raters

All values refer to the FMS composite score unless otherwise stated

AT athletic trainer, *CI* confidence interval, *ExpAT* experienced athletic trainer, F female, *FMS* functional movement screen, *ICC* intraclass correlation coefficient, κ kappa, *M* male, *NA* not applicable, *NCAA* National Collegiate Athletic Association, *S&C* strength and conditioning

score each participant. The original LESS protocol requires video recording of tests, with subsequent scoring by assessors from the footage, and this methodology has associated costs, from both a financial and a time perspective. A real-time scoring system to overcome these restrictive issues was developed by Padua et al. [32]. Three raters' real-time scoring of 43 participants was compared, and the resulting ICC values ranged from 0.72 (95 % CI 0.42–0.88) to 0.81 (95 % CI 0.56–0.92), suggesting moderate to good inter-rater reliability for the real-time version. Taken collectively, the initial evidence is promising with regard to the reliability of the test.

3.4 Single-Leg Squat Screens

Five studies were identified that explored either the intra- or inter-rater reliability of movement screens containing various single-leg squat tests (Table 5) [35–39]. The single-leg squat task [36], the single-leg mini squat [35, 37, 38], and another assessment comprising two 'functional tasks'—the single-leg squat and lateral step-down—have been investigated [39]. While these three screens differ slightly in their protocols, the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during single leg

Table 4 Studies that reported the intra- and/or inter-rater reliability of the Landing Error Scoring System

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score (95 % CI)
Smith et al. [31]	10	High school/collegiate athletes	2 Raters (occupation unspecified)	2	ICC 0.97	2	ICC 0.92
Padua et al. [32]	19 M, 24 F	Military personnel	3 Athletic trainers	NA	NA	3	Rater 1 vs. 2 ICC 0.81 (0.56–0.92) Rater 1 vs. 3 ICC 0.72 (0.42–0.88) Rater 1 vs. combined 2 and 3 ICC 0.79 (0.64–0.88)
Onate et al. [33]	19 F	NCAA Division 1 soccer players	2 Athletic trainers	NA	NA	2	ICC 0.84
Padua et al. [34]	25 M, 25 F	Military personnel	2 Raters (occupation unspecified)	1	ICC 0.91	2	ICC 0.84

All values refer to the LESS composite score

CI confidence interval, *F* female, *ICC* intraclass correlation coefficient, *LESS* Landing Error Scoring System, *M* male, *NA* not applicable, *NCAA* National Collegiate Athletic Association

squatting. Three of the five identified studies investigated the intra-rater reliability, with kappa values ranging from 0.13 to 0.80, representing poor to substantial agreement. These results are difficult to interpret, since the kappa values within each individual study varied so widely. Differences in the populations observed between studies may explain some of the variance; however, comprehensive participant information was not reported in all instances. Similarly, differences in the precise protocols and scoring criteria for each screen variation may have contributed to the inconsistent results. These values were derived from the data collected by only seven raters in total, and this relatively small evidence source likely contributes to the uncertain findings. Similarly, as can be seen from Table 5, the inter-rater kappa values ranged from 0.00 to 0.92, making conclusions difficult to draw with regard to the reliability of these screening tools.

3.5 Drop Vertical Jump Screens

The drop vertical jump, whereby an individual drops from a raised surface to the floor and immediately jumps vertically as high as possible, is a common screening test performed to identify movement patterns thought to be associated with risk of injury [40–42]. However, quantification of performance on such tests is typically achieved via objective analysis of joint angles and 'separation distance' between the knees and ankles [40–43]. This type of assessment was considered separate to the primarily subjective process of movement screening discussed in this review. However, the literature search identified three studies that described the intra- and inter-rater reliability of drop vertical jump screen variations that conformed to the definition of a movement screen previously outlined here (Table 6) [44–46]. Of the three studies, one used the first-order agreement coefficient (AC1) statistic to analyze reliability [45]. The AC1 values can be interpreted in the same way as described above for kappa values [47]. As with the identified single-leg squat screens, the drop vertical jump screens differ in their protocols; however, the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during landing. Intra-rater reliability ranged from moderate to almost perfect; however, much greater variation existed between raters. Across the three studies, AC1 and kappa values ranged from 0.32 to 0.92, representing fair to almost perfect agreement. The poorest intra- and inter-rater reliability was reported by Whatman et al. [45], suggesting that perhaps the protocols and scoring criteria adopted by the other two studies are superior [44, 46]. The results from the identified studies are mixed; therefore, further research utilizing consistent test protocols and scoring criteria are required to elucidate the reliability of these screening tools.

3.6 Tuck Jump Assessment

The tuck jump assessment created by Myer et al. [48] was designed to assess the movement quality associated with repeated jumping and landing and requires an individual to perform repeated tuck jumps for 10 s. The plyometric aspect of this exercise task is relevant, since it has been reported that injury prevention interventions lacking this explosive component have demonstrated limited success in

Table 5 Studies that reported the intra- and/or inter-rater reliability of the identified single-leg squat screens

References	Participation information		Rater information	Intra-rater reliability		Inter-reliability	
	Sample size	Occupation/ sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Junge et al. [35]	72	Students (children)	2 Physiotherapy students	NA	NA	2	Weighted κ values for each scoring category ranged from 0.54 to 0.86
Crossley et al. [36]	15	Unknown (adult)	3 Physical therapists, 1 'expert panel'	3	Rater 1 κ 0.80 Rater 2 κ 0.70 Rater 3 κ 0.60	4	κ values ranged from 0.60 to 0.80
Örtqvist et al. [37]	33	Students (children)	2 Physiotherapists	1	*K* 0.48 (0.16–0.79)	2	κ 0.57 (0.30–0.85)
Ageberg et al. [38]	8 M, 17 F	Unknown (adult)	2 Physical therapists	NA	NA	2	κ 0.92 (0.75–1.08)
Chmielewski et al. [39]	7 M, 18 F	Unknown (adult)	2 Physical therapists, 1 athletic trainer	3	Weighted κ values for each test and each scoring method ranged from 0.13 to 0.68	3	Weighted κ values for each test and each scoring method ranged from 0.00 to 0.55

CI confidence interval, *F* female, κ kappa, *M* male, *NA* not applicable

Table 6 Studies that reported the intra- and/or inter-rater reliability of the identified drop vertical jump screens

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Nilstad et al. [44]	60 F	Soccer players (elite)	3 Physiotherapists	NA	NA	3	κ values ranged from 0.52–0.92
Whatman et al. [45]	12 M, 11 F	Variety of undisclosed sports (youth)	66 Physiotherapists	26	All raters AC1 0.60 (range 0.14–0.92) Raters >14 years' experience AC1 0.65 (range 0.22–0.91) Raters <10 years' experience AC1 0.56 (range 0.20–0.83)	66	All raters AC1 0.34 (95 % CI 0.22–0.47) Raters >14 years' experience AC1 0.36 (95 % CI 0.22–0.50) Raters 10–14 years' experience AC1 0.37 (95 % CI 0.21–0.53) Raters 5–9 years' experience AC1 0.33 (95 % CI 0.33–0.55) Raters <5 years' experience AC1 0.32 (95 % CI 0.19–0.46)
Ekegren et al. [46]	40 F	Soccer players (regional level youth)	3 Physiotherapists	3	Rater 1 κ 0.80 (95 % CI 0.65–1.00) Rater 2 κ 0.85 (95 % CI 0.72–1.00) Rater 3 κ 0.75 (95 % CI 0.58–1.00)	3	Time point 1 κ 0.80 (95 % CI 0.62–0.98) Time point 2 κ 0.77 (95 % CI 0.59–0.95)

AC1 first-order agreement coefficient, *CI* confidence interval, *F* female, κ kappa, *M* male, *NA* not applicable

reducing knee injury [48–50]. Two studies were identified that established the intra- and inter-rater reliability of the tuck jump assessment [51, 52]. Herrington et al. [51] revealed good intra-rater reliability for two raters, with kappa values ranging from 0.81 to 1.00. Similarly, they reported good inter-rater reliability, with a kappa value of 0.88. In contrast, Dudley et al. [52] reported poor to moderate intra- and inter-rater reliability, with ICC values ranging from 0.44 to 0.72. One possible explanation for the discrepancy between these findings—at least with regard to intra-rater reliability—may relate to the differences in sample sizes. Dudley et al. [52] viewed videos of 40 participants, whereas Herrington et al. [51] only assessed ten subjects. This may have resulted in recall bias, with the raters investigated by Herrington et al. [51] potentially remembering the previous scores of the ten participants when scoring their videos for the second time. However, since one of the creators of the tuck jump assessment was a co-author and rater within the Herrington et al. [51] article, more extensive training and experience could also have contributed to the superior reliability values. The tuck jump assessment is unique amongst movement screens in that it requires the participants to perform repeated plyometric movements, with the creators proposing that the increased sport-specificity of the task may aid in highlighting injury risk. However, this assessment currently only demonstrates face validity and this should be remembered when taking this assertion into consideration. The nature of this assessment means that it may be of particular interest to practitioners working in jumping and landing sports such as netball and basketball.

3.7 Athletic Ability Assessment

The recently developed AAA, which consists of nine subtests, is currently used within numerous high-performance environments (unpublished observation) to assess athletes' movement patterns [53]. The nine subtests include a prone hold, lateral hold, overhead squat (with 10-kg bar), single-leg squat off a box, walking lunge (with a 20-kg bar), single-leg forward hop, lateral bound, push-up, and chin-up. McKeown et al. [53] reported excellent intra- and inter-rater reliability [ICC values of 0.97 (90 % CI 0.92–0.99) and 0.96 (90 % CI 0.94–0.98), respectively]. These authors also observed a strong correlation ($r = 0.94$) between athletes' overall AAA scores when assessed live and via video recording, indicating that either method is viable. The AAA maximum score is 117, and a detailed scoring system that stipulates criteria based on body segment is provided, allowing for more precise assessment than some other movement screens since the continuum of possible scores is large.

3.8 Conditioning Specific Movement Tasks

The CSMT screen was developed to aid in the assessment of young rugby union players' readiness to enter elite academies [54]. The CSMT screen comprises six subtests, including an overhead squat (with a 20-kg bar), Romanian deadlift (with a 20-kg bar), single-leg squat, double-leg to single-leg landing, a 40-m sprint, and countermovement jump. A four-point scoring system, similar to that employed by the FMS, was used to rate the quality of the six movements. Intra-rater kappa values ranged from 0.61 to 1.00, indicating substantial to excellent agreement [54]. Similarly, inter-rater kappa values ranged from 0.62 to 1.00. The reliability values are based on the scores given by only two raters, which should be remembered when interpreting the findings. Further investigations utilizing adult populations, greater number of raters, athletes from different sports, and raters of varying experience are required before definitive conclusions can be made with regard to the reliability of the CSMT screen.

3.9 Netball Movement Screening Tool

Another sport-specific screen, the NMST, was designed to assess movement quality in patterns relevant to the sport of netball. The NMST comprises ten subtests, including a squat, lunge with a twist, a bend and pull movement, push-up, single-leg squat, vertical jump (landing on both legs), vertical jump (landing on one leg), broad jump, the SEBT, and an active straight-leg raise. Reid et al. [55] reported intra- and inter-rater ICC values of 0.96 (95 % CI 0.91–0.98) and 0.84 (95 % CI 0.65–0.93), respectively. These values suggest excellent agreement within and between raters; however, the results were based on the scores given by only two examiners. The age of the netball players assessed by Reid et al. [55] ranged between 13 and 17 years, so further reliability studies conducted with adult players are needed to establish the applicability of the results to this population.

3.10 Physical Performance Measures

The 16-PPM is made up of 16 subtests, ten of which are quantitative in nature, e.g., measured in distance or number of repetitions completed [56]. While these ten subtests do not meet the aforementioned definition of movement screening exercises, the 16-PPM also includes six qualitative subtests that do assess how well an athlete performs the required movement. The six qualitatively scored subtests, which are all performed without external load, include an overhead squat, downward dog, single-leg squat, shoulder mobility assessment, active straight-leg raise, and Beighton hypermobility assessment. The following reliability values refer only to the six qualitative subtests. Intra-rater reliability, reported as weighted kappa values, varied between expert and novice raters according to Tarara et al. [56]. Weighted kappa values ranged from 0.32 to 0.81 for the expert rater, representing fair to almost perfect agreement. In contrast, the two novice raters' weighted kappa values ranged from −0.09 to 0.78, indicating poor to substantial agreement between test occasions. As such, it would appear that training is required for raters administering the 16-PPM to ensure consistent scoring. Little information was given as to the occupation or level of qualification of the expert rater, so it is unclear how much training may be required to achieve an acceptable level of consistency. Inter-rater reliability varied widely, with weighted kappa values ranging between 0.24 and 0.93 for individual subtests, representing fair to almost perfect agreement. Taking all the qualitative subtests into account, the 16-PPM appears to be a moderately reliable tool for assessing movement competency if administered by expert raters.

3.11 Star Excursion Balance Test Movement Quality Screen

The SEBT involves the objective measurement of unilateral reach distance of the lower extremity in various directions [57]. One article was identified that applied subjective movement quality criteria to the SEBT [58]. In its original form, the SEBT does not take into account how somebody achieves their score and reports only the objective reach distance in centimeters. Incorporating an assessment of an individual's movement quality during this test may provide additional useful information to practitioners. In the identified study, scoring criteria related to knee, pelvis, and trunk position were used by three physical therapists to score 100 university students [58]. Intra-rater reliability was not assessed, while inter-rater kappa values ranged from 0.18 to 0.60, representing slight to moderate agreement. As information related to within-rater variation was lacking, no judgment can currently be made as to the usefulness of the movement quality version of the SEBT.

4 Injury-Prediction Ability of Movement Screens

Studies that employed a prospective cohort or case–control design and investigated the association between outcome score and injury were identified for two movement screens: the FMS and LESS. Movement screening is widely used by elite sporting organizations in an attempt to detect injury risk [2]. Given this, it is important that the efficacy of movement screens in achieving this goal is understood. That only two of the ten identified screens have any supporting evidence as to their association with injury risk demonstrates that much work is needed to support this practice.

4.1 Methodological Quality Assessment for Injury-Prediction Studies

The methodological quality of each paper that investigated the ability of a movement screen to predict injury was assessed using a previously validated checklist for retrospective and prospective studies [59]. Specifically, an amended version was used as described by McCall et al. [60], since not all of the questions included in the full checklist were relevant for cohort studies. The questions excluded were only appropriate for intervention studies. For the purposes of this review, the questions included were 1, 2, 3, 5, 6, 7, 10, 11, 12, 18, 20, 21, 22, and 25 as previously used [60, 61]. Following the protocol outlined by McCall et al. [60], a percentage score was awarded for each article (see ESM Appendix S2 for raw data). A 'level of evidence' was then awarded based on the procedure outlined by the Scottish Intercollegiate Guidelines Network (SIGN) [62]. Scientific levels of evidence range from one to four according to the type of study. For example, cohort and case–control studies are level two. Levels one and two can score an additional mark of '++', '+', and '−' dependent on the judged quality and risk of bias. Percentage cut-off scores were used to determine if a paper was either of high quality with very low risk of bias (≥75 %), well conducted with low risk of bias (50–74 %), or low quality with high risk of bias (<50 %) [60]. A graded recommendation following the SIGN guidelines was given for each of the two movement screens that have had their injury predictive value investigated. The assignment of the graded recommendation was based on the levels of evidence of the relevant studies and the considered subjective judgment of the present authors. Graded recommendations were as follows: A: strong recommendation, B: moderate recommendation, C: weak recommendation, or D: insufficient evidence to make a specific recommendation [60].

4.2 Functional Movement Screen

A total of 18 articles were included that investigated the link between FMS score and injury risk (Table 7) [17, 19, 27, 63–77]. Ten [19, 27, 64, 70–74, 76, 77] of the 18 studies reported an association between the FMS composite score and injury. It should be noted that one of these studies appears to have reported an incorrect odds ratio (OR) based on the data presented, and the conclusions should be interpreted with caution [19]. Kiesel et al. [77] were the first to investigate the link between FMS score and injury; they followed 46 American Football players over the course of a pre-season (4.5 months). All players completed the FMS at the start of pre-season, and any subsequent injuries that met the defined criteria were recorded. These authors found that the greatest specificity and sensitivity were obtained when a cut-off score of 14 was used. Specificity and sensitivity are measures of the true-negative and true-positive rate, respectively [78]. In the case of this study, the specificity value displayed the proportion of non-injured athletes with a score >14, while the sensitivity value displayed the proportion of injured athletes with a score ≤14. The closer both measures are to a value of 1, the more robust the tool as a predictive instrument. An OR of 11.67 for those scoring ≤14 compared with those scoring >14 was reported by Kiesel et al. [77], and this suggests a significant association between FMS composite score and injury risk. The specificity and sensitivity values were 0.91 and 0.54, respectively. This revealed that, while the proportion of true negatives to false negatives was high, the proportion of true positives to false positives was relatively even. Despite a very large OR of 11.67, only around half of the subsequent injuries were predicted by an FMS score of ≤14. Interestingly, this seminal article by Kiesel et al. [77] is often cited within the scientific literature and explains why a cut-off score of 14 is commonly used when researching the link between FMS and injury risk. Seven articles [27, 64, 70, 71, 73, 74, 76] have since replicated the finding that individuals achieving an FMS composite score of ≤14 have an increased likelihood of experiencing an injury; however, the degree of the relationship varies between studies. Differences in the number of participants, length of follow-up period, and sport/occupation of participants may have contributed to the inconsistencies in strength of relationship between FMS score and injury likelihood. In contrast, eight of the identified studies found no link between FMS composite score and injury risk [17, 63, 65–69, 75]. However, three of these studies [68, 69, 75] utilized very small sample sizes and this may explain the lack of association between FMS score and injury. There may simply have been too few injuries among the participants during the follow-up periods for any association to be observed. As such, the findings of these three studies [68, 69, 75] should carry minimal weight when making any judgment about the predictive value of the FMS.

Due to the inconsistency in findings, the graded recommendation for the FMS is ‘D’. A number of factors contribute to the ambiguity of the collective findings. First, the definition of injury was not consistent among the identified articles. Indeed, this is a common issue in sports medicine at large [79]. Kiesel et al. [77] classified an injury as membership of the injured reserve group and a time-loss of 3 weeks—presumably meaning that only relatively serious injuries were recorded. No details of injured reserve membership criteria or details of the specific injuries experienced were provided. In contrast, O’Connor et al. [76] defined injury as any damage to the body during training that resulted in an individual seeking medical care. This broad definition could have encompassed very minor injuries. McGill et al. [75] only considered back injuries that resulted in missed game play. Such variability in the classification of injuries makes it difficult to compare the results between each study. Similarly, the length of the follow-up period varied widely between studies, with the shortest reported window of observation being 6 weeks and the longest being 2 years [74, 75]. In some instances, the precise length of the injury-tracking period was not specified [19, 27, 67, 70]. It has been previously recommended that epidemiological sports injury studies should follow participants for at least 1 year, as this allows sufficient time for accumulation of exposure and injury events [79, 80]. Unfortunately, most of the identified studies followed participants for less than this time period, and this should be a consideration for future research.

Other relevant considerations that have been ignored by the vast majority of studies are accounting for exposure time and training load. These represent very influential confounding variables that are essential to drawing meaningful conclusions from future prospective studies. Interestingly, a number of populations were investigated by the included studies: athletes, military personnel, elite police officers, and firefighters. For instance, amongst the athlete group, individuals ranged from recreationally active to elite professionals. Given the range of occupations and performance levels of participants, it is perhaps to be expected that an inconsistent relationship between FMS score and injury should be observed when all studies are viewed collectively. The injury patterns between sports and occupations differ [81–83], hence the predictive value of the FMS may not be consistent across all populations. The use of the FMS composite score has been questioned since it is not a unitary construct and, as a result, may be a misleading value [84, 85]. Instead, it has been proposed that using the individual sub-test scores when analyzing FMS performance may be preferable. However, as is shown in Table 7, of the 18 prospective studies, ten reported an association between the composite score and injury likelihood, so it should not be disregarded entirely.

4.3 Landing Error Scoring System

Two studies investigating the link between LESS score and injury were identified through the literature search (Table 7) [31, 86]. Both studies prospectively screened participants before tracking them over the course of a sporting season. Smith et al. [31] did not report any

Table 7 Studies that investigated the relationship between movement screen scores and injury

Movement screen and references	Participant information			Association between scores and injury	Quality score (%)	Level of evidence
	Sample size	Occupation/sport	Age, years[a]			
FMS						
Bardenett et al. [63]	77 M, 90 F	Cross country, American Football, soccer, swimming, tennis and volleyball athletes (high school)	15.2	No association between composite score and injury	87	2++
Garrison et al. [64]	160	Swimming/diving, rugby and soccer athletes (NCAA Division I)	17–22	Score $\leq$13 = OR 9.52 (95% CI: 4.16–21.79)	80	2++
Hotta et al. [17]	84 M	Middle- and long-distance runners (collegiate)	20.0 ± 1.1	No association between composite score and injury Runners scoring $\leq$3 on the deep squat and active straight leg raise components = OR of 9.7 (95 % CI 2.1–44.4)	80	2++
McGill et al. [65]	53 M	Elite task force police officers	37.9 ± 5.0	No association between composite score and injury	80	2++
Teyhen et al. [66]	188 M	US army rangers	23.3 ± 3.7	No association between composite score and injury	73	2+
Warren et al. [67]	89 M, 78 F	Basketball, cross country, American Football, golf, T&F, tennis, volleyball, soccer and swimming/diving athletes (NCAA Division I)	20.6 ± 1.6 (injured) 20.0 ± 1.4 (non-injured)	No association between composite score and injury	80	2++
Zalai et al. [68]	20 M	Soccer players (professional)	23.0 ± 3.0	No association between composite score and injury Players who suffered an ankle injury received a lower score for the hurdle step sub-test ($p < 0.05$) Players who suffered a knee injury received a lower score for the deep squat sub-test ($p < 0.05$)	73	2+
Dossa et al. [69]	20 M	Ice-hockey players (elite youth)	16–20	No association between composite score and injury	80	2++
Kiesel et al. [70]	238 M	American Football players (professional)	Unknown	Injured vs. non-injured groups' mean scores 16.9 vs. 17.4 ($p < 0.05$) Score $\leq$14 = RR 1.87 (95 % CI 1.20–2.96) Players with at least one asymmetry had an RR of 1.80 (95 % CI 1.11–2.74)	60	2+
Knapik et al. [71]	770 M, 275 F	Coast guard cadets	18.1 ± 0.7 (M) 17.9 ± 0.7 (F)	M: score $\leq$11 = RR 1.64 (95 % CI 1.17–2.32) F: score $\leq$14 = RR 1.93 (95 % CI 1.27–2.95)	73	2+
Letafatkar et al. [19]	50 M, 50 F	Soccer, handball, and basketball players (recreational)	22.6 ± 3.0	Score <17 = OR 4.7	73	2+
Shojaedin et al. [72]	50 M, 50 F	Soccer, handball and basketball players (recreational)	22.6 ± 3.0	Score $\leq$17 = OR 4.7	53	2+
Butler et al. [73]	108	Firefighters	Unknown	Score $\leq$14 = OR 8.31 (95 % CI 3.2–21.6)	60	2+
Lisman et al. [74]	874 M	Marine officer candidates	22.4 ± 2.7	Score $\leq$14 = OR 2.04 (95 % CI 1.32–3.15)	60	2+
McGill et al. [75]	14 M	Basketball players (collegiate)	20.4 ± 1.6	No association found between composite score and injury	73	2+
O'Connor et al. [76]	874 M	Marine officer candidates	18–30	Score $\leq$14 = RR 1.5 ($p < 0.05$)	67	2+
Chorba et al. [27]	38 F	Soccer, volleyball, and basketball players (NCAA Division II)	19.2 ± 1.2	Score $\leq$14 = OR 3.85 (95 % CI 0.98–15.13)	80	2++

Table 7 continued

Movement screen and references	Participant information			Association between scores and injury	Quality score (%)	Level of evidence
	Sample size	Occupation/sport	Age, years[a]			
Kiesel et al. [77]	46 M	American Football players (professional)	Unknown	Injured vs. non-injured groups' mean scores 14.3 vs. 17.4 ($p < 0.05$) Score ≤14 = OR 11.67 (95 % CI 2.47–54.52)	53	2+
LESS						
Padua et al. [86]	348 M, 481 F	Soccer players (elite youth)	13.9 ± 1.8	ACL injured vs. non-injured groups' mean scores 6.2 vs. 4.4 ($p < 0.05$) Score ≥5 = RR 10.7 for indirect and non-contact ACL injury	73	2+
Smith et al. [31]	29 M, 63 F	Lacrosse, soccer, basketball, American Football, field hockey, gymnastics (high school/collegiate)	18.3 ± 2.0	No association between score and non-contact ACL injury	87	2++

ACL anterior cruciate ligament, *CI* confidence interval, *F* female, *FMS* functional movement screen, *LESS* Landing Error Scoring System, *M* male, *NCAA* National Collegiate Athletic Association, *OR* odds ratio, *RR* risk ratio, *T&F* track and field, *2+* well-conducted study with low risk of bias, *2++* high-quality study with very low risk of bias

[a] Data are presented as mean ± standard deviation or range

significant relationship between LESS score and subsequent injury, whereas Padua et al. [86] did. Those ranked by Smith et al. [31] as 'poor' (scoring >6) displayed an OR of 3.62 compared with those ranked as 'excellent' (scoring ≤4) but the 95 % CI crossed 1 (0.87–15.11), indicating that the groups most likely did not differ in their risk of injury. However, Smith et al. [31] only included grade III (complete tear) non-contact anterior cruciate ligament (ACL) injuries in their analysis and, as such, it is not clear if the LESS score was associated with any other type of injury. The LESS protocol involves whole body movement, so the outcome score may potentially display an association with other injury types. The apparent lack of connection between the LESS and ACL injury reported by Smith et al. [31] is surprising, since the screen assesses the degree of knee valgus and flexion during landing, which are both relevant factors to both patellofemoral pain and ACL injury [87]. It was suggested that the narrow range of recorded scores (only 0–11 out of a possible 19) could have contributed to the lack of association with injury [31]. The authors also postulated that the screen may have superior predictive ability with regard to injury among less well-trained or less physically mature individuals undergoing rapid neuromuscular development. This may be due to differences in proprioceptive awareness and strength among these groups compared with more physically mature, well-trained individuals. This theory is somewhat supported by the findings of Padua et al. [86], who observed an almost 11-fold greater risk of ACL injury among individuals with scores of ≥5 compared with those scoring <5. The average age of the participants followed by Padua et al. [86] was 14 years compared with 18 years for the cohort observed by Smith et al. [31]. It may be that the LESS does have some injury-predictive ability but only amongst young populations and in certain sports. Further research is required among both younger and older populations before any firm conclusions can be made regarding that suggestion. Despite a theoretical link between the LESS and lower body injury, especially ACL injury, the evidence is currently ambiguous. As only two studies have prospectively investigated the ability of the LESS to predict injury and they reported conflicting results, the graded recommendation for this movement screen is 'D'.

5 Limitations and Recommendations for Future Research

When interpreting the results of the identified articles, it is important the reader be cognizant of a number of common limitations. The majority of reliability studies were categorized as methodologically poor. While the ICC or kappa scores reported often indicated good to excellent agreement within and between raters, the true value of these findings can be questioned because of the aforementioned methodological quality of the studies. In future studies investigating the reliability of movement screens, rater information such as occupation, years of experience, and number of tests performed should be included to allow for a more thorough interpretation of the results. In addition, larger sample sizes would help improve the methodological quality of future reliability studies. Similarly, future studies investigating the ability of movement screens to predict injury should clearly define what an 'injury' is and state the length of the observation period to allow contextual appraisal of the results. In some sports, such as soccer and rugby, established guidelines for injury reporting already exist [79, 80]. None of the studies investigating the link between movement screening score and injury reported or accounted for the exposure time of the participants. This is a crucial point that must be considered by future studies; without this information, a significant confounding variable is being ignored. All else being equal, the less time a player spends training and playing, then the less opportunity they have to get injured. Readers are not currently able to determine from the current research whether individuals with supposed poorer movement ability are actually at increased risk of injury because of that or simply because their exposure time is greater. Another issue to consider is that if the individual responsible for recording injuries knows the movement screening scores, then an element of bias may exist. Ideally, the individual recording the injury occurrence should be blinded to the outcome of the movement screen.

6 Conclusion

The majority of movement screens identified through the literature search lack a substantial evidence base in relation to both their reliability and their ability to predict injury. However, due to its extensive research base, the FMS is the only movement screen that has consistently demonstrated good intra- and inter-rater reliability. In addition, some studies have suggested possible predictive ability with regard to injury risk for the FMS and LESS; however, this is not a unanimous finding. Based purely on the reported ICCs and kappa values, all identified screens appear to have good reliability with the exception of the various single-leg squat screens and the SEBT movement quality screen. Further research is warranted to verify the initial reliability values for the identified movement screens, since the evidence base is still limited and the majority of the identified reliability studies were classified as methodologically poor. None of the identified movement screens have enough supporting evidence to justify them being heralded as injury prediction tools. Overall, movement screening may be useful for practitioners to enhance their holistic knowledge of an athlete, but it seems the subjectivity of scoring makes it difficult to apply these results to injury prediction with any degree of certainty.

Compliance with Ethical Standards

Funding No funding was provided that contributed to the development of this manuscript. Robert McCunn and Hugh Fullagar are supported by 'Science and Health in Football' scholarships funded by the DAAD (German Academic Exchange Service).

Conflict of interest Robert McCunn, Tim Meyer, Hugh Fullagar, and Karen aus der Fünten declare that they have no conflicts of interest with the content of this review. Ian McKeown declares that he is the primary developer of the athletic ability assessment (AAA) movement screen; however, he does not stand to receive financial gain as a result. Additionally, the AAA is neither a patented commodity nor is it protected by copyright.

References

1. Mendez-Villanueva A, Buchheit M. Football-specific fitness testing: adding value or confirming the evidence? J Sports Sci. 2013;31(13):1503–8.
2. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. Br J Sports Med. 2014;48(18):1352–7.
3. McGill S, Frost D, Andersen J, et al. Movement quality and links to measures of fitness in firefighters. Work. 2013;45(3):357–66.
4. Frost D, Andersen J, Lam T, et al. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics. 2013;56(4):637–49.
5. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. Sports Med. 1992;14(2):82–99.
6. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function-part 2. Int J Sports Phys Ther. 2014;9(4):549–63.
7. Giblin S, Collins D, Button C. Physical literacy: importance, assessment and future directions. Sports Med. 2014;44(9):1177–84.
8. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function—part 1. Int J Sports Phys Ther. 2014;9(3):396–409.
9. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport. 2003;4(3):122–8.
10. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217–38.
11. Smith CA, Chimera NJ, Wright NJ, et al. Interrater and intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):982–7.
12. Parenteau-G E, Gaudreault N, Chambers S, et al. Functional movement screen test: a reliable screening test for young elite ice hockey players. Phys Ther Sport. 2014;15(3):169–75.
13. Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. J Strength Cond Res. 2012;26(2):408–15.
14. Shultz R, Anderson SC, Matheson GO, et al. Test–retest and interrater reliability of the functional movement screen. J Athl Train. 2013;48(3):331–6.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
16. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651–7.
17. Hotta T, Nishiguchi S, Fukutani N, et al. Functional movement screen for predicting running injuries in 18–24 year-old competitive male runners. J Strength Cond Res. 2015. [Epub ahead of print].
18. Gulgin H, Hoogenboom B. The functional movement screening (fms)™: an inter-rater reliability study between raters of varied experience. Int J Sports Phys Ther. 2014;9(1):14–20.

19. Letafatkar A, Hadadnezhad M, Shojaedin S, et al. Relationship between functional movement screening score and history of injury. Int J Sports Phys Ther. 2014;9(1):21–7.
20. Elias JE. The inter-rater reliability of the functional movement screen within an athletic population using untrained raters. J Strength Cond Res. 2013. [Epub ahead of print].
21. Gribble PA, Brigle J, Pietrosimone BG, et al. Intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):978–81.
22. Frohm A, Heijne A, Kowalski J, et al. A nine-test screening battery for athletes: a reliability study. Scand J Med Sci Sports. 2012;22(3):306–15.
23. Klusemann MJ, Pyne DB, Fay TS, et al. Online video-based resistance training improves the physical capacity of junior basketball athletes. J Strength Cond Res. 2012;26(10):2677–84.
24. Teyhen DS, Shaffer SW, Lorenson CL, et al. The functional movement screen: a reliability study. J Orthop Sports Phys Ther. 2012;42(6):530–40.
25. Butler RJ, Plisky PJ, Kiesel KB. Interrater reliability of videotaped performance on the functional movement screen using the 100-point scoring scale. Athl Train Sports Health Care. 2012;4(3):103–9.
26. Schneiders AG, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. Int J Sports Phys Ther. 2011;6(2):75–82.
27. Chorba RS, Chorba DJ, Bouillon LE, et al. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. N Am J Sports Phys Ther. 2010;5(2):47–54.
28. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. J Strength Cond Res. 2010;24(2):479–86.
29. 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. 2013. [Epub ahead of print].
30. McKeown I, Ball N. Current practices of long term athlete development of junior athletes in high performance sport environments. J Aust Strength Cond. 2013;21(1):4–12.
31. Smith HC, Johnson RJ, Shultz SJ, et al. A prospective evaluation of the Landing Error Scoring System (LESS) as a screening tool for anterior cruciate ligament injury risk. Am J Sports Med. 2012;40(3):521–6.
32. Padua DA, Boling MC, Distefano LJ, et al. Reliability of the landing error scoring system-real time, a clinical assessment tool of jump-landing biomechanics. J Sport Rehabil. 2011;20(2):145–56.
33. Onate J, Cortes N, Welch C, et al. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. J Sport Rehabil. 2010;19(1):41–56.
34. Padua DA, Marshall SW, Boling MC, et al. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am J Sports Med. 2009;37(10):1996–2002.
35. Junge T, Balsnes S, Runge L, et al. Single leg mini squat: an inter-tester reproducibility study of children in the age of 9–10 and 12–14 years presented by various methods of kappa calculation. BMC Musculoskelet Disord. 2012;13:203.
36. Crossley KM, Zhang WJ, Schache AG, et al. Performance on the single-leg squat task indicates hip abductor muscle function. Am J Sports Med. 2011;39(4):866–73.
37. Örtqvist M, Moström EB, Roos EM, et al. Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children. Knee Surg Sports Traumatol Arthrosc. 2011;19(12):2060–6.
38. Ageberg E, Bennell KL, Hunt MA, et al. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskelet Disord. 2010;11:265.
39. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. J Orthop Sports Phys Ther. 2007;37(3):122–9.
40. Noyes FR, Barber-Westin SD, Fleckenstein C, et al. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. Am J Sport Med. 2005;33(2):197–207.
41. Barber-Westin SD, Galloway M, Noyes FR, et al. Assessment of lower limb neuromuscular control in prepubescent athletes. Am J Sport Med. 2005;33(12):1853–60.
42. Dingenen B, Malfait B, Nijs S, et al. Can two-dimensional video analysis during single-leg drop vertical jumps help identify non-contact knee injury risk? A one-year prospective study. Clin Biomech. 2015. [Epub ahead of print].
43. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. J Sport Rehabil. 2012;21(1):7–11.
44. Nilstad A, Andersen TE, Kristianslund E. Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening. J Orthop Sport Phys. 2014;44(5):358–65.
45. Whatman C, Hume P, Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Phys Ther Sport. 2013;14(3):168–74.
46. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. J Orthop Sports Phys Ther. 2009;39(9):665–74.
47. Gwet KL. Handbook of inter-rater reliability; the definitive guide to measuring the extent of agreement among multiple raters. 3rd ed. Gaithersburg: Advanced Analytics, LLC; 2012.
48. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. Athl Ther Today. 2008;13(5):39–44.
49. Herman DC, Weinhold PS, Guskiewicz KM, et al. The effects of strength training on the lower extremity biomechanics of female recreational athletes during a stop-jump task. Am J Sport Med. 2008;36(4):733–40.
50. Söderman K, Werner S, Pietilä T, et al. Balance board training: prevention of traumatic injuries of the lower extremities in female soccer players? A prospective randomized intervention study. Knee Surg Sports Traumatol Arthrosc. 2000;8(6):356–63.
51. Herrington L, Myer GD, Munro A. Intra and inter-tester reliability of the tuck jump assessment. Phys Ther Sport. 2013;14(3):152–5.
52. Dudley LA, Smith CA, Olson BK, et al. Interrater and intrarater reliability of the tuck jump assessment by health professionals of varied educational backgrounds. J Sports Med (Hindawi Publ Corp). 2013;2013:Article ID 483503.
53. McKeown I, Taylor-McKeown K, Woods C, et al. Athletic ability assessment: a movement assessment protocol for athletes. Int J Sports Phys Ther. 2014;9(7):862–73.
54. Parsonage JR, Williams RS, Rainer P, et al. Assessment of conditioning-specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. J Strength Cond Res. 2014;28(6):1497–506.
55. Reid DA, Vanweerd RJ, Larmer PJ, et al. The inter and intra rater reliability of the Netball Movement Screening Tool. J Sci Med Sport. 2014;18(3):353–7.
56. Tarara DT, Hegedus EJ, Taylor JB. Real-time test–retest and interrater reliability of select physical performance measures in physically active college-aged students. Int J Sports Phys Ther. 2014;9(7):874–87.
57. Plisky PJ, Rauh MJ, Kaminski TW, et al. Star excusrsion balance test as a predictor of lower extremity injury in high school basketball players. J Orthop Sports Phys Ther. 2006;36(12):911–9.

58. Ness BM, Taylor AL, Haberl MD, et al. Clinical observation and analysis of movement quality during performance on the star excursion balance test. Int J Sports Phys Ther. 2015;10(2):168–77.
59. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377–84.
60. McCall A, Carling C, Davison M, et al. Injury risk factors, screening tests and preventative strategies: a systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. Br J Sports Med. 2015;49(9):583–9.
61. Freckleton G, Pizzari T. Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. Br J Sports Med. 2013;47(6):351–8.
62. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001;323(7308):334–6.
63. Bardenett SM, Micca JJ, DeNoyelles JT, et al. Functional movement screen normative values and validity in high school athletes: can the FMS be used as a predictor of injury? Int J Sports Phys Ther. 2015;10(3):303–8.
64. Garrison M, Westrick R, Johnson MR, et al. Association between the functional movement screen and injury development in collegiate athletes. Int J Sports Phys Ther. 2015;10(1):21–8.
65. McGill S, Frost D, Lam T, et al. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. Ergonomics. 2015. [Epub ahead of print].
66. Teyhen DS, Shaffer SW, Butler RJ, et al. What risk factors are associated with musculoskeletal injury in US army rangers? A prospective prognostic study. Clin Orthop Relat Res. 2015;473(9):2948–58.
67. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. J Sport Rehabil. 2015;24(2):163–70.
68. Zalai D, Panics G, Bobak P, et al. Quality of functional movement patterns and injury examination in elite-level male professional football players. Acta Physiol Hung. 2015;102(1):34–42.
69. Dossa K, Cashman G, Howitt S, et al. Can injury in major junior hockey players be predicted by a pre-season functional movement screen—a prospective cohort study. J Can Chiropr Assoc. 2014;58(4):421–7.
70. Kiesel KB, Butler RJ, Plisky PJ. Prediction of injury by limited and asymmetrical fundamental movement patterns in american football players. J Sport Rehabil. 2014;23(2):88–94.
71. Knapik JJ, Cosio-Lima LM, Reynolds KL, et al. Efficacy of functional movement screening for predicting injuries in coast guard cadets. J Strength Cond Res. 2014;29(5):1157–62.
72. Shojaedin SS, Letafatkar A, Hadadnezhad M, et al. Relationship between functional movement screening score and history of injury and identifying the predictive value of the FMS for injury. Int J Inj Contr Saf Promot. 2014;21(4):355–60.
73. Butler RJ, Contreras M, Burton LC, et al. Modifiable risk factors predict injuries in firefighters during training academies. Work. 2013;46(1):11–7.
74. Lisman P, O'Connor FG, Deuster PA, et al. Functional movement screen and aerobic fitness predict injuries in military training. Med Sci Sports Exerc. 2013;45(4):636–43.
75. McGill SM, Andersen JT, Horne AD. Predicting performance and injury resilience from movement quality and fitness scores in a basketball team over 2 years. J Strength Cond Res. 2012;26(7):1731–9.
76. O'Connor FG, Deuster PA, Davis J, et al. Functional movement screening: predicting injuries in officer candidates. Med Sci Sports Exerc. 2011;43(12):2224–30.
77. 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;2(3):147–58.
78. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. BMJ. 1994;308(6943):1552.
79. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Br J Sports Med. 2006;40(3):193–201.
80. Fuller CW, Molloy MG, Bagate C, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. Br J Sports Med. 2007;41(5):328–31.
81. Almeida SA, Williams KM, Shaffer RA, et al. Epidemiological patterns of musculoskeletal injuries and physical training. Med Sci Sports Exerc. 1999;31(8):1176–82.
82. Wolf BR, Ebinger AE, Lawler MP, et al. Injury patterns in Division I collegiate swimming. Am J Sports Med. 2009;37(10):2037–42.
83. Bathgate A, Best JP, Craig G, et al. A prospective study of injuries to elite Australian rugby union players. Br J Sports Med. 2002;36(4):265–9.
84. Kazman JB, Galecki JM, Lisman P, et al. Factor structure of the functional movement screen in marine officer candidates. J Strength Cond Res. 2014;28(3):672–8.
85. Li Y, Wang X, Chen X, et al. Exploratory factor analysis of the functional movement screen in elite athletes. J Sports Sci. 2015;33(11):1166–72.
86. Padua DA, DiStefano LJ, Beutler AI, et al. The landing error scoring system as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. J Athl Train. 2015;50(6):589–95.
87. Myer GD, Ford KR, Di Stasi SL, et al. High knee abduction moments are common risk factors for patellofemoral pain (PFP) and anterior cruciate ligament (ACL) injury in girls: is PFP itself a predictor for subsequent ACL injury? Br J Sports Med. 2015;49(2):118–22.

BJSM Online First, published on June 8, 2016 as 10.1136/bjsports-2016-096413

CORRESPONDENCE

Screening for risk factors: if you liked it then you should have put a number on it

The usefulness of screening tests to predict injury has recently been questioned.[1 2] However, from our point of view, it is important that screening not be completely demonised. A screening test may not accurately identify who will get injured (albeit providing a likelihood) but this should not result in obsolescence. Usefulness is not solely dependent on predictive ability. As Bahr[1] highlights: causative relationships (ie, injury risk factors) can be identified using screening tests, for example, eccentric hamstring weakness.[3] Should we therefore completely abolish screening tests because we cannot definitively say that poor eccentric hamstring strength will 100% result in a strain injury? Screening puts a number on an attribute, which allows us to quantify injury risk and, in turn, modify the design of injury prevention strategies. Clearly, the syntax surrounding screening tests needs to be adapted. Bahr[1] is absolutely correct that the practice of applied practitioners using screening results to categorise their athletes into intervention and 'control' groups is not supported by the evidence, but let us not throw the baby out with the bathwater.

Robert McCunn, Tim Meyer

Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany

Correspondence to Robert McCunn, Institute of Sport and Preventive Medicine, Saarland University, Campus Bldg B8.2, 66123 Saarbrücken, Germany; robert.mccunn@uni-saarland.de

Twitter Follow Robert McCunn at @RobertMcCunn

Competing interests None declared.

Provenance and peer review Not commissioned; internally peer reviewed.

To cite McCunn R, Meyer T. *Br J Sports Med* Published Online First: [*please include* Day Month Year] doi:10.1136/bjsports-2016-096413

Accepted 17 May 2016

Br J Sports Med 2016;**0**:1. doi:10.1136/bjsports-2016-096413

REFERENCES

1 Bahr R. Why screening tests to predict injury do not work-and probably never will…: a critical review. *Br J Sports Med* Published Online First: 19 Apr 2016 doi:10.1136/bjsports-2016-096256
2 McCunn R, aus der Fünten K, Fullagar HHK, *et al*. Reliability and association with injury of movement screens: a critical review. *Sports Med*. Published Online First: 31 Dec 2015.
3 Timmins RG, Bourne MN, Shield AJ, *et al*. Short biceps femoris fascicles and eccentric knee flexor weakness increase the risk of hamstring injury in elite football (soccer): a prospective cohort study. *Br J Sports Med* Published Online First: 16 Dec 2015 doi:10.1136/bjsports-2015-095362

SCIENCE AND MEDICINE IN FOOTBALL, 2016
http://dx.doi.org/10.1080/24733938.2016.1256554

Routledge
Taylor & Francis Group

POINT

Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement?

Robert McCunn[a], John A Sampson[b], Matthew Whalan[b,c,d] and Tim Meyer[a]

[a]Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany; [b]Centre for Human and Applied Physiology, School of Medicine, University of Wollongong, Wollongong, New South Wales, Australia; [c]Player Welfare Department, Football South Coast, New South Wales, Australia; [d]Wollongong Wolves FC, Wollongong, New South Wales, Australia

ABSTRACT
In 2006 a consensus statement recommending how football injury data should be collected and reported was published. These recommendations have provided a useful framework for research over the last 10 years. However, many questions related to the underlying methodology of studies concerned with injury epidemiology and prevention in football still exist. This is particularly true for research conducted in non-professional environments. The present point-counterpoint article highlights some of these issues and asks the question: are we in need of an updated consensus statement?

KEYWORDS
Epidemiology; injury; semi-professional; amateur; soccer

Football participation carries an inherent risk of injury. Consistent and accurate records of injury are imperative for planning, evaluating and delivering injury prevention programmes such as the FIFA 11+ (van Mechelen et al. 1992). In this regard, Fuller et al.'s (2006) consensus statement serves research within full-time professional football well by streamlining research methods and providing definitions/recommendations for injury data collection. Furthermore, Fuller et al.'s (2006) definition of injury (any physical complaint sustained as a result of football activities) and the proposed injury incident data recording sheets are appropriate for both full- and part-time players. However, problems are encountered when applying some of the guidelines to "part-timers", who cumulatively compose the largest playing group (e.g., youth academy, semi-professional and recreational/community level players). In our view, there are two main issues at the forefront of injury research conducted within part-time groups that require clarification: (1) who should record the data and (2) how should injury severity be quantified? A number of solutions to each question may exist. This point-counterpoint article presents various options for discussion in an attempt to ensure future injury research conducted within part-time playing groups is of high methodological quality.

Data collection

Collecting accurate and reliable data is essential if one hopes to present valid results that allow consistent comparisons across the literature. However, within existing studies of part-time playing groups, there are discrepancies. For example, the data has been collected by various individuals including (1) coaches (Ekstrand & Hilding 1999; Froholdt et al. 2009), (2) parents (Emery et al. 2005; Emery & Meeuwisse 2006), (3) medical staff (Brito et al. 2012; aus der Fünten et al. 2014; Herrero et al. 2014; Silvers et al. 2015) and (4) academic researchers (McNoe & Chalmers 2010; Schmikli et al. 2011; Hammes et al. 2016). The varying depth of medical knowledge and skills among these parties may consequently lead to inconsistencies in data collection and present difficulties when comparing outcomes. For example, all parties would most likely correctly record certain types of "dramatic" injuries such as bone breaks; however, discrepancies may arise when more ambiguous cases occur, e.g., minor muscle strains. Indeed, certain groups may over-report or under-report the incidence of injuries depending on their relationship with the players and their specific role within the club. For example, players may withhold physical complaints from their coach for fear of not being selected to play in matches or student research assistants may not be assertive enough when attempting to obtain relevant injury information from players they are unfamiliar with. Medical practitioners (e.g., doctors and physiotherapists) are undoubtedly the most qualified to diagnose and record injuries. However, the majority of part-time teams do not have immediate access to medical practitioners. An injury recording method that relies on individuals with this level of expertise would therefore, in many cases, make large-scale research projects untenable. Alternatively, player self-reporting of exposure time and injury incidence (e.g., via online resources) removes the requirement for any third party to record the information. However, self-reporting arguably raises questions over compliance and data accuracy. In our view, the most accurate data can be collected from a third party present at training and on match days. Hence, we suggest each team must appoint a primary data collector who would record basic details such as exposure, injury location, incident description and symptoms. Much of the research

CONTACT Robert McCunn bob.mccunn@me.com

highlighted earlier has applied this method; however, we propose that this method should be improved by stipulating that the nominated individual must posses a minimum standard of medical knowledge and undergo training with respect to injury data collection procedures. Such an approach would ensure the recording of detailed injury descriptions, facilitating retrospective injury diagnosis by trained medical professionals (Hammes et al. 2016). Allowing non-experts such as parents, coaches and university students to contribute in this manner offers a viable strategy to ensure high-quality data is collected, and for research teams to conduct meaningful large-scale projects. The next step in this process would be for the scientific community to agree upon the extent and content of the required training. Some variability in the methodology between studies is inevitable. Hence, the importance of meticulously describing data collection protocols cannot be overstated, since it is vital in allowing readers to judge the quality of the results presented.

Injury quantification

Reporting injury incidence per 1000 h of exposure as recommended by Fuller et al. (2006) seems appropriate for part-time playing groups. However, Fuller et al.'s (2006) guidelines for recording injury severity (days lost from full participation in training/match play) are problematic in a part-time environment where, in contrast to full-time athletes, players are not seen by medical staff on a daily basis. Indeed, several days between training/match days may pass when the part-time player is not seen by anyone associated with the club and the opportunity to observe a "return to full participation" may only arise once or twice per week. One potential method of determining injury severity is via the player self-reporting recovery by providing the date they believe they were fit to return to full participation. Similarly, players could report perceived injury severity, e.g., using a Likert-type scale via an online questionnaire to indicate how serious they believe any given case is (Clarsen et al. 2014). Alternatively, follow-up appointments with injured players by medical practitioners (either via phone or in-person) on non-training/playing days could be used. However, each of these options poses compliance and logistical implementation issues. We suggest that the definition proposed by Fuller et al. (2006) should apply even though adopting such an approach would almost certainly overestimate injury severity. The greatest impact of overestimation would be observed at the less serious end of the spectrum. For example, in the event of a player suffering an injury on a Saturday match day, missing their only scheduled training session the following Tuesday and returning to play the subsequent Saturday match day, a six-day lay-off would be recorded. However, the player may have been fit to train/play by Wednesday meaning the injury should have been classified as a three-day time-loss injury. As such, some "minimal" injuries may be erroneously recorded as "mild" (Fuller et al. 2006). However, upholding the current definition consistently across the literature would, at least, provide comparable data sets.

Conclusion

In summary, while a number of solutions to the highlighted questions may exist, it is important to acknowledge that each is somewhat flawed. Agreement surrounding the recommendations for injury data collection procedures among part-time playing groups is thus crucial. Such consensus will allow comparisons between studies and ensure that the conclusions drawn from future research are meaningful. Any updated consensus statement should carefully consider the logistical implications for researchers when making recommendations. Herein, the arguments have been presented with a focus on football. However, the issues presented apply across sports and as such epidemiologists and applied practitioners from different domains would add value to this discussion.

Disclosure statement

No potential conflict of interest was reported by the authors.

Related Articles

For the counterpoints to this point please see: Caroline Finch, 'Injury data collection in lower leagues needs to be targeted specifically to those settings', http://dx.doi.org/10.1080/24733938.2016.1256555;

Ben Clarsen, 'Current severity measures are insufficient for overuse injuries', http://dx.doi.org/10.1080/24733938.2016.1256579;

Martin Hägglund, 'Data collection procedures for football injuries in lower leagues: Is there a need for an updated consensus statement?' http://dx.doi.org/10.1080/24733938.2016.1256581.

References

aus der Fünten K, Faude O, Lensch J, Meyer T. 2014. Injury characteristics in the German professional male soccer leagues after a shortened winter break. J Athl Train. 49:786–793.

Brito J, Malina RM, Seabra A, Massada JL, Soares JM, Krustrup P, Rebelo A. 2012. Injuries in Portuguese youth soccer players during training and match play. J Athl Train. 47:191–197.

Clarsen B, Rønsen O, Myklebust G, Flørenes TW, Bahr R. 2014. The Oslo Sports Trauma Research Center questionnaire on health problems: a new approach to prospective monitoring of illness and injury in elite athletes. Br J Sports Med. 48:754–760.

Ekstrand J, Hilding J. 1999. The incidence and differential diagnosis of acute groin injuries in male soccer players. Scand J Med Sci Sports. 9:98–103.

Emery CA, Meeuwisse WH. 2006. Risk factors for injury in indoor compared with outdoor adolescent soccer. Am J Sports Med. 34:1636–1642.

Emery CA, Meeuwisse WH, Hartmann SE. 2005. Evaluation of risk factors for injury in adolescent soccer: implementation and validation of an injury surveillance system. Am J Sports Med. 33:1882–1891.

Froholdt A, Olsen OE, Bahr R. 2009. Low risk of injuries among children playing organized soccer: a prospective cohort study. Am J Sports Med. 37:1155–1160.

Fuller CW, Ekstrand J, Junge A, Andersen TE, Bahr R, Dvorak J, Meeuwisse WH. 2006. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Br J Sports Med. 40:193–201.

Hammes D, aus der Fünten K, Bizzini M, Meyer T. 2016. Injury prediction in veteran football players using the Functional Movement Screen™. J Sports Sci. 34:1371–1379.

Herrero H, Salinero JJ, Del Coso J. 2014. Injuries among Spanish male amateur soccer players: a retrospective population study. Am J Sports Med. 42:78–85.

McNoe BM, Chalmers DJ. 2010. Injury in community-level soccer: development of an injury surveillance system. Am J Sports Med. 38:2542–2551.

Schmikli SL, de Vries WR, Inklaar H, Backx F. 2011. Injury prevention target groups in soccer: injury characteristics and incidence rates in male junior and senior players. J Sci Med Sport. 14:199–203.

Silvers-Granelli H, Mandelbaum B, Adeniji O, Insler S, Bizzini M, Pohlig R, Junge A, Snyder-Mackler L, Dvorak J. 2015. Efficacy of the FIFA 11+ injury prevention program in the collegiate male soccer player. Am J Sports Med. 43:2628–2637. [E-pub ahead of print].

van Mechelen W, Hlobil H, Kemper HC. 1992. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. Sports Med. 14:82–99.

SCIENCE AND MEDICINE IN FOOTBALL, 2017
http://dx.doi.org/10.1080/24733938.2017.1283436

Functional Movement Screen (FMS™) score does not predict injury in English Premier League youth academy football players

Florence Newton[a,d], Alan McCall[a,b], Desmond Ryan[a], Colin Blackburne[a], Karen aus der Fünten[c], Tim Meyer[c], Colin Lewin[a] and Robert McCunn[c]

[a]Research & Development Department, Arsenal Football Club, London, UK; [b]Research & Development Department, Edinburgh Napier University, Edinburgh, UK; [c]Institute of Sport and Preventive Medicine, Saarland University, Saarbrücken, Germany; [d]Institute of Sport, Exercise and Health, University College London, London, UK

ABSTRACT

Purpose: despite being commonly used, the interaction between Functional Movement Screen (FMS™) score and injury in any elite football population has not been studied. The aim of the present study was to investigate the relationship between FMS™ score and non-contact injury among elite youth players from a Premier League football academy.

Materials and methods: eighty-four players were screened during the pre-season period and non-contact injuries recorded prospectively for the entirety of the 2013/14 football season. Logistic regression analysis was utilized to explore the relationships between the individual sub-tests of the FMS™ and injury. Receiver operating characteristic (ROC) curves were used to assess the predictive value of the FMS™ composite score.

Results: logistic regression revealed no relationships between score achieved on the individual sub-tests and injury. ROC curves indicated poor predictive ability of the composite score. Players scoring below the identified cut-off values (≤14 or ≤15 depending on injury type considered) were 0.66 (95%CI: 0.40-1.10), 0.70 (95%CI: 0.32-1.57) and 1.52 (95%CI: 0.50-4.61) times as likely to suffer 'any', 'overuse' and 'severe' injuries respectively than those who scored above the identified cut-off values.

Conclusions: there was no relationship between FMS™ score and injury. It was unable to predict any non-contact injury among English Premier League youth academy players.

Practical implications: The present findings suggest that the FMS™ should not be used for risk stratification among young elite soccer players since the composite score was unrelated to injury likelihood. However, the FMS™ may be useful in other ways. For example, it may provide useful information to applied practitioners when designing strength-training programs for groups of players they are unfamiliar with, as is often the case at the start of a new season.

ARTICLE HISTORY
Accepted 8 December 2016

KEYWORDS
Screening; soccer; risk; adolescent; elite

Introduction

Injuries in male elite youth football players have been shown to range from 2.0 to 19.4 injuries/1000 h of total exposure (9.5–48.7 injuries/1000 h of match exposure) with strains and sprains among the most common injury types mainly occurring in the upper leg, knee and ankle (Pfirrmann et al. 2016). Injuries in elite youth players are of particular concern to elite youth academy directors as time lost from training and matches has the potential to negatively affect the development of technical, tactical, physical and mental qualities of players. Indeed, a large-scale epidemiological study of elite male youth French players showed that those incurring more severe injuries were less likely to obtain a professional contract (Le Gall et al. 2009). As such, strategies aimed at reducing the risk of injury during the development period of young elite footballers should be emphasised.

While injury risk is multifactorial and complex (Bittencourt et al. 2016), one injury risk screening tool that is commonly used and deemed important by premier league football teams is the Functional Movement Screen (FMS™) (McCall et al. 2014). The purpose of this screening tool is to assess the movement quality of an individual (Cook et al. 2006a). Movement quality is not well defined but one definition put forward is the ability to "maintain correct posture and joint alignment in addition to balance while performing selected movements" (McCunn et al. 2016). The FMS™ has displayed moderate-to-good intra- and inter-rater reliability (McCunn et al. 2016; Moran et al. 2016). The underlying theory behind movement screening is that "poor" movement quality may be a contributing factor to non-contact injury. To date, evidence relating to any potential relationship between non-contact injury and FMS score is conflicting (McCall et al. 2015; McCunn et al. 2016).

Despite its widespread use within professional football clubs, only one prospective cohort study has investigated the relationship between the FMS™ and injury in top-league football players (Zalai et al. 2015). Indeed, none has been conducted with elite youth footballers. A recent systematic review that sought to outline the evidence behind the practices and perceptions of elite football clubs' injury prevention strategies concluded that insufficient evidence existed to make any recommendation in relation to the FMS™ (McCall

CONTACT Robert McCunn robert.mccunn@uni-saarland.de

et al. 2015). Therefore, the aim of the present study was to determine whether a causative relationship existed between FMS™ score and injury among male players from an English Premier League youth football academy.

Methods

Experimental design

The present study followed a prospective cohort design. Players meeting the inclusion criteria were assessed using the FMS™ during the preseason period. Injury surveillance was performed over the entirety of the subsequent season (2013/14) and all injury events recorded in accordance with the recommendations provided by Fuller et al. (2006).

Participants

Eighty-four male players registered with an English Premier League football club youth academy agreed to participate in the present study (age 13.0 ± 1.3 years, height 167.0 ± 9.4 cm, body mass 55.8 ± 11.4 kg). Inclusion criteria required players to be registered with the club for the entirety of the observation season (2013/14), injury free at the initiation of the preseason period (1st June) and eligible for the under 12, -13, -14, -15 or -16 squads. Participant assent and written parental consent were obtained prior to all testing procedures. The study was approved by the University College London Research Ethics Committee and conformed to the Declaration of Helsinki.

Procedures

All FMS™ testing was conducted by United Kingdom Strength and Conditioning Association-accredited coaches or chartered physiotherapists. All testers had multiple years experience in conducting such assessments and undertook a recap of all procedures prior to testing each year. Standardised written instructions that followed the original test guidelines were provided for all raters and were delivered verbatim when instructing participants (Cook et al. 2006a, 2006b). Official FMS™ test kit was used. Each participant completed all seven subtests sequentially in the following order: deep squat, hurdle step, in-line lunge, shoulder mobility, active straight leg raise, trunk stability push up and rotatory stability.

All injuries sustained during matches and training sessions were recorded and monitored by club physiotherapists in accordance with the recommendations provided by Fuller et al. (2006). Injury was defined as any physical complaint sustained by a player resulting from a football match or football training session that resulted in time loss. However, only non-contact injuries were included for analysis. Non-contact injuries included were also further categorised into two additional groups: overuse and severe. An overuse injury was one that was characterised by repeated microtrauma without a single, identifiable event while severe injuries were those that resulted in a time loss of more than 28 days (Fuller et al. 2006).

Statistical analyses

Data were analysed using SPSS Statistics version 22 (SPSS, Inc., Chicago, Illinois, USA) and MedCalc for Windows, version 16.4.3 (MedCalc Software, Ostend, Belgium). FMS™ composite scores were compared between injured and uninjured players (using three injury definitions: all non-contact, overuse and severe) using Mann–Whitney *U*-tests. In addition, Cohen's *d* effect sizes (ES) were calculated and interpreted as trivial ($0 \le ES \le 0.2$), small ($0.2 < ES \le 0.6$), moderate ($0.6 < ES \le 1.2$), large ($1.2 < ES \le 2.0$) and very large ($2.0 < ES \le 4.0$) (Cohen 1992; Hopkins 2002). It was assumed that training and match exposure time between players in injured and non-injured groups was largely similar. Based on historical data from the academy in question, we estimated that 50% of the players included in the present study would suffer a non-contact, football-related injury during the observation period. Given this estimation, a statistical power of 0.8 required a total sample size of $n = 80$ to detect a moderate effect (ES = 0.65) of FMS™ composite score between injured and uninjured players (G*Power Version 3.1, Kiel, Germany). Binomial logistic regression was used to examine the relationship between injury and potential risk factors including FMS™ composite score, each individual subtest score, number of asymmetries displayed during the test and age group. Spearman's rank correlation coefficient test was used to detect multicollinearity between independent variables. If two variables demonstrated a strong correlation ($r_s > .90$), then one was selected and the other not included in any further analysis. Each risk factor was examined independently via univariable analysis and those with a *P* value <.10 were investigated further in a multivariable model (Engebretsen et al. 2010). Assuming the aforementioned estimation of injury incidence, a statistical power of 0.8 required a sample size of $n = 82$ to detect an odds ratio of 2 (per unit increase for each independent variable) using logistic regression (G*Power Version 3.1, Kiel, Germany). Receiver operating characteristic (ROC) curves were used to assess the predictive value of the FMS™ composite score for each injury definition and identify cut-off scores that maximised specificity and sensitivity. The identified cut-off scores were used to create 2 × 2 contingency tables and calculate relative risks (RRs) with associated CI. Additionally, positive likelihood ratios were calculated to allow contextual appraisal of injury risk after a positive test. The limit for the alpha error to be considered significant was set at $P < 0.05$.

Results

Overall FMS™ scores for injured and uninjured players (using all definitions of injury) are presented in Table 1. Spearman's rank

Table 1. Comparison of FMS™ composite scores between injured and uninjured players for all injury definitions.

Type of injury	Injured	Uninjured	Effect size
All non-contact	15.8 ± 1.8 *n* = 38	15.3 ± 2.7 *n* = 46	0.22
Overuse	16.1 ± 1.8 *n* = 24	15.3 ± 1.9 *n* = 60	0.43
Severe	15.7 ± 1.8 *n* = 11	15.5 ± 1.9 *n* = 73	0.11

Data as mean ± SD.

Table 2. Univariable logistic regression analyses for each injury definition.

Type of injury	Variable	Odds ratio	95%CI	*P* value
All non-contact	FMS™ composite score	1.16	0.92–1.47	0.20
	Asymmetries (number)	0.90	0.53–1.54	0.70
	Age group	1.28	0.95–1.73	0.11
	Deep squat	0.75	0.31–1.82	0.53
	Hurdle step	0.94	0.43–2.08	0.89
	In-line lunge	1.58	0.61–4.10	0.35
	Shoulder mobility	1.68	0.87–3.28	0.13
	Active straight leg raise	1.41	0.66–3.01	0.38
	Trunk stability push-up	1.34	0.69–2.61	0.38
	Rotary stability	1.13	0.32–4.02	0.85
Overuse	FMS™ composite score	1.26	0.97–1.64	0.08
	Asymmetries (number)	1.04	0.58–1.86	0.90
	Age group	1.14	0.82–1.58	0.44
	Deep squat	1.15	0.44–3.00	0.78
	Hurdle step	1.25	0.52–3.00	0.61
	In-line lunge	1.43	0.51–3.98	0.49
	Shoulder mobility	2.10	0.93–4.76	0.08
	Active straight leg raise	1.77	0.76–4.14	0.19
	Trunk stability push-up	1.38	0.65–2.92	0.40
	Rotary stability	0.68	0.16–2.96	0.61
Severe	FMS™ composite score	1.06	0.76–1.48	0.74
	Asymmetries (number)	1.25	0.60–2.61	0.55
	Age group	0.94	0.61–1.46	0.79
	Deep squat	1.20	0.34–4.32	0.78
	Hurdle step	0.42	0.13–1.40	0.16
	In-line lunge	0.72	0.16–3.14	0.66
	Shoulder mobility	1.28	0.48–3.44	0.63
	Active straight leg raise	0.68	0.22–2.06	0.49
	Trunk stability push-up	2.04	0.66–6.36	0.22
	Rotary stability	2.63	0.51–13.67	0.25

CI: Confidence interval; FMS: functional movement screen.

correlation coefficients revealed no multicollinearity existed between any independent variable included in the logistic regression analyses. The results from the univariable logistic regression analyses are presented in Table 2. None of the predictor variables met the inclusion criteria for further investigation in a multivariable model for all non-contact and severe injuries. The composite score and shoulder mobility subtest did meet the inclusion criteria for further investigation when overuse injuries were considered. However, when included together in a multivariable analysis, no statistically significant relationships were observed. ROC curves for all non-contact (area under the curve [AUC] [95%CI]: 0.59 [0.47–0.72], $P = 0.14$), overuse (AUC [95%CI]: 0.63 [0.50–0.77], $P = 0.06$) and severe (AUC [95%CI]: 0.52 [0.34–0.70], $P = 0.84$) injuries revealed no statistically significant results. A cut-off score of ≤15 for any non-contact and severe injuries was identified while a threshold of ≤14 maximised specificity and sensitivity when considering overuse events. The positive likelihood ratios and RR values using the identified cut-off scores for each injury definition are presented in Table 3.

Table 3. Positive likelihood ratio and relative risk values using the identified FMS™ cut-off score for each injury definition.

	+ve LR	Chance of injury before screening (%)	Chance of injury after +ve screening (scoring ≤cut-off) (%)	RR (95%CI)
All non-contact	0.66	45	30	0.66 (0.40–1.10)
Overuse	0.71	29	20	0.70 (0.32–1.57)
Severe	1.28	13	17	1.52 (0.50–4.61)

+ve: Positive; CI: confidence interval; LR: likelihood ratio; RR: relative risk.

Discussion

The main findings of the present study revealed that in elite male youth football players competing in an English Premier League Academy: (1) there were no differences in FMS™ composite score observed between injured and uninjured groups regardless of injury definition used ($P > 0.05$ and trivial-small ESs) (Tables 1 and 2) and (2) no relationships with FMS™ score and non-contact, overuse or severe injuries existed; hence, the FMS™ had poor predictive ability.

FMS™ score is not related to injury in youth elite soccer players

No relationship between the FMS™ composite score and injury (all non-contact, overuse and severe) was observed. As the FMS composite score is made up from seven individual tests, some of which likely have greater relevance to football than others (e.g. the shoulder mobility vs. lower limb tests for outfield players), it was decided at the outset that possible relationships between injury and individual subtests would be investigated also. However, despite separating the FMS™ into its individual sub-sets, not only were no relationships found (for any injury definitions), but no statistically significant relationships were also observed between injury and any of the independent variables (including age group and FMS™ asymmetries).

What about predicting injuries?

Establishing a relationship between an attribute and injury is useful as it highlights a risk factor, which may in turn help inform the content of prevention strategies. However, predictive ability is even more appealing from a practical perspective (Bahr 2016). The most appropriate statistical measures that should be used to determine the predictive ability of a test include ROC curve analysis and likelihood ratios (Pepe et al. 2004; Opar et al. 2015; Bahr 2016; Whiteley 2016). A screening tool with excellent diagnostic accuracy would allow confident grouping of "at-risk" players who could subsequently be targeted with specific injury prevention interventions.

In the present study, ROC curve analysis revealed that the screening tool had poor predictive ability for any injury type (whether non-contact, overuse or severe). The AUC of an ROC curve provides an indication of the predictive ability of a diagnostic tool. An AUC = 0.5 indicates that a diagnostic tool has no predictive value while an AUC = 1 indicates a perfect test that results in no false positives or negatives (Hajian-Tilaki 2013). The ROC curves created in the current study produced AUCs between 0.52 and 0.63 depending on the injury definition used. These values are low and indicate that the FMS™ was likely not any better at predicting which players got injured than chance alone, i.e. randomly assigning players to high/low risk groups.

Further statistical calculations providing insight to the diagnostic accuracy of screening tools include specificity and sensitivity in addition to positive and negative predictive values. However, while these values are relevant, they are not as readily interpretable as a comparison of pre- and post-test odds of injury. Another relevant value for assessing the usefulness of a diagnostic tool is the positive likelihood ratio, which allows calculation of the post-test odds of injury after a positive test (an FMS™ score below the identified cut-off value) (Whiteley 2016). Likelihood ratios allow the calculation of these odds and offer practitioners clear information as to the usefulness of the screening tool in question. In the present study, positive likelihood ratios below a value of one indicated a reversal of the expected outcome and revealed a seemingly protective effect of scoring below the identified cut-off value (≤14 or ≤15 depending on injury type considered) on the FMS™ in the context of all non-contact and overuse injuries. Similarly, RR values ranged from 0.66 to 0.70 (Table 3) indicating a reduced likelihood of suffering any non-contact or overuse injury after scoring below the identified cut-off value. However, when considering this seemingly counterintuitive result, it is important to note that 95%CIs for the RR values crossed one in all instances. What is clear, however, is that among the present sample of elite male youth soccer players, an FMS™ composite score below the identified cut-off values was not associated with increased injury risk.

Does this mean we shouldn't use the FMS?

While the FMS™ may not be useful as a screening tool for highlighting elevated susceptibility to injury in elite male youth academy footballers, it does not necessarily render the screening tool completely useless. Indeed, it may provide other useful information. Its wide use among the world's top-league football clubs alludes to its appeal and perceived usefulness (McCall et al. 2014). Fuller et al. (2016) reported that young Australian rules' football players were 1.5 times more likely to report pain during the FMS™ if they had suffered an injury the previous season than if they had not. The authors postulated that the FMS™ might be useful for highlighting players who have not fully recovered from previous injuries.

The National Strength and Conditioning Association's recent position statement on long-term athlete development highlights the importance of structured youth strength and conditioning programmes focusing on aspects such as fundamental movement ability (Lloyd et al. 2016). In the context of Premier League youth academies, strength and conditioning practitioners and physiotherapists are faced with the challenge of a continual turnover of players each year. The FMS™ may offer a quick, logistically viable and systematic method of quantifying movement competency and in doing so help determine readiness for introduction to formalised resistance training and progressions to more advanced exercise techniques. This may be particularly helpful in guiding the physical development plans of newly recruited players whom club support staff are not familiar with.

Limitations

While the present study represents a novel addition to the literature regarding injury risk in elite male youth footballers, there are some limitations. First, exposure data were not available for the participants in the present study. This meant that no additional statistical procedures could be used (e.g. survival analysis or Cox proportional hazard modelling) which would have provided additional insight into the relationship between FMS™ score and injury (Bahr & Holme 2003; Finch & Marshall 2016). Differences in exposure time between injured and uninjured groups may have contributed to the findings. While it was assumed that exposure time between injured and uninjured groups was largely similar, this could not be empirically confirmed in the present study. It may be that players who achieved better FMS™ scores generally displayed superior overall athleticism and were selected to play more frequently. It must be stressed that such a hypothesis is purely speculation; however, it is one theory that may help explain the seemingly counter-intuitive results. The greater the exposure time, the greater the potential for suffering an injury. In addition, players performed strength and conditioning sessions throughout the observation season and such intervention may have mitigated the potential risk associated with scoring poorly on the FMS™. However, since all players were included in this aspect of training, the protective effect should have been equally apparent in all individuals regardless of FMS™ score. Another limitation of the present study is that multiple injuries to the same player were not taken into account. This has been highlighted as an issue that needs to be addressed to advance the value of such prospective cohort studies; however, the lack of exposure data once again precluded such survival analysis (Finch & Marshall 2016). Finally, these results may only be a reflection of the present team and future work using larger samples including multiple teams is necessary.

Conclusion

The present results question the efficacy of the FMS™ for highlighting young male elite football players at increased risk of injury. The FMS™ is not recommended for this purpose. Readers should be cognizant that this conclusion relates to the FMS™ specifically and does not necessarily apply to other movement screening tools. The FMS™ may be too generic, a test to highlight soccer-specific injury risk in male elite youth players. However, there may be other benefits to performing the assessment. For example, the FMS™ may help guide applied practitioners in the appropriate prescription of physical development programmes for large squads of players they are unfamiliar with. Future research should not only seek to add to this initial evidence for elite youth football players but also provide further insight through incorporation of exposure data and in doing so include multiple injuries to the same player within the statistical analysis.

Acknowledgements

The authors would like to acknowledge Terassa Taylor-Kaveney for her valuable input during the preparation phase of the study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Bahr R. 2016. Why screening tests to predict injury do not work—and probably never will…: a critical review. Br J Sports Med. 50:776–780.

Bahr R, Holme I. 2003. Risk factors for sports injuries – a methodological approach. Br J Sports Med. 37:384–392.

Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. 2016. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition – narrative review and new concept. Br J Sports Med. 50:1309–1314. [Epub ahead of print].

Cohen J. 1992. A power primer. Psychol Bull. 112:155–159.

Cook G, Burton L, Hoggenboom B. 2006b. Pre-participation screening: the use of functional movements as an assessment of function – part 2. N Am J Sports Phys Ther. 1:132–139.

Cook G, Burton L, Hoogenboom B. 2006a. Pre-participation screening: the use of fundamental movements as an assessment of function – part 1. N Am J Sports Phys Ther. 1:62–72.

Engebretsen AH, Myklebust G, Holme I, Engebretsen L, Bahr R. 2010. Intrinsic risk factors for hamstring injuries among male soccer players: A prospective cohort study. Am J Sports Med. 38:1147–1153.

Finch CF, Marshall SW. 2016. Let us stop throwing out the baby with the bathwater: towards better analysis of longitudinal injury data. Br J Sports Med. 50:712–715.

Fuller CW, Ekstrand J, Junge A, Andersen TE, Bahr R, Dvorak J, Hägglund M, McCrory P, Meeuwisse WH. 2006. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Clin J Sport Med. 16:97–106.

Fuller JT, Chalmers S, Debenedictis TA, Townsley S, Lynagh M, Gleeson C, Zacharia A, Thomson S, Magarey M. 2016. High prevalence of dysfunctional, asymmetrical, and painful movement in elite junior Australian football players assessed using the functional movement screen. J Sci Med Sport. [Epub ahead of print]..

Hajian-Tilaki K. 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 4:627–635.

Hopkins WG. 2002. A scale of magnitudes for effect statistics. A New view of Statistics. [cited 2016 Sep 05]. Available from: http://www.sportsci.org/resource/stats/effectmag.html

Le Gall F, Carling C, Reilly T. 2009. A comparison of injury in professional and non-professional male graduate youth soccer players. In: Reilly, T, Korkusuz, F, editors. Science and football VI: the proceedings of the sixth world congress on science and football. 2007 Jan 15-20. Antalya: Turkey.

Lloyd RS, Cronin JB, Faigenbaum AD, Haff GG, Howard R, Kraemer WJ, Micheli LJ, Myer GD, Oliver JL. 2016. National strength and conditioning association position statement on long-term athletic development. J Strength Cond Res. 30:1491–1509.

McCall A, Carling C, Davison M, Nedelec M, Le Gall F, Berthoin S, Dupont G. 2015. Injury risk factors, screening tests and preventative strategies: a systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. Br J Sports Med. 49:583–589.

McCall A, Carling C, Nedelec M, Davison M, Le Gall F, Berthoin S, Dupont G. 2014. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. Br J Sports Med. 48:1352–1357.

McCunn R, Aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. 2016. Reliability and association with injury of movement screens: a critical review. Sports Med. 46:763–781.

Moran RW, Schneiders AG, Major KM, Sullivan SJ. 2016. How reliable are functional movement screening scores? A systematic review of rater reliability. Br J Sports Med. 50:527–536.

Opar DA, Williams MD, Timmins RG, Hickey J, Duhig SJ, Shield AJ. 2015. Eccentric hamstring strength and hamstring injury risk in Australian footballers. Med Sci Sports Exerc. 47:857–865.

Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 159:882–890.

Pfirrmann D, Herbst M, Ingelfinger P, Simon P, Tug S. 2016. Analysis of injury incidences in male professional adult and elite youth soccer players: a systematic review. J Athl Train. 51:410–424.

Whiteley R. 2016. Screening and likelihood ratio infographic. Br J Sports Med. 50:837–838.

Zalai D, Panics G, Bobak P, Csáki I, Hamar P. 2015. Quality of functional movement patterns and injury examination in elite-level male professional football players. Acta Physiol Hung. 102:34–42.

IJSPT

ORIGINAL RESEARCH

THE INTRA- AND INTER-RATER RELIABILITY OF THE SOCCER INJURY MOVEMENT SCREEN (SIMS)

Robert McCunn[1]
Karen aus der Fünten[1]
Andrew Govus[2]
Ross Julian[1]
Jan Schimpchen[1]
Tim Meyer[1]

ABSTRACT

Background/purpose: The growing volume of movement screening research reveals a belief among practitioners and researchers alike that movement quality may have an association with injury risk. However, existing movement screening tools have not considered the sport-specific movement and injury patterns relevant to soccer. The present study introduces the Soccer Injury Movement Screen (SIMS), which has been designed specifically for use within soccer. Furthermore, the purpose of the present study was to assess the intra- and inter-rater reliability of the SIMS and determine its suitability for use in further research.

Methods: The study utilized a test-retest design to discern reliablility. Twenty-five (11 males, 14 females) healthy, recreationally active university students (age 25.5 ± 4.0 years, height 171 ± 9 cm, weight 64.7 ± 12.6 kg) agreed to participate. The SIMS contains five sub-tests: the anterior reach, single-leg deadlift, in-line lunge, single-leg hop for distance and tuck jump. Each movement was scored out of 10 points and summed to produce a composite score out of 50. The anterior reach and single-leg hop for distance were scored in real-time while the remaining tests were filmed and scored retrospectively. Three raters conducted the SIMS with each participant on three occasions separated by an average of three and a half days (minimum one day, maximum seven days). Rater 1 re-scored the filmed movements for all participants on all occasions six months later to establish the 'pure' intra-rater (intra-occasion) reliability for those movements.

Results: Intraclass correlation coefficient (ICC) values for intra- and inter-rater composite score reliability ranged from 0.66-0.72 and 0.79-0.86 respectively. Weighted kappa values representing the intra- and inter-rater reliability of the individual sub-tests ranged from 0.35-0.91 indicating fair to almost perfect agreement.

Conclusions: Establishing the reliability of the SIMS is a prerequisite for further research seeking to investigate the relationship between test score and subsequent injury. The present results indicate acceptable reliability for this purpose; however, room for further development of the intra-rater reliability exists for some of the individual sub-tests.

Keywords: Assessment, association football, kinematic, screening

Level of evidence: 2b

[1] Saarland University, Institute of Sport and Preventive Medicine, Saarbrücken, Germany
[2] University of Bedfordshire, Institute of Sport Science and Physical Activity, Bedford, United Kingdom

CORRESPONDING AUTHOR
Robert McCunn
Institute of Sport and Preventive Medicine
Saarland University
Campus Bldg. B8.2
66123 Saarbrücken, Germany
Phone: +49 681 302 70410
Fax: +49 (0) 681 302 4296
E-mail: bob.mccunn@me.com

INTRODUCTION

The proliferation of movement screening research and its widespread use in professional soccer reveals a belief among practitioners and researchers alike that movement quality may have an association with injury risk.[1, 2] Movement quality is ill defined but relates to the ability of an individual to perform a given movement in a controlled manner while demonstrating good or acceptable technique. Exactly what constitutes good technique remains a topic of debate. While it is arguable that no 'correct' movement pattern exists for any given exercise there are certain characteristics that may be undesirable, such as restricted range of motion and an inability to control coordinated movements. The rationale behind movement screening is that such limitations may result in acute injuries or contribute to insidious overuse complaints.[3-5]

Numerous screens exist; however, the supporting evidence with regard to both their reliability and association with injury varies widely in both volume and methodological quality.[1] The majority of such research has focused on the Functional Movement Screen (FMS™), which has demonstrated good reliability but conflicting relationships with injury likelihood.[1, 6] The FMS™ was designed as a 'general' movement assessment tool and has been used within a wide range of sports and professional domains including the military and emergency services.[7-9] In contrast, some screens such as the Landing Error Scoring System (LESS) have been designed with the intention of identifying those at an increased risk of a particular type of injury, for example, anterior cruciate ligament rupture.[10] In addition, some have been designed for use within particular sports, for example, netball and rugby union.[3, 11] Despite the popularity of movement screening within professional soccer, no soccer-specific tool currently exists.[2] The present study introduces the Soccer Injury Movement Screen (SIMS), which has been designed specifically for use with soccer athletes. The movements contained within the assessment were selected to reflect the most common sites (lower extremities) and types (strains and sprains) of soccer-related injury and hence they primarily tax the mobility and stability of the ankle, knee and hip joints in addition to the strength and flexibility of the surrounding musculature.[12] When selecting the individual sub-tests, priority was given to movements previously proposed within the scientific literature as potentially associated with injury likelihood.

The efficacy of screening tests that seek to identify or predict which players will get injured has recently been questioned.[13] In the context of sports-related injuries the idea that a single attribute such as movement quality for example, could be predictive is unlikely.[14] As a result, the ultimate objective of the SIMS will be to investigate whether a *causative relationship* exists between movement quality and injury. Any potential relationship between movement quality and injury is unlikely to be substantial enough to justify the SIMS being considered 'predictive' but it may help inform the content of injury prevention programs by highlighting risk factors.[15]

There is reason to expect that a causative relationship between movement quality and injury may exist since some authors have reported poor FMS™ scores preceding subsequent injury.[8, 16] However, numerous studies utilizing the same movement screening tool have not observed any link.[17] The SIMS may eventually demonstrate a stronger association to injury risk than the FMS™ due to its more explicit scoring criteria (Appendix 2) focusing on specific aspects of each movement. Furthermore, the FMS™ includes movements targeting the upper limbs, which have limited relevance for soccer players, whereas the SIMS concentrates on the lower limbs only.

Before any prospective cohort studies can be conducted using the SIMS its reliability must first be established. The reliability of an assessment tool is of critical importance since it is a pre-requisite for test validity.[18] Therefore, the purpose of the present study was to test the intra- and inter-rater reliability of the SIMS and determine its suitability for use in further research.

METHODS

Participants

Twenty-five (11 males, 14 females) healthy, recreationally active university students (age 25.5 ± 4.0 years, height 171 ± 9 cm, weight 64.7 ± 12.6 kg) agreed to participate in the present study. Inclusion criteria required participants to be aged between 18-40 years of age, free of injury (any physical condition that precluded them from completing the assessment) and recreationally active. Information pertaining to the study protocol and requirements were provided for each participant before written informed consent was collected. The study was approved by the local ethics committee (ref number: 270/15, Ärztekammer des Saarlandes, Saarbrücken, Germany) and conformed to the Declaration of Helsinki.

Raters

Three raters carried out the SIMS in the present study; all possessed postgraduate sport science qualifications and had previous professional experience delivering movement assessments. In addition, Rater 1 was an accredited strength and conditioning coach with both the United Kingdom Strength and Conditioning Association (UKSCA) and the National Strength and Conditioning Association (NSCA). Prior to the present study all raters conducted pilot testing using the SIMS with 10 participants. The pilot testing incorporated two 2-hour sessions where raters reviewed the test instructions (Appendix 1), the scoring criteria (Appendix 2) and familiarized themselves with the camera positioning (Figure 1). In addition, three more two-hour sessions were conducted where raters practiced scoring video footage and discussed the interpretation of the scoring criteria. In total, rater training amounted to ~12 hours (10 classroom-based and two field-based).

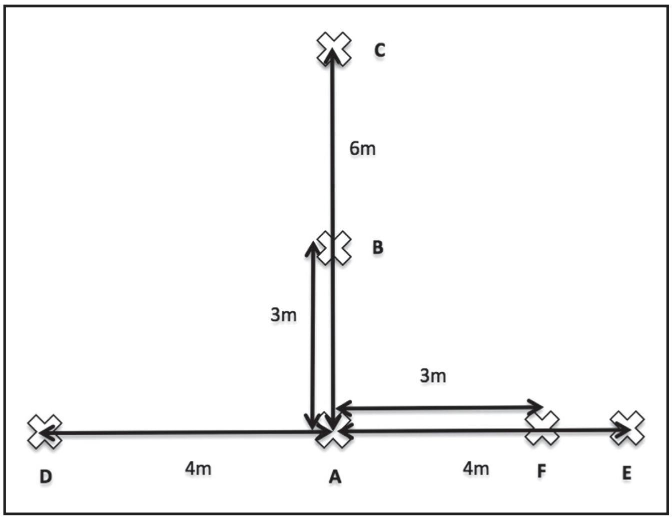


Figure 1. *Schematic showing the equipment setup for the SIMS*

For all movements the participants start at A. Anterior reach: measuring tape is fixed to the floor between A and B; Single-leg deadlift: camera at B (portrait) and E (landscape) when standing on right leg, camera at B (portrait) and D (landscape) when standing on left leg; In-line lunge: camera at B (portrait) and E (landscape) when right leg forward, camera at B (portrait) and D (landscape) when left leg forward; Single-leg hop for distance: measuring tape is fixed to the floor between A and C; Tuck jump: taped cross on floor at A (60x60cm), camera at B (portrait) and F (portrait).

Design

The present study utilized a test-retest design. Participants performed the SIMS on three occasions separated by an average of 3.5 days (minimum one day, maximum seven days). The SIMS contains five sub-tests: the anterior reach (AR), single-leg deadlift (SLDL), in-line lunge (ILL), single-leg hop for distance (SLHD) and tuck jump (TJ) (Figure 2). Raters 1 and 2 scored all participants whereas Rater 3 only scored 15 of the 25 (for reasons unrelated to the study). Raters scored two of the five movements (AR and SLHD) included in the SIMS in real-time on each occasion. The remaining three movements (SLDL, ILL and TJ) were filmed from both the frontal and sagittal planes using iPhone 4S devices (Apple Inc., California, USA) and scored retrospectively. These sub-tests were scored from video footage, as opposed to in real-time; to allow raters to view the movements in slow motion and increase the likelihood of identifying errors. A minimum of one week separated the scoring of participants' filmed movements for occasions one, two and three respectively in an attempt to reduce the risk of rater bias (i.e. remembering the previous scores given). Scores for occasions one, two and three were compared within each rater to investigate 'real-world' intra-rater (inter-occasion) reliability. Scores were also compared between raters for each occasion to assess inter-rater reliability. Rater 1 re-scored the filmed movements for all participants on all occasions six months later to establish the 'pure' intra-rater (intra-occasion) reliability for those movements.

Soccer Injury Movement Screen (SIMS)

Detailed descriptions of each movement contained within the SIMS and associated scoring criteria are outlined in Appendices 1 and 2. The ILL is the same in its setup as when performed as part of the FMS™ albeit it is scored differently, while the tuck jump is

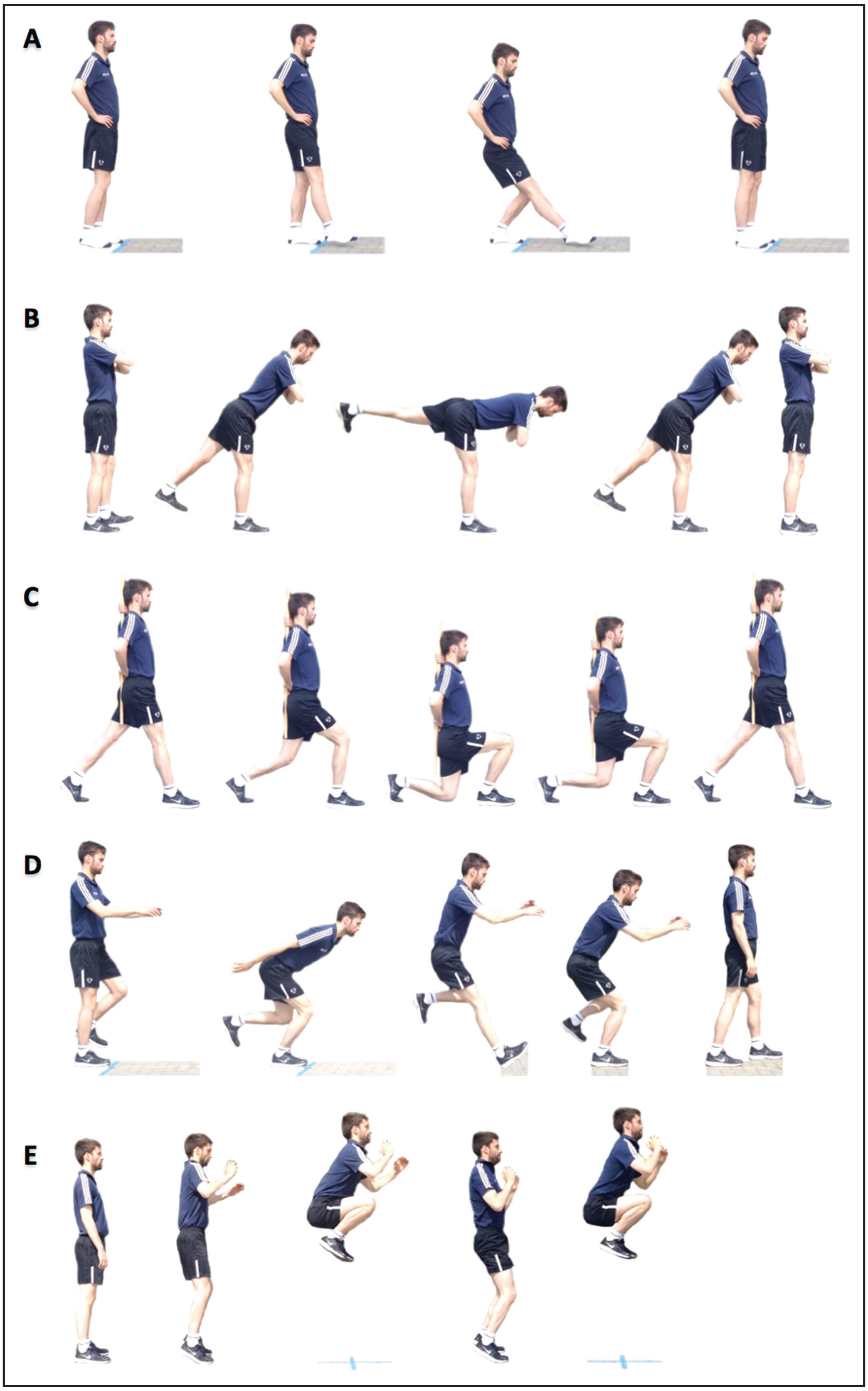


Figure 2. *Demonstration cards that were shown to participants along with verbal instructions prior to test execution*

A: anterior reach; B: single-leg deadlift; C: in-line lunge; D: single-leg hop for distance; E: tuck jump.

performed and scored exactly as described by Myer et al.[5, 19] A standardized five minute warm up was completed before each occasion and included dynamic bodyweight exercises (e.g. squats, walking lunges, hamstring walkouts, diagonal hop and holds). The assessments were performed outdoors on a hard, rubberized sports court during summertime in dry temperate weather conditions. Participants were instructed to wear tight fitting sports clothing and the same training shoes on each occasion. The five component movements were performed in sequential order starting with the AR followed by the SLDL, ILL, SLHD and TJ. Prior to each sub-test participants were read the test instructions (Appendix 1) verbatim and shown demonstration cards (Figure 2). Participants were then allowed three practice attempts for each sub-test where any obvious miscommunication or misunderstandings relating to how to execute the movements were clarified. Time to complete the assessment was 10-15 minutes per participant.

Each component movement was scored out of 10 points resulting in a theoretical maximum composite score of 50 when the score from each sub-test is summed. A higher score indicated poorer performance; hence, zero was the theoretical 'best' score while 50 was the 'worst'. The AR and SLHD scoring criteria were objective in nature and were based on reach and jump distance respectively. In contrast, the SLDL, ILL and TJ relied on subjective assessment of movement quality. Raters were allowed to watch the clips of the filmed movements, both in real-time speed and slow motion, as many times as they deemed necessary to make an accurate judgment when scoring.

Statistical analyses

Descriptive data are presented as means $\pm$ standard deviation. Reliability statistics are accompanied with 95% confidence intervals (CI). Data were analysed using R statistics program (R Core Development Team 2014) and MedCalc for Windows, version 16.4.3 (MedCalc Software, Ostend, Belgium). Comparison of composite and individual sub-test scores between male and female participants was performed using the Mann-Whitney U statistic. Cohen's *d* effect size (ES) was also calculated to compare male and female participants and was interpreted as follows: ≤0.2, trivial; 0.21-0.60, small; 0.61-1.2, moderate; 1.21-2.0, large; 2.1-4.0, very large.[20, 21] Two way mixed model intraclass correlation coefficients ($ICC_{3,1}$), weighted kappas (quadratic) and minimal detectable change (MDC) were used to determine the intra- and inter-rater reliability of the composite score. MDC values were calculated at both a 95% and 80% level of confidence in order to provide applied practitioners with the means to identify 'true' changes in test performance. Typically, MDC values are calculated to reflect a 95% confidence interval; however, this results in very conservative estimates of how much a test score has to change to be considered real and may be of limited usefulness in the applied setting where small improvements/decrements in test performance can be meaningful.[22] MDC values at lower levels of confidence (e.g. 80%) can be calculated and are useful to applied practitioners who may be willing to rely on more liberal estimates of test score changes. In addition, weighted kappas (quadratic) were used to determine intra- and inter-rater reliability of each individual subtest. ICC values were interpreted according to the following criteria: <0.40, poor; 0.40-0.59, fair; 0.60-0.74, good; ≥0.75, excellent.[23] Similarly, weighted kappa values were interpreted according to the guidelines outlined by Landis and Koch[24]: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect. Alpha was set at $p \leq 0.05$.

RESULTS

Composite scores were not significantly different between males (18.3) and females (15.3) (Table 1). Only the SLDL scores differed between genders (males = 4.3, females = 1.8) (Table 1).

$ICC_{3,1}$, weighted kappa and MDC values for intra-rater (inter-occasion) reliability are presented in Table 2. Weighted kappa values for the individual subtests ranged from fair to substantial (0.35-0.77). With regard to the composite score, weighted kappa values were interpreted as substantial (0.63-0.68) while the ICCs were classified as good (0.66-0.72) for each rater.

$ICC_{3,1}$ and weighted kappa values for inter-rater reliability are presented in Table 3. Weighted kappa values for the individual subtests ranged from mod-

Table 1. *Mean values (reported in arbitrary units) and comparison of test scores between males and females.*					
	Overall (n=25)	Males (n=11)	Females (n=14)	*p-value*	Male vs female effect size (qualitative inference)
Composite score (mean ± SD)	16.6 ± 4.9	18.3 ± 3.0	15.3 ± 5.8	0.080	0.6 (Small)
AR (mean ± SD)	1.7 ± 1.8	2.1 ± 2.3	1.4 ± 1.3	0.648	0.4 (Small)
SLDL (mean ± SD)	2.9 ± 2.1	4.3 ± 2.0	1.8 ± 1.5	<0.01	1.4 (Large)
ILL (mean ± SD)	2.6 ± 1.5	2.5 ± 1.5	2.6 ± 1.6	0.825	0.1 (Trivial)
SLHD (mean ± SD)	4.1 ± 2.3	4.2 ± 1.9	4.0 ± 2.7	0.718	0.1 (Trivial)
TJ (mean ± SD)	5.4 ± 1.3	5.2 ± 1.0	5.5 ± 1.6	0.534	0.2 (Trivial)

Test scores drawn from Rater 1 on the third testing occasion. AR= anterior reach, ILL= in-line lunge, SLDL= single-leg deadlift, SLHD= single-leg hop for distance, TJ= tuck jump

Table 2. *Summary of intra-rater (inter-occasion) reliability values. Values in brackets represent the 95% confidence intervals.*									
	Weighted kappa						$ICC_{3,1}$		
	AR	SLDL	ILL	SLHD	TJ	Composite score	Composite score	MDC @ 95% confidence	MDC @ 80% confidence
Rater 1	0.47 (0.17-0.77)	0.77 (0.67-0.87)	0.64 (0.52-0.77)	0.44 (0.26-0.61)	0.58 (0.43-0.73)	0.68 (0.54-0.81)	0.71 (0.52-0.85)	7.0	4.5
Rater 2	0.46 (0.22-0.69)	0.68 (0.55-0.81)	0.48 (0.30-0.66)	0.35 (0.15-0.55)	0.58 (0.44-0.72)	0.64 (0.49-0.80)	0.72 (0.54-0.85)	7.5	4.9
Rater 3	0.39 (0.02-0.77)	0.68 (0.55-0.81)	0.63 (0.49-0.77)	0.36 (0.11-0.61)	0.45 (0.26-0.65)	0.63 (0.45-0.80)	0.66 (0.38-0.86)	6.7	4.4

AR= anterior reach, ICC= intra-class correlation coefficient, ILL= in-line lunge, MDC= minimum detectable change, SLDL= single-leg deadlift, SLHD= single-leg hop for distance, TJ= tuck jump

Table 3. *Summary of inter-rater reliability values (between all three raters). Values in brackets represent the 95% confidence intervals.*							
	Weighted kappa						$ICC_{3,1}$
	AR	SLDL	ILL	SLHD	TJ	Composite score	Composite score
Occasion 1	0.83 (0.72-0.95)	0.51 (0.35-0.66)	0.71 (0.58-0.85)	0.84 (0.69-1.00)	0.60 (0.40-0.81)	0.78 (0.68-0.88)	0.79 (0.58-0.92)
Occasion 2	0.76 (0.62-0.90)	0.48 (0.29-0.66)	0.70 (0.56-0.84)	0.91 (0.85-0.97)	0.43 (0.18-0.68)	0.81 (0.71-0.90)	0.86 (0.70-0.95)
Occasion 3	0.59 (0.33-0.84)	0.64 (0.50-0.79)	0.58 (0.41-0.75)	0.91 (0.86-0.97)	0.50 (0.35-0.65)	0.79 (0.70-0.87)	0.79 (0.58-0.92)

AR= anterior reach, ICC= intra-class correlation coefficient, ILL= in-line lunge, SLDL= single-leg deadlift, SLHD= single-leg hop for distance, TJ =tuck jump

erate to almost perfect (0.43-0.91). With regard to the composite score weighted kappa values ranged from substantial to almost perfect (0.78-0.81) while the ICCs were classified as excellent (0.79-0.86) for all three occasions.

Weighted kappa scores for 'pure' intra-rater (intra-occasion) reliability are presented in Table 4. The kappa values were evaluated as almost perfect for the SLDL (0.90) and ILL (0.85) while the TJ value was interpreted as substantial (0.73).

DISCUSSION

Overall, the present results indicate sufficient reliability for the SIMS to be considered useful for further research and applied practitioners alike. The intra-rater reliability of the SIMS composite score was classed as substantial and good for all raters based upon the weighted kappa and ICC scores respectively (Table 2). The MDC values calculated

Table 4. *Summary of intra-rater (intra-occasion) reliability values for video-taped movements. Values in brackets represent the 95% confidence intervals.*			
	Weighted kappa		
	SLDL	ILL	TJ
Rater 1	0.90 (0.86-0.95)	0.85 (0.80-0.91)	0.73 (0.62-0.83)

ILL= in-line lunge, SLDL= single-leg deadlift, TJ= tuck jump at an 80% level of confidence demonstrate that if a one-point increase or decrease in each sub-test were observed a 'real' change in composite score would have likely occurred. The inter-rater reliability was classified as substantial to almost perfect when considering the weighted kappa values and excellent according to the ICCs (Table 3). The SLDL sub-test was the only movement where a discrepancy in scores between males and females was apparent (Table 1). Male participants regularly cited hamstring inflexibility as a limiting factor during this task whereas female participants rarely mentioned this. Females generally display superior hamstring flexibility as compared to men.[25] This difference in hamstring flexibility between males and females may potentially explain the gender difference in SLDL score observed in the present study.

The AR portion of the Y-balance test has previously been investigated as a risk factor with limb asymmetry >4 cm equating to a 2.3 – 2.7 times greater likelihood of non-contact injury among basketball and track and field athletes.[26, 27] The scoring criteria used in this assessment (Appendix 2) required the rater to assign a score (0 – 10) based on the difference in reach distance between limbs. The reason for limiting the scoring range to a maximum of 10 points (a reach asymmetry of ≥10 cm) was to maintain equal weighting between all five sub-tests (each of which was scored out of 10). The scoring criteria were clearly objective for this sub-test and therefore did not directly assess movement quality. However, it was decided that the AR warranted inclusion in the SIMS regardless of not directly assessing movement quality, due to the promising evidence surrounding its relationship to injury.[26, 27] The test reflects a number of physical qualities including neuromuscular control, strength and ankle stability: all of which are likely contributors to movement quality.[1, 26, 27] Therefore, while this sub-test did not assess movement quality directly the variable that was measured (difference in reach distance) is likely a reasonable surrogate marker. Ankle injuries occur frequently within soccer therefore the anterior reach may be a promising tool for highlighting increased risk of such events.[28] The intra-rater weighted kappa values for the AR ranged from fair to moderate (Table 2). In contrast, the inter-rater values ranged from moderate to almost perfect (Table 3). The difference between the intra- and inter-rater weighted kappa values suggests that the scoring criteria were clear but that a large proportion of the variation in the test scores stemmed from the participants and/or the influence of time between testing occasions. As such, additional participant familiarization with the test may help improve the intra-rater reliability.

While the SLDL is multifaceted in its demands, eccentric strength and flexibility of the hamstrings are clearly primary aspects of the movement due to the flexion of the hip with an extended knee on the standing leg. Both eccentric strength and flexibility of the hamstrings have been proffered as injury risk factors within soccer players.[29, 30] Hence, the ability to perform the SLDL with a high degree of movement quality may indicate proficiency in these important attributes (hamstring flexibility and eccentric strength). The intra-rater SLDL weighted kappa values for each rater represented substantial agreement (Table 2) while the inter-rater reliability values ranged from moderate to substantial (Table 3). These findings suggest that while raters were very consistent in their scoring of the SLDL within themselves there is opportunity for improvement in the between-rater agreement. Such a scenario is somewhat inevitable when considering subjective scoring criteria; however, more detailed guidelines on what constitutes a movement 'error' may help improve consensus between raters in the future.

The ILL, or split squat, is a widely used exercise within soccer both during warm-up routines and resistance training sessions.[31, 32] According to Cook et al.[34] the ILL focuses on the "stresses simulated during rotational, decelerating and lateral type movements". All of these movement patterns are frequently observed during soccer match play.[34] The ability to perform this exercise correctly is important to ensure players do not use compensatory movements that potentially cause or exacerbate acute and overuse injuries. When performing the ILL the same test setup was used as with the FMS™; however, the scoring criteria utilized in the current research (Appendix 2) differed.[33] The alternative scoring criteria were employed with the intention of explicitly outlining the potential movement flaws and hence enhancing clinical usefulness of the results. Both intra- and inter-rater reliability of the ILL ranged from moderate to substantial (Tables 2 & 3). The weighted kappa values reported in the present investigation are in keeping with those observed in studies of the FMS™ version of the ILL.[35-37] The more detailed scoring criteria adopted by the SIMS as compared with the FMS™ did not appear to adversely affect the reliability yet will provide practitioners with a clearer indication of where any potential movement dysfunction originates from.

It is important for soccer-specific movement assessments to incorporate explosive actions such as jumping and landing since they occur frequently during match play and often precede serious injury.[34, 38] While bilateral, vertical drop jumps have long been used for injury risk stratification[39, 40] many explosive soccer-specific actions are unilateral in nature and involve horizontal as well as vertical displacement (for example: kicking, changing direction and landing after a header).[34] The scoring criteria for the SLHD were objective and incorporated both the jump distance and the between limb difference in jump distance (Appendix 2) with each of these aspects weighted equally. The precise distances that characterized the different scoring ranges were based on pilot testing conducted with recreationally active university students and therefore may not be applicable to professional or youth soccer players. Revised criteria may need to be established for higher-level athletes. The authors opted for objective, as opposed to subjective, scoring in this instance due to recent evidence suggesting jump distance as a risk factor for non-contact hamstring injury.[41] While the intra-rater weighted kappa values ranged from fair to moderate the inter-rater values indicated almost perfect agreement between raters (Tables 2 & 3). The discrepancy between the intra- and inter-rater weighted kappa values suggests that a large proportion of the variation in the test scores stemmed from the participants and/or the influence of time between testing occasions rather than the application of the scoring criteria per se.

Allowing more jump attempts may increase the likelihood of maximum jump distance being reached and a plateau in performance occurring, which may in turn help improve reliability. On 32 of the 75 SLHD tests scored by Rater 1, (25 participants on three occasions) participants recorded their best jump distance (for that occasion) on their last attempt. Similarly, 15 of the 25 participants recorded their best jump distances overall on testing occasion 3. In addition, 12 of the 25 participants scored by Rater 1 recorded their best between limb difference score on their third testing occasion. This demonstrates that incorporating a number of familiarization sessions on multiple days prior to testing may improve reliability for the same reasons highlighted previously (plateauing of performance). However, it should be remembered that the more attempts allowed and the more familiarization sessions performed the greater the potential for fatigue to influence test performance and the less practically feasible the assessment may become. There may be a trade-off between improved reliability and the feasibility of using the SIMS as a screening tool in the applied environment. A recent systematic review by Hegedus et al.[42] assessed the methodological quality of studies exploring the reliability and validity of commonly used field-expedient screening tests such as the SLHD. They found no studies of satisfactory methodological quality reporting the reliability of the SLHD precluding comparison of the current results to previous findings.

The TJ assessment has been proposed as a field-expedient assessment of lower limb neuromuscular control.[19] It is unique as an assessment of movement quality since it requires the participant to continuously perform plyometric vertical jumps for 10 seconds.[19] While it is unlikely a player would replicate this precise activity during match-play the taxing nature of the test means it is likely to expose potentially injurious lower-limb movement patterns (particularly those associated with the onset of fatigue) that other, typically lower intensity assessments may not highlight. It has been suggested as a particularly useful tool for highlighting knee valgus movement during landing, which has been proposed as a risk factor for anterior cruciate ligament (ACL) injury.[19, 43] Considering the long-term sequelae associated with ACL injury the authors judged the TJ worthy of inclusion in the SIMS.[44, 45] Both the intra- and inter-rater weighted kappa values represented moderate agreement within and between raters (Tables 2 & 3). While this indicates acceptable reliability the weighted kappa values calculated are lower than previously reported by Myer et al.[19] However, Myer et al.[19] only assessed 10 participants and so raters may have remembered the previous scores given, leading to recall bias. In addition, they scored the same video footage twice as opposed to scoring participants on two separate occasions. The scoring criteria (Appendix 2) are inherently subjective but reliability may be improved by adding some objective guidelines to certain scoring items. For example, one of the scoring items asks: "was there a pause between jumps"? This could potentially be changed to: "was there a pause, lasting longer than one second (or another defined time period), between jumps"? Such amendments may improve consistency of scoring within and between raters. However, future research is needed to assess the difference in reliability when objective instructions are given compared with when they are not.

In an effort to separate some of the sources of variation within the test-retest design, one rater scored all the filmed movements (SLDL, ILL and TJ) from each testing occasion twice. This removed the influence of variation in test performance stemming from the participants and revealed the 'pure' intra-rater, or intra-occasion, reliability. The weighted kappa values for the SLDL and ILL represented almost perfect agreement while the score for the TJ indicated substantial reliability (Table 4). These higher weighted kappa values (as compared to those reported in Table 2) are not surprising since they reflect only the variation in scoring associated with the rater. These results suggest that improvements in the 'real world' intra-rater reliability are more likely to arise from aspects related to the participants rather than the raters. Bearing this in mind, future strategies aimed at improving the intra-rater reliability of the SIMS further may include extended participant familiarization with the test and allowing them to read the scoring criteria. Explicitly explaining the scoring criteria for the FMS™ to participants elicited improved scores.[46] This suggests that ambiguity related to what is being asked of participants during movement screening may influence their test execution and potentially contribute to variation in performance.

A number of limitations should be considered when interpreting the results of the present study. Perhaps most importantly, the pilot testing conducted to establish the scoring ranges for the SLHD (Appendix 2) were based on recreationally active university students' scores. As such, it may be necessary to revise this aspect of the scoring criteria in the future if the SIMS is used with professional soccer players. Similarly, if the SIMS were to be utilized with youth soccer players then amendments to the scoring criteria may be necessary. In addition, the results presented here are from only 25 participants, which, is a relatively modest sample size for assessing reliability according to Terwee et al[47]; however, the scores from three trials were included, rather than the usual two in an effort to improve the credibility of the conclusions. Furthermore, the raters represented a homogenous group. All were PhD students with postgraduate degrees in sport science. Further research may be needed to assess the reliability of the SIMS when conducted by other groups of raters, for example, undergraduate students or sports coaches.

CONCLUSIONS

Until now, no movement screen has been developed specifically for use among soccer players. The SIMS composite score demonstrated good to excellent intra- and inter-rater reliability. However, the intra-rater reliability of the individual sub-tests ranged from fair to substantial indicating scope for further improvement. Establishing the reliability of the SIMS is a prerequisite for further research seeking to investigate the relationship between test score and subsequent injury. The present results indicate at least acceptable reliability for this purpose.

REFERENCES

1. McCunn R, aus der Fünten K, Fullagar HHK, et al. Reliability and association with injury of movement screens: A critical review. *Sports Med*. 2016;46(6):763-781.
2. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. *Br J Sports Med*. 2014;48(18):1352-1357.
3. Reid DA, Vanweerd RJ, Larmer PJ, et al. The inter and intra rater reliability of the Netball Movement Screening Tool. *J Sci Med Sport*. 2015;18(3):353-357.
4. Padua DA, Marshall SW, Boling MC, et al. The Landing Error Scoring System (LESS) is a valid and

reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. *Am J Sports Med*. 2009;37(10):1996-2002.

5. Cook G, Burton L, Hoogenboom B. Pre-participation screening: The use of fundamental movements as an assessment of function – part 1. *N Am J Sports Phys Ther*. 2006;1(2):62-72.
6. Moran RW, Schneiders AG, Major KM, et al. How reliable are Functional Movement Screening scores? A systematic review of rater reliability. *Br J Sports Med*. 2016;50(9):527-536.
7. McGill S, Frost D, Lam T, et al. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. *Ergonomics*. 2015;58(10):1682-1689.
8. Lisman P, O'Connor FG, Deuster PA, et al. Functional movement screen and aerobic fitness predict injuries in military training. *Med Sci Sports Exerc*. 2013;45(4):636-643.
9. 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;2(3):147-158.
10. Padua DA, DiStefano LJ, Beutler AI, et al. The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. *J Athl Train*. 2015;50(6):589-595.
11. Parsonage JR, Williams RS, Rainer P, et al. Assessment of conditioning-specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. *J Strength Cond Res*. 2014;28(6):1497-1506.
12. Ekstrand J, Hägglund, Waldén M. Injury incidence and injury patterns in professional football: the UEFA injury study. *Br J Sports Med*. 2011;45(7):553-558.
13. Bahr R. Why screening tests to predict injury do not work – and probably never will...: A critical review. *Br J Sports Med*. 2016;50(13):776-780.
14. Cook C. Predicting future physical injury in sports: It's a complicated dynamic system. *Br J Sports Med*. 2016 [Epub ahead of print].
15. McCunn R, Meyer T. Screening for risk factors: If you liked it then you should have put a number on it. *Br J Sports Med*. 2016 [Epub ahead of print].
16. Garrison M, Westrick R, Johnson MR, et al. Association between the functional movement screen and injury development in college athletes. *Int J Sports Phys Ther*. 2015;10(1):21-28.
17. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. *J Sport Rehabil*. 2015;24(2):163-170.
18. Batterham AM, George KP. Reliability in evidence-based clinical practice: A primer for allied health professionals. *Phys Ther Sport*. 2003;4(3):122-128.
19. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. *Athl Ther Today*. 2008;13(5):39-44.
20. A new view of statistics: Effect magnitudes. http://www.sportsci.org/resource/stats/effectmag.html Published 2002. Updated August 7, 2006. Accessed August 1, 2016.
21. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159.
22. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.
23. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessment*. 1994;6(4):284-290.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
25. Marshall PW, Siegler JC. Lower hamstring extensibility in men compared to women is explained by differences in stretch tolerance. *BMC Musculoskelet Disord*. 2014;15:223.
26. Smith CA, Chimera NJ, Warren M. Association of y balance test reach asymmetry and injury in division I athletes. *Med Sci Sports Exerc*. 2015;47(1):136-141.
27. Plisky PJ, Rauh MJ, Kaminski TW, et al. Star excursion balance test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther*. 2006;36(12):911-919.
28. Walls RJ, Ross KA, Fraser EJ, et al. Football injuries of the ankle: A review of injury mechanisms, diagnosis and management. *World J Orthop*. 2016;7(1):8-19.
29. Timmins RG, Bourne MN, Shield AJ, et al. Short biceps femoris fascicles and eccentric knee flexor weakness increase the risk of hamstring injury in elite football (soccer): A prospective cohort study. *Br J Sports Med*. 2015 [Epub ahead of print].
30. van Beijsterveldt AM, van de Port IG, Vereijken AJ, et al. Risk factors for hamstring injuries in male soccer players: A systematic review of prospective studies. *Scand J Med Sci Sports*. 2013;23(3):253-262.
31. Bizzini M, Dvorak J. FIFA 11+: An effective programme to prevent football injuries in various player groups worldwide – a narrative review. *Br J Sports Med*. 2015;49(9):577-579.
32. Owen AL, Wong DP, Dellal A, et al. Effect of an injury prevention program on muscle injuries in elite professional soccer. *J Strength Cond Res*. 2013;27(12):3275-3285.

33. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: The use of fundamental movements as an assessment of function – part 1. *Int J Sports Phys Ther*. 2014;9(3):396-409.
34. Stølen T, Chamari K, Castagna C, et al. Physiology of soccer: An update. *Sports Med*. 2005;35(6):501-536.
35. Teyhen DS, Shaffer SW, Lorenson CL, et al. The Functional Movement Screen: A reliability study. *J Orthop Sports Phys Ther*. 2012;42(6):530-540.
36. Schneiders AG, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther*. 2011;6(2):75-82.
37. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. *J Strength Cond Res*. 2010;24(2):479-486.
38. Waldén M, Krosshaug T, Bjørneboe J, et al. Three distinct mechanisms predominate in non-contact anterior cruciate ligament injuries in male professional football players: A systematic video analysis of 39 cases. *Br J Sports Med*. 2015;49(22):1452-1460.
39. Nilstad A, Andersen TE, Kristianslund E, et al. Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening. *J Orthop Sports Phys Ther*. 2014;44(5):358-365.
40. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J Orthop Sports Phys Ther*. 2009;39(9):665-674.
41. Goossens L, Witvrouw E, Vanden Bossche L, et al. Lower eccentric hamstring strength and single leg hop for distance predict hamstring injury in PETE students. *Eur J Sport Sci*. 2015;15(5):436-442.
42. 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;49(10):642-648.
43. Ford KR, Myer GD, Hewett TE. Valgus knee motion during landing in high school female and male basketball players. *Med Sci Sports Exerc*. 2003;35(10):1745-1750.
44. Waldén M, Hägglund M, Magnusson H, et al. ACL injuries in men's professional football: A 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after ACL rupture. *Br J Sports Med*. 2016;50(12):744-750.
45. Øiestad BE, Engebretsen L, Storheim K, et al. Knee osteoarthritis after anterior cruciate ligament injury: A systematic review. *Am J Sports Med*. 2009;37(7):1434-1443.
46. 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;29(11):3037-3044.
47. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.

Appendix 1. *Description of the Soccer Injury Movement Screen (SIMS).*

Movement name	Rationale/perceived usefulness	Instructions
Pre-assessment	N/A	"For each exercise you have three practice attempts and three scored attempts on each leg. In the case of the tuck jump you have three practice jumps followed by the scored 10 second effort."
Anterior reach	- Provides an indication of ankle mobility (dorsiflexion) - Highlights limb asymmetry (ankle mobility and/or leg strength) - Provides an indication of single-leg control (e.g. motor control and balance)	"Remove your shoes. Place the big toe of your standing leg so it is touching the back of the taped line. Place hands on your hips. Reach the toes of the other leg as far along the measuring tape as possible – hovering around 5 centimeters off the ground. You must keep your standing foot in contact with the floor throughout, e.g. you cannot rise up on to your toes. Try to hover at the point of maximal reach for a couple of seconds to allow scoring. You must return to the start position for the attempt to be counted. Likewise, you must maintain balance throughout each attempt for the score to be recorded."*
Single-leg deadlift	- Provides an indication of ability to simultaneously flex and extend at the hip with extended knees while maintaining neutral spinal alignment - Provides an indication of hamstring flexibility - Provides an indication of single-leg control (e.g. motor control and balance)	"Put your shoes back on. Tuck your t-shirt into your shorts. Stand on the middle of the cross, taped on the floor, and cross arms over your chest. Imagine a straight line between your head and your right heel. Try to hinge at the hip while keeping that line straight until parallel to the floor. Try to keep your standing leg (left) extended. Return to the start position with both feet touching the floor between each repetition." Switch the words 'right' and 'left' when instructing the participant when testing the other side.

* If available, a slider device (e.g. Y Balance Test Kit™) can be used to perform the anterior reach.

Movement name	Rationale/perceived usefulness	Instructions
In-line lunge	- Provides an indication of ability to simultaneously flex and extend at the hip with flexed knees while maintaining neutral spinal alignment - Provides an indication of lower limb motor control and balance	As per instructions from Functional Movement Screen (Cook et al. 2006a) (see reference list for full article details). "Place your left toes so they are touching the back of the taped line. Place the heel of your right foot xx centimeters (as marked by instructor)** directly in front of your left foot. Hold the dowel behind your back gripping it with your left hand at your neck and your right hand at your lower back. Make sure the dowel is touching your head, upper back and buttocks. While maintaining an upright posture, descend into a lunge touching your left knee to the floor. Maintain contact with the dowel at the head, upper back and bum throughout. Return to the start position with knees fully extended between each repetition." Switch the words 'right' and 'left' when instructing the participant when testing the other side.
Single-leg hop for distance	- Provides an indication of lower-limb unilateral power - Highlights limb asymmetry (lower-limb power and/or ankle stability and/or lower-limb eccentric strength) - Provides an indication of single-leg control	"Place the toes of the jumping leg so they are touching the back of the taped line. Jump as far as you can while still able to stick the landing on the same leg and hold your position to allow measurement. You must record three successful scored jumps on each leg and you will receive as many attempts as necessary to achieve this."
Tuck jump	- Allows quick assessment of bilateral knee control during plyometric activity - Highlights limb asymmetry (lower-limb power and/or hip mobility)	As per instructions from Myer et al. (2008) (see reference list for full article details). "Stand on the middle of the cross taped on the floor with feet shoulder width apart. Upon signal from the tester, perform continuous vertical jumps on the spot for 10 seconds making sure to lift your knees towards your chest so that your upper thighs are parallel with the floor each time. Try to perform as many jumps as possible."

** Foot placement is determined by measuring the distance from the floor to the tibial tuberosity (shin length).

APPENDIX 2. SCORING CRITERIA

General rater instructions

Record each participant's height, weight and tibial tuberosity height (distance from the floor to their tibial tuberosity). If a participant cannot physically perform any test due to pain then they should be considered injured, this should be reported to the relevant club staff members and the test should be postponed.

Scoring guidelines for the anterior reach and single-leg hop for distance (objective assessments)

Anterior reach

Measure the distance (in centimeters) from the start line to the most distal part of the foot of the reaching leg. Round to the nearest centimeter. Three repetitions are performed on each leg and reach distance should be recorded for each attempt. The maximum reach distances achieved by each leg should be used to calculate the difference between left and right. The maximum theoretical score achievable is 10 and this would represent a 'poor' score. In contrast, the theoretical minimum score is zero and this would represent a 'good' score.

Difference in reach distance (cm) between legs	Test score
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
≥10	10

Single-leg hop for distance

Measure the distance (in centimeters) from the start line to the heel of the jumping/landing leg. Round to the nearest centimeter. Three repetitions are performed on each leg and jump distance should be recorded for each attempt. Both jump distance and limb symmetry are taken into account when assigning a test score. The maximum jump distance achieved on each leg should be summed and used to calculate the score. Combine the scores for jump distance and jump symmetry to produce the final score out of 10. The maximum theoretical score achievable is 10 and this would represent a 'poor' score. In contrast, the theoretical minimum score is zero and this would represent a 'good' score.

Sum of right and left best jump distances (cm): Males	Females	Test score
< 320	< 220	5
321-340	221-240	4
341-360	241-260	3
361-380	261-280	2
381-400	281-300	1
> 400	> 300	0

Difference between best right and left jumps (cm)	Test score
> 20	5
17-20	4
13-16	3
9-12	2
4-8	1
< 4	0

Scoring guidelines for the single-leg deadlift, in-line lunge and tuck jump (subjective assessments)

- If an error occurs once and the rater judges it to be egregious then it should be scored as an error.
- If an error (but only to a minor extent) is observed once then it should not be scored.
- If the same error (but only to a minor extent) is observed twice then it should be scored as an error.

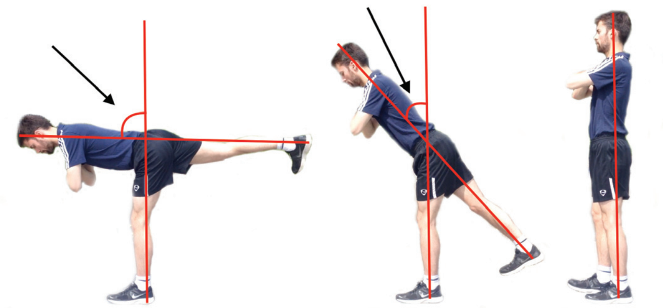
Defining specifically what constitutes "minor extent" or "egregious" is not possible. These judgments are left to the discretion of each individual rater. An important consideration is that raters are consistent in their judgments within themselves.

Single-leg deadlift

The score for this test is based on the 'movement quality' criteria outlined below. Three repetitions are performed on each leg. The maximum theoretical score achievable is 10 and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Both legs are scored and the average of both right and left scores is assigned to the individual.

Item #		
1	Is external hip rotation (standing leg) visible?	Yes = 1 No = 0
2	Does lumbar spine remain neutral?	Yes = 0 No = 1
3	Does thoracic spine remain neutral?	Yes = 0 No = 1
4	Does knee of raised leg remain extended throughout?	Yes = 0 No = 1
5	Is upper and lower body movement synchronized?	Yes = 0 No = 1
6	Is footprint maintained?	Yes = 0 No = 1
7	Is hip abduction (standing leg) present?	Yes = 1 No = 0
8	Does the standing leg knee remain extended throughout?	Yes = 0 No = 1
9	Parallel to floor position achieved? Parallel (90°) = 0, 89°-45° = 1, < 45° = 2 (all relative to the stance leg hip flexion angle)	

In relation to item #9 – the angle being assessed is displayed in the following diagram:



In-line lunge

The score for this test is based on the 'movement quality' criteria outlined below. Three repetitions are performed on each side. The maximum theoretical score achievable is eight and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Both legs are scored and the average of both right and left scores is assigned to the individual. To generate a score out of 10 multiply the fractional score out of eight by 10 e.g. if an individual displays four out of eight possible errors then the score out of 10 is: (4/8)x10 = 5. The reason for generating a score out of 10 is to maintain the same weighting between the five sub-tests.

Item #		
1	Does dowel remain vertical in frontal plane throughout?	Yes = 0 No = 1
2	Does torso rotation (transverse plane) occur?	Yes = 1 No = 0
3	Does dowel remain vertical in sagittal plane throughout?	Yes = 0 No = 1
4	Does back knee touch the floor?	Yes = 0 No = 1
5	Does heel of front foot lift off the floor?	Yes = 1 No = 0
6	Is footprint maintained throughout?	Yes = 0 No = 1
7	Are the three dowel contact points with body maintained?	Yes = 0 No = 1
8	Does knee valgus occur during the movement?	Yes = 1 No = 0

Tuck jump

Mark a cross on the floor using tape (two 60cm strips that intersect). The score for this test is based on the 'movement quality' criteria outlined below. The maximum theoretical score achievable is 10 and this would indicate 'poor' movement quality. In contrast, the theoretical minimum score is zero and this would indicate 'good' movement quality. Myer et al. (2008) created the tuck jump assessment and any further clarification on scoring procedures can be sought from their original article (see reference list for full article details).

Item #		
1	Was there knee valgus at landing?	Yes = 1 No = 0
2	Do thighs reach parallel (peak of jump)?	Yes = 0 No = 1
3	Were thighs equal side-to-side (during flight)?	Yes = 0 No = 1
4	Was foot placement shoulder width apart?	Yes = 0 No = 1
5	Was foot placement parallel (front to back)?	Yes = 0 No = 1
6	Was foot contact timing equal?	Yes = 0 No = 1
7	Was there excessive contact landing noise?	Yes = 1 No = 0
8	Was there a pause between jumps?	Yes = 1 No = 0
9	Did technique decline prior to 10 seconds?	Yes = 1 No = 0
10	Were landings in same footprint (within taped cross)?	Yes = 0 No = 1