Utility of FMS to understand injury incidence in sports: current perspectives

Meghan Warren1
Monica R Lininger1
Nicole J Chimera2
Craig A Smith1,3

1Department of Physical Therapy and Athletic Training, Northern Arizona University, Flagstaff, AZ, USA; 2Department of Athletic Training, Daemen College, Amherst, NY, USA; 3Smith Performance Center, Tuscon, AZ, USA

Abstract: The Functional Movement Screen (FMS) is a popular movement screen used by rehabilitation, as well as strength and conditioning, professionals. The FMS, like other movement screens, identifies movement dysfunction in those at risk of, but not currently experiencing, signs or symptoms of a musculoskeletal injury. Seven movement patterns comprise the FMS, which was designed to screen fundamental movement requiring a balance between stability and mobility. The 7 movement patterns are summed to a composite FMS score. For an instrument to have wide applicability and acceptability, there must be high levels of reliability, validity, and accuracy. The FMS is certainly a reliable tool, and can be consistently scored within and between raters. Although the FMS has high face and content validity, the criterion validity (discriminant and convergent) is low. Additionally, the FMS does not appear to be studying a single construct, challenging the use of the summed composite FMS score. The accuracy of the FMS in screening for injury is also suspect, with low sensitivity in almost all studies, although specificity is higher. Finally, within the FMS literature, the concepts of prediction and association are conflated, combined with flawed cohort studies, leading to questions about the efficacy of the FMS to screen for injury. Future research on the use of the FMS, either the composite score or the individual movement patterns, to screen for injury or injury risk in adequately powered, well-designed studies are required to determine if the FMS is appropriate for use as a movement screen.

Keywords: movement screen, prediction, sensitivity, athletes

Introduction

“Movement screen” has become a term used to define screening via movement analysis to identify dysfunction linked to increased risk of musculoskeletal injury or pathology. Screening detects disease or pathology in an individual that is not currently showing signs or symptoms of that disease.1,2 Movement screening has the same purpose as any other typical screen – to stratify those at increased injury risk from those not at risk. Stratification allows for targeted intervention in the “at-risk” group with the intent to lower injury occurrence.3 For example, Lamontagne et al4 evaluated squat performance in those with and without the diagnosis of hip impingement. Those with hip impingement did not squat as deeply compared with those without hip impingement (41.5%±12.5% of leg length vs 32.3±6.8%; P=0.037). From this, a cut-point to stratify those with hip impingement from those without may be developed and used for screening people who participate in activities requiring deep squat (eg, volleyball or powerlifting). A positive test (those that do not squat deep) would lead to further
testing, possible intervention to prevent injury, and limit future disability from a difficult to treat musculoskeletal pathology that may not initially present as painful or limiting.

Movement screens are developed using different methods including known pathomechanics and injury mechanisms, sport and fundamental movement requirements, and expert opinion. Three broad categories of movement screens have been reported in the literature: jump testing, balance with reaching tests, and movement quality against a standard criterion. Jump testing has been evaluated extensively to determine those at risk for anterior cruciate ligament rupture; balance with reaching tests, and movement quality compared against a standard criterion. Balance with reaching tests, such as the Star Excursion and Y Balance Test (YBT), have been used to associate poor performance with increased rate of injury in high school and collegiate athletes. Movement quality compared against a standard criterion has been used in multiple component assessment tools, including tasks such as squatting, balance, lunge, push-up, and arm movement. These include the foundation matrix, batteries of different movements, and the movement competency screen. Perhaps the most well-known of these type of movement screens, the Functional Movement Screen (FMS), is the focus of this manuscript.

The intent of this manuscript was not to thoroughly review all of the literature on the FMS; several excellent meta-analyses cover measurement aspects of the FMS in detail. Rather the purpose of this manuscript was to review measurement and psychometric indices of the FMS to assist researchers and clinicians with the interpretation of the evidence on FMS.

What is FMS?
The FMS is a commonly studied and used movement screen to assess injury risk in athletic and nonathletic populations. The FMS was designed to screen performance with fundamental movements, requiring a balance between stability and mobility while moving through a proximal to distal sequence. Seven individual movement patterns comprise the FMS, including deep squat, hurdle step, in-line lunge, shoulder mobility, active straight leg raise, trunk stability push-up, and rotatory stability. Each movement pattern is qualitatively assessed by a rater, and scored 3–0 based on degree of compensatory movements required to complete the movement or the presence of pain (3= ability to correctly complete the movement pattern without any predefined compensations; 2= performing the movement with any one of the movement pattern-specific compensations; 1= inability to perform the movement pattern; and 0= presence of pain during any portion of the movement pattern). The scores for the individual movement patterns are than summed to determine the overall composite FMS score.

The FMS is well represented in published literature. PubMed and CINAHL searches (“functional movement screen” or FMS and injury) were conducted for all papers from January 01, 2006 to February 02, 2018 (limited to English language and human studies), in addition to handsearching reference lists, and resulted in 118 primary (eg, prospective cohort study, experimental trials) and secondary studies (eg, meta-analysis, critically appraised topic). The year 2006 was selected as the beginning of the search as that is when the first papers that described the FMS were published. The research questions the authors posed in these papers covered a wide range, including normative values, differences in scores by several characteristics (eg, sex, body composition, skill level, reliability and validity, as well as prediction and association with injury). The FMS has been studied in diverse samples, from middle school students to middle-aged people, from professional athletes to military personnel to civil servants (eg, police and firefighters). Most of the primary (ie, no systematic reviews/meta-analyses) studies have been conducted using samples of collegiate/university athletes (n=22), military personnel (n=13), physically active people (n=14). FMS studies have been conducted in 16 countries, with the majority conducted in the United States (over 60%), but also in Europe (eg, Hungary, Ireland, Sweden), Africa (eg, South Africa), Asia (eg, Qatar, Israel, China), and the Americas (eg, Brazil, Canada, and Australia).

The FMS is also a popular movement screening tool that is utilized frequently in clinical settings. Although this is difficult to quantify empirically, FMS is part of the initial screening for the National Hockey League, national sport clubs internationally (eg, Polish National Handball team), and physical therapy and athletic performance clinics in the US and worldwide. Through www.functionalmovement.com (distributor of FMS), 22 training courses for FMS were offered in April 2018 around the world, in addition to online certification courses.

Several studies have established normative values for the summed composite FMS score in a variety of samples (Table 1). The average (mean or median) of the composite FMS score ranged between 13 and 16. Of these 11 studies, 5 also assessed normative values for the individual movement patterns, but caution is warranted with interpretation. Two of the studies reported mean and SD values for each movement pattern, despite an ordinal level of measurement. This cal-
calculation is potentially problematic because ordinal variables do not have the same properties as interval and ratio levels of measurement; median would be the appropriate measure of central tendency with variability expressed as interquartile range for ordinal variables. Two studies included graphs with the distribution of the movement pattern scores,29,30 and 1 listed the percent of scores for each movement pattern.22 Although several studies have been conducted reporting normative values, many of these studies are small (ie, only 4 with samples over 500 people)22–26 and were conducted in groups for which there are fewer studies about injury association/prediction (except for military). Therefore, establishing population-based norms in the athletic population is still an area that requires future study to better understand the FMS.

### Measurement properties for instruments

A screening tool that is useful in research as well as clinical settings must be both reliable and valid. Researchers have often referred to reliability by different terms such as reproducibility, consistency, and stability. In brief, reliability is defined as the degree to which a measurement method yields consistent results under identical conditions. Validity, on the other hand, refers to the degree to which a measurement method accurately measures what it is intended to measure. For the FMS, validity is typically assessed by comparing the scores to those of other movement quality assessments, such as the Shuttle Run Test (SRT). Reliability can be assessed using various statistical measures, including inter-rater reliability (IRR) and intraclass correlation coefficient (ICC). However, it is important to note that these measures may be affected by the sample size and the number of raters involved. Therefore, future studies should aim to include a larger sample size and use multiple raters to enhance the reliability of the FMS.

### Table 1 Normative values of the FMS

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Sample description</th>
<th>Age: mean ± SD or range (years)</th>
<th>FMS score: mean ± SD</th>
<th>Movement pattern scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abraham et al22</td>
<td>1,005</td>
<td>Adolescents</td>
<td>10–17</td>
<td>All: 14.6±2.5</td>
<td>DS: 14.9±2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 14.2±2.2</td>
</tr>
<tr>
<td>Bardenett et al25</td>
<td>167</td>
<td>High school athletes</td>
<td>13–18</td>
<td>All: 14.6±2.5</td>
<td>DS: 13.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 13.1</td>
</tr>
<tr>
<td>Smith et al31</td>
<td>94</td>
<td>High school male athletes</td>
<td>15.5±1.2</td>
<td>All: 15.5±1.2</td>
<td>DS: 2.0±0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.3</td>
</tr>
<tr>
<td>Marques et al29</td>
<td>103</td>
<td>Elite soccer players</td>
<td>14–20</td>
<td>All: 15.4±2.4</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.3</td>
</tr>
<tr>
<td>Agresta et al32</td>
<td>45</td>
<td>Healthy distance runners</td>
<td>22–54</td>
<td>All: 15.5±1.9</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>Loudon et al28</td>
<td>43</td>
<td>Running athletes</td>
<td>22.2±3.0</td>
<td>All: 15.4±2.4</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>Fox et al27</td>
<td>62</td>
<td>Male Gaelic field sport athletes</td>
<td>22.2±3.0</td>
<td>All: 15.5±1.9</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>Schneiders et al30</td>
<td>209</td>
<td>Active population</td>
<td>21.9±3.7</td>
<td>All: 15.7±1.9</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18–40</td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>de la Motte et al24</td>
<td>1,037</td>
<td>Military applicants</td>
<td>20.8±3.0</td>
<td>All: 14.7±1.8</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>Orr et al24</td>
<td>1,512</td>
<td>Australian state police force</td>
<td>28.0±8.0</td>
<td>All: 15.2±2.4</td>
<td>DS: 2.0±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS: 2.0±0.5</td>
</tr>
<tr>
<td>Perry and Koehle23</td>
<td>622</td>
<td>Middle-aged adults</td>
<td>50.9±10.8</td>
<td>All: 14.1±2.9</td>
<td>DS: 2.0±0.5</td>
</tr>
</tbody>
</table>

**Abbreviations:** FMS, Functional Movement Screen; ♂, male; ♀, female; DS, deep squat; HS, hurdle step; ILL, in-line lunge; SM, shoulder mobility; SLR, active straight leg raise; PU, trunk stability push-up; RS, rotatory stability.
as agreement, replication, consistency. However, from a purely psychometric perspective, reliability is the degree to which the measurement is free from error.85 The measurement is the dependent or outcome variable that is collected. Error can be from 2 difference sources: systematic change (bias) and random error. Systematic error is when the scores change in a similar orientation between repeated measures.85 Random error is from mechanical, biological, or protocol differences that are unpredictable.85 Reliability can be estimated over time (test–retest), for internal consistency (interrater, Cronbach α), and for equivalence (alternative forms).

Validity is the degree to which the instrument measures the construct intended.86 While there are many forms of construct validity (Figure 1), this manuscript will focus on face and content (translation validity) along with convergent and discriminant (criterion validity). Translation validity centers on the true meaning of the construct (face and content validity), whereas criterion is the degree of relationship (association or prediction) between the tool and an external source (convergent and discriminant validity).

Accuracy (validity) can be calculated in diagnostic studies using sensitivity and specificity. As a review, sensitivity is the ability of a tool to correctly identify an individual when he/she truly has the condition. Specificity is the ability of a test to correctly identify those who do not have the condition of interest. It is important to understand that a reliable instrument is not necessarily a valid tool. An instrument may come to the same result repeatedly, but that result may not be accurate. Therefore, reliability is a necessary but not sufficient condition for validity.87

Reliability

Seventeen studies have been published that have assessed the reliability of the FMS.16–18,30,41,61,88–97 Of these, 1 is a systematic review published in 201697 and 2 are systematic reviews with meta-analyses, published in 201667 and 2017.18 These studies have been conducted in experienced and nonexperienced raters, using physical therapists, athletic trainers, strength and conditioning coaches, and undergraduate and graduate students as raters. These studies have assessed real time as well as video recordings of the FMS performance. Overall, these studies conclude the FMS can be scored consistently, both within (intra-rater) and between (interrater) scorers. In our opinion, the research question of the reliability of the FMS has been thoroughly answered, and this is an area that requires no future study.

Validity

Construct validity

The scoring for the FMS includes summing each of the individual movement pattern scores to a single composite construct to quantify movement dysfunction.8 Although there may be benefit for the ease of interpretation and even though the total score has been used extensively in research, evidence does not appear to support a unidimensional construct, or single summed composite score, for the FMS.62,80,98 Four factor analyses42,62,80,98 assessed the constructs with FMS. In a 2014 exploratory factor analysis using data from 934 Marine Officer candidates, Kazman et al62 failed to identify a unidimensional structure for FMS. Reasons for this proposed to be heterogeneity with the compensations for each individual movement pattern, the inclusion of pain as a specific scoring criterion, and different facets being measured with each individual movement pattern. These results of no evidence of a unidimensional construct were replicated in 2015 using data from 290 elite athletes from Chinese national teams.80 In 2016, 2 factor analyses were completed. Koehle et al98 completed exploratory and confirmatory factor analyses using retrospective data from 1,113 individuals. These data resulted in 2 factors, a basic movement factor, with shoulder mobility and active straight leg raise, and a complex movement factor, including deep squat, hurdle step, in-line lunge, and the trunk stability push-up, with rotary stability loaded onto both factors. This 2-factor model was tested in a confirmatory factor analysis including data from 176 division 1 student-athletes.42 Support for a 1-factor model was found, but also stated that the 1 factor model performed similarly to the 2-factor model. In summary, 1 out of the 4 factors analyses found support for
a 1-factor model, or a single summed composite score. One study found that a 2-factor model performed best, with the factors revealed that were similar to the original description of the FMS. Finally, 2 of the 4 studies could not determine a construct and concluded that each individual movement pattern should be assessed independently.

Translation validity

Face validity
A screening tool that has a high level of face validity will measure the construct of interest. This is a subjective assessment of construct validity, but is often reported by researchers. For the FMS, face validity has been demonstrated in 2 ways. As stated before, this is a very popular movement screening instrument with over 20 certification courses in 1 month alone all over the world. Additionally, since 2006, there have been 118 papers published that used the FMS. This wide acceptance of the FMS as a movement screening tool in research and clinically also provides evidence for face validity.

Content validity
Content validity measures how well an instrument captures all possible facets of a construct. This is normally determined through expert opinion of those in the field of interest. Content validity of the FMS was demonstrated because the developer of the FMS, Gray Cook is a board-certified orthopedic physical therapy clinical specialist. In addition, the initial papers describing the FMS included Lee Burton, who is an athletic trainer, and Barb Hogenboom, who is a board-certified sports physical therapist and athletic trainer. The combined expertise in movement analysis from these authors bolsters the content validity of FMS.

Criterion validity
The criterion validity of the FMS has been assessed in a single study. Whiteside et al compared the FMS scoring to prespecified kinematic thresholds for 6 of the 7 FMS movement patterns (shoulder mobility excluded) in 11 collegiate athletes. The participants completed the 6 tests of the FMS while assessed by a rater, as well as a kinematic motion capture system. The rater used the standard 4-level scoring for the FMS, while the kinematic analysis used specific criteria for scoring. For example, in the in-line lunge movement pattern for the FMS, 1 of the criteria for the standard scoring was “dowel contact maintained” during the movement pattern. The kinematic equivalent of this was “range of distance from the top hand origin to C7 and of distance from bottom hand to sacrum both <5 cm, across the entire exercise.” Poor agreement was found between the 2 methods. Using the example above, the percent agreement was 50.0% with a κ of 0.20. Evidence for criterion validity is low, and the authors recommended caution with using the FMS and previously published scoring criteria.

Discriminant validity
Discriminant validity is the ability of a tool to differentiate between 2 people, for example those in 1 group compared to another. A cut-point score from the composite FMS of 14 was published in the first study that used FMS to better understand the value for identifying athletes at a higher risk of injury. Although 14 is not used as the cut-point for every study using the FMS, it is utilized most often (Table 2). Beach et al and Frost et al compared small samples with high (>14) and low (<14) composite FMS scores with back loading mechanics, and lumbar movement and frontal knee motion. In a sample of 15 high and low scorers, Beach et al found no association with group and peak low back loading magnitudes with lifting. Frost et al reported less motion with the high scoring group, but substantial interindividual variability with lumbar flexion, extension, lateral side bending, rotation, and frontal knee motion. The authors further concluded, “current FMS scoring criteria may be insensitive to potentially risky movement (eg, uncontrolled frontal plane knee motion) behavior.” Only 1 study found different lower extremity mechanics by the FMS scoring criteria, but only studied 1 movement pattern – the deep squat. In a sample of 28 individuals, 9 scored 1 on the deep squat, 9 scored 2, and 10 scored 3. Significantly greater peak ankle dorsiflexion and/or dorsiflexion excursion, knee flexion and sagittal plane knee joint excursion, and hip flexion and sagittal plane hip joint excursion, as well as greater knee and hip extension moments were found in the individuals who scored a 3 compared with those who scored lower. Despite this positive finding, the preponderance of findings does not lend support for discriminant validity of the FMS, although additional study of individual movement patterns is needed.

Convergent validity
Often thought of as an opposite form of discriminant validity, convergent validity assesses if 2 instruments are measuring the same underlying construct. This is commonly represented with a high correlation between 2 different tools or constructs. Six research articles assessed the relationship between the FMS (or a portion of the test) between the FMS and another assessment tool. Teyhen et al evaluated the
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Injury definition</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collegiate athletes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garrison et al49</td>
<td>Multiple sports</td>
<td>Med att +1 day time loss</td>
<td>0.67</td>
<td>0.73</td>
<td>NR</td>
</tr>
<tr>
<td>Warren et al50</td>
<td>Multiple sports</td>
<td>Med att</td>
<td>0.54</td>
<td>0.46</td>
<td>0.48</td>
</tr>
<tr>
<td>Dorrel et al59</td>
<td>Multiple sports</td>
<td>Med + time loss</td>
<td>0.61 Value (95% CI)</td>
<td>0.49 (0.41–0.57)</td>
<td>0.56 (0.49–0.63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe (3 week time loss)</td>
<td>0.65 (0.43–0.81)</td>
<td>0.45 (0.39–0.51)</td>
<td>0.53 (0.41–0.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Musculoskeletal</td>
<td>0.62 (0.52–0.70)</td>
<td>0.49 (0.41–0.57)</td>
<td>0.54 (0.47–0.61)</td>
</tr>
<tr>
<td>Mokha et al45</td>
<td>Male soccer, and female rowing, soccer, volleyball</td>
<td>Med att +1 day time loss</td>
<td>0.26</td>
<td>0.59</td>
<td>0.36</td>
</tr>
<tr>
<td>Bond et al44</td>
<td>Male and female basketball</td>
<td>Med att</td>
<td>0.14</td>
<td>0.86</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1–9 day time loss</td>
<td>0.17</td>
<td>0.87</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10+ day time loss</td>
<td>0.28</td>
<td>0.88</td>
<td>0.43</td>
</tr>
<tr>
<td>Chorba et al51</td>
<td>Female basketball, soccer, volleyball</td>
<td>Med att</td>
<td>0.58</td>
<td>0.74</td>
<td>NR</td>
</tr>
<tr>
<td>Hotta et al47</td>
<td>Male runners</td>
<td>Med att +4 week time loss</td>
<td>0.73</td>
<td>0.46</td>
<td>0.65</td>
</tr>
<tr>
<td>Clay et al46</td>
<td>Female rowers</td>
<td>Med att +1 day time loss</td>
<td>0.29</td>
<td>0.92</td>
<td>NR</td>
</tr>
<tr>
<td>Walbright et al52</td>
<td>Female basketball and volleyball</td>
<td>Med att +1 day time loss</td>
<td>0.18–1.00</td>
<td>0–0.88</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Professional athletes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kiesel et al39</td>
<td>American football</td>
<td>Med att +1 day time loss</td>
<td>0.54</td>
<td>0.91</td>
<td>NR</td>
</tr>
<tr>
<td>Kiesel et al110</td>
<td>American football</td>
<td>Time loss</td>
<td>0.27</td>
<td>0.87</td>
<td>NR</td>
</tr>
<tr>
<td>Tee522</td>
<td>Rugby union</td>
<td>&gt;28 day time loss</td>
<td>0.62</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Other athletes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bardenett et al55</td>
<td>High school athletes – multiple sports</td>
<td>Med att +1 day time loss</td>
<td>0.56</td>
<td>0.38</td>
<td>0.50</td>
</tr>
<tr>
<td>Chalmers et al46</td>
<td>Elite Junior Australian football athletes</td>
<td>Med att +1 day time loss</td>
<td>0.65</td>
<td>0.36</td>
<td>0.51</td>
</tr>
<tr>
<td>Duke et al58</td>
<td>Experienced male rugby union athletes</td>
<td>Any</td>
<td>0.54</td>
<td>0.95</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First half of the season</td>
<td>0.36</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second half of the season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dossa et al74</td>
<td>Elite junior hockey</td>
<td>Med att</td>
<td>0.50</td>
<td>0.70</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Military</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosio-Lima et al55</td>
<td>Coast Guard Maritime Security Response Team candidates</td>
<td>Med att</td>
<td>0.40</td>
<td>0.86</td>
<td>NR</td>
</tr>
<tr>
<td>O’Connor et al55</td>
<td>Marine officer candidates</td>
<td>Any</td>
<td>0.45</td>
<td>0.78</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serious</td>
<td>0.12</td>
<td>0.94</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overuse</td>
<td>0.13</td>
<td>0.90</td>
<td>0.58</td>
</tr>
<tr>
<td>Knapik et al66</td>
<td>Coast Guard recruits</td>
<td>Med att</td>
<td>Men: 0.55</td>
<td>Men: 0.49</td>
<td>Men: 0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women: 0.60</td>
<td>Women: 0.61</td>
<td>Women: 0.59</td>
<td></td>
</tr>
<tr>
<td>Everard et al64</td>
<td>Military recruits</td>
<td>1 day time loss</td>
<td>0.23</td>
<td>0.77</td>
<td>0.43</td>
</tr>
<tr>
<td>Kodesh et al57</td>
<td>Female soldiers</td>
<td>2 day time loss</td>
<td>0.42</td>
<td>0.63</td>
<td>0.51</td>
</tr>
<tr>
<td>Bushman et al58</td>
<td>Light infantry brigade combat soldiers</td>
<td>Any</td>
<td>0.33</td>
<td>0.82</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Traumatic</td>
<td>0.28</td>
<td>0.77</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overuse</td>
<td>0.37</td>
<td>0.81</td>
<td>0.61</td>
</tr>
<tr>
<td>Bushman et al69</td>
<td>Light infantry brigade combat soldiers</td>
<td>Any</td>
<td>0.08–0.22</td>
<td>0.90–0.99</td>
<td>0.52–0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Traumatic</td>
<td>0.02–0.19</td>
<td>0.87–0.98</td>
<td>0.50–0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overuse</td>
<td>0.03–0.24</td>
<td>0.09–0.99</td>
<td>0.51–0.58</td>
</tr>
<tr>
<td><strong>Other samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McGill et al122</td>
<td>Police officers</td>
<td>Back injury without known mechanism</td>
<td>0.28</td>
<td>0.76</td>
<td>NR</td>
</tr>
<tr>
<td>Peate et al33</td>
<td>Firefighters</td>
<td>Any injury</td>
<td>0.36</td>
<td>0.71</td>
<td>NR</td>
</tr>
<tr>
<td>Butleret al24</td>
<td>Firefighters</td>
<td>3 day time loss</td>
<td>0.84</td>
<td>0.62</td>
<td>NR</td>
</tr>
<tr>
<td>Shojaedin et al55</td>
<td>University students</td>
<td>Any injury</td>
<td>0.51</td>
<td>0.83</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knee injury</td>
<td>0.14</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ankle injury</td>
<td>0.13</td>
<td>0.95</td>
<td></td>
</tr>
</tbody>
</table>

Notes: aCut-point of <15 vs >15 used. bCut-point of <13 vs >13 used. cCut-point of <14.5 vs >14.5 used.

Abbreviations: AUC, area under the curve; CI, confidence interval; FMS, Functional Movement Screen; NR, not reported; Med att, medically attended.
relationship between the FMS and the YBT in active duty service members finding that a better performance on the FMS was associated with a greater anterior reach on the YBT. Bond et al.\textsuperscript{44} and Paszkewicz et al.\textsuperscript{107} compared the FMS to a mobility test. Bond et al.\textsuperscript{44} suggested that while the newly constructed Basketball-Specific Mobility Test and the FMS may be appropriate for detecting musculoskeletal deficiencies, this screen and test do not accurately determine injury risk in basketball players at the collegiate setting. Two systematic reviews have been published containing convergent validity measures.\textsuperscript{108,109} Whittaker et al.\textsuperscript{109} proposed that only 4 of the 17 included articles assessed the relationships between risk factors. Finally, McCall et al.\textsuperscript{108} reviewed the “Top 3” risk factors, screening tests, and preventative exercises used in premier football (soccer) athletes. The FMS, psychological questionnaires, and isokinetic testing all received a “D” (insufficient evidence to assign a specific recommendation) rating for level of evidence. Due to the limited number of rigorous studies assessing the relationship of the FMS with another instrument, convergent validity is low.

**Accuracy**

**Sensitivity and specificity**

The FMS has been studied extensively related to how well the test can identify those at a risk for injury. Although most researchers report both sensitivity (true positive rate) and specificity (true negative rate), for the association between FMS composite score and injury, the more important metric is sensitivity. High sensitivity means a high number of true positives relative to the number of false negatives (ie, those at a risk for injury for whom the FMS did not identify them at a risk). High specificity means a high number of true negatives relative to the number of false positives (ie, those identified at a risk of injury by the FMS who in actuality did not get injured). In the case of injury, we want to identify as many people as possible who are at risk of injury to work to minimize that risk with injury prevention efforts. If those who are not at a risk get identified at a risk (false-positive), there may be an overuse of resources (ie, putting people in injury prevention programs they may not need), but missing those at a risk of injury (false negatives) could have more serious consequences.

Overall, the sensitivity of the FMS is low (Table 2). For any medically attended injury, the sensitivity ranged from 0.33 to 0.58; when time loss was included in the injury definition, the sensitivity was higher generally, except for 4 studies with very low sensitivity (0.26 and 0.29 in 2 studies of collegiate athletes),\textsuperscript{45,46} 0.27 in professional football players,\textsuperscript{110} and military recruits.\textsuperscript{44} The specificity was much higher, with 15/23 studies reporting specificity higher than 0.70. This means there was a low false-positive rate relative to the true negative, and the FMS may more correctly classify people who did not sustain an injury.

**Relationship between FMS and injury**

By far, the area of largest study for FMS is in the association/prediction of musculoskeletal injury with some facet of the FMS. This includes the composite score, individual movement patterns, asymmetries in movement patterns, or pain with movement patterns. Since 2007, 39 papers have been published on this topic. In addition, 3 systematic reviews and meta-analyses,\textsuperscript{18–20} 2 systematic reviews,\textsuperscript{108,109} and 1 review\textsuperscript{111} have also been published. The most comprehensive systematic review and meta-analysis including 24 studies focusing on the FMS composite score and injury association was published in March 2017; therefore, another comprehensive review of the literature is not needed and readers are referred to Moran et al.\textsuperscript{20} In conclusion, this systematic review stated, “the level of evidence for strength of association between FMS composite score and subsequent injury is not sufficient to support the use of FMS composite score as an injury prediction tool.”\textsuperscript{20} It is important to note several things about this systematic review. First, of the 24 studies reviewed, only 25% (n=6) received a grade of “good” for quality assessment tool, 8% (n=2) received a grade of “acceptable,” and the remaining 67% were graded as “low” quality. Additionally, only 3 studies conducted with military/police samples were sufficiently similar to be combined for meta-analysis. Finally, the assessment of the use of FMS as an “injury prediction tool” was made using odd ratios (ORs), relative risks, positive likelihood ratio, focusing on statistical significance; only in 2 cases was area under the curve (AUC) considered. Since this systematic review and meta-analysis was published, another 9 studies have been published examining the association between the FMS composite score and injury.\textsuperscript{44,46,58,59,64,65,72,112,113} Two of the studies in rugby union players reported a significant association;\textsuperscript{58,72} the other 7 supported the conclusion of Moran et al.\textsuperscript{20} Tee et al\textsuperscript{72} found significantly lower FMS composite score in professional rugby union players who had a subsequent injury compared to those who did not sustain an injury. Duke et al\textsuperscript{58} reported significant OR between FMS composite score ≤14 vs >14 and any injury for the first half (OR =10.42 [95% confidence interval {CI}: 1.28–84.75]) and second half (OR =4.97 [95% CI: 1.02–24.19]) of the season.
To most accurately interpret this literature however, a distinction must be made about prediction and causation. Since 2007, there have been 36 studies published using either the FMS composite score, individual movement patterns, or pain with FMS movement patterns to better understand risk profiles of the sample. Of these 36, 27 use the word “prediction” in the title or purpose statement. Unfortunately, many of these papers present ORs as the effect measure and focus on the statistical performance of that effect measure, as was illustrated above with Moran et al. A smaller (n=9 studies), but more recent, systematic review and meta-analysis on FMS composite score and injury included a table titled, “Injury Predictive Value Studies.” This table only listed OR with 95% CI and focused the discussion on statistical significance only. OR can provide a measure of association between FMS score and injury, but cannot discriminate between who will and will not have an outcome (ie, injury). 114

AUC and Akaike Information Criterion are some of the more frequently used metrics for prediction. Of these, Akaike Information Criterion was not used in any of the previously published papers that mentioned prediction. AUC is used to discriminate between 2 disease states (eg, injured or not injured) and is defined as the probability that given 2 participants, one who will develop a disease and the other who will not, the model will assign a higher probability of disease to the former. AUC is used to estimate the risk of an adverse outcome (eg, injury) based on a risk profile (eg, FMS composite score). The AUC for the studies on FMS composite score and injury is very low (Table 2). For many studies, the AUC is <0.50, meaning that those with lower FMS scores may, in fact, have a lower probability for injury. Only 2 studies, Tee et al 27 (AUC =0.73) and Hotta et al 47 (AUC =0.65), may be considered predictive of injury. Additionally, 13 of the studies summarized in Table 2 did not include AUC despite many of them having prediction in the title or purpose statement. Therefore the conclusion from Moran et al 20 is correct, but the low predictive ability should be interpreted from sensitivity, specificity, and AUC rather than OR from the previously published studies.

**Future directions for the FMS and injury**

Although normative data exists for FMS; population-based normative data in athletic populations is lacking. Given the popularity of the FMS as a movement analysis tool in athletic and physically active populations, greater study is needed on larger sport-specific samples to create normative data sets. To date, the FMS as a single construct has been evaluated; however, the construct and criterion validity for all of the movement patterns has yet to be investigated. Further, asymmetries in movement patterns may be more associated with injury than the FMS composite score; this area deserves additional study in a variety of sport-related populations. Many of the previously published prospective cohort studies with the composite FMS scores have been low quality. In addition to the items cited in the previous review, these observational studies have been flawed by unadjusted analyses only with poor statistical methodology, small sample size with no a priori sample size or post hoc power calculations in the absence of nonsignificant results, and no discussion about possible effect modification. Rigorous, adequately powered studies that fully adjust confounding are required to fully understand the association of FMS movement patterns and/ or asymmetries with injury. Prediction studies also require sufficient sample size and control of all other variables that may predict injuries (eg, sex, previous injury, skill level). These studies may be sport-specific or encompass a variety of sports.

Future studies on the clinical applicability of the FMS are also needed. A recent item analysis of the FMS suggested that the FMS can possibly be classified into 2 categories – screening for flexibility and evaluation of motor performance. We recently demonstrated that FMS composite score and movement patterns were affected by lower extremity range of motion and core function. While the FMS may identify movement dysfunction, the clinical application and implementation of targeted exercise program progression to improve FMS performance needs additional study to ensure consistency among practitioners. Additionally, FMS movement patterns may be affected by understanding or knowledge of how to perform the specific movement; thus, the FMS may not accurately represent movement dysfunction. Future research should discern between actual and implied movement dysfunction when using the FMS as a screening tool; this would provide greater clinical relevance to using the FMS in clinical settings.

**Conclusion**

The FMS is a popular movement screening tool that can be consistently scored by raters with varied background and training. The FMS also has good translation validity with low criterion validity and sensitivity. Despite the popularity of FMS to describe injury risk, the concepts of prediction and association are conflated in the literature. There is limited evidence for success in injury prediction; the association with injury is more mixed. Much of the previous work on
the relationship between FMS and injury is flawed by methodological limitations. Adequately powered, well-designed, prospective cohort studies are required to allow final determination of FMS as a movement screen to identify those who are higher risk of injury and need for intervention.

**Author contributions**
Dr Warren takes responsibility for the integrity of the information contained in this paper. The study was conceived and designed by all the authors. All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

**Disclosure**
The authors report no conflicts of interest in this work.

**References**


