Determination of the Functional Movement Screen to Predict Musculoskeletal Injury in Intercollegiate Athletics

Brian William Wiese, MS, ATC, LAT; Jonathan Keith Boone, MS, ATC, FMS; Carl Gino Mattacola, PhD, ATC; Patrick Owen McKeon, PhD, ATC, CSCS; and Timothy Lee Uhl, PhD, ATC, PT

ABSTRACT

Functional movement insufficiencies may place an individual at increased risk of injury. The Functional Movement Screen (FMS) is used to identify movement deviations. Limited research has been conducted to verify the predictive value of the FMS test to identify injury risk. Our purpose was to investigate the predictive usefulness of the FMS. FMS scores of 144 National Collegiate Athletic Association Division I football athletes were obtained during preseason. Participants were tracked prospectively over the course of a single season and all sport-related injuries were evaluated, documented, and tracked by an athletic trainer. A cutoff score was determined using a receiver operator characteristic curve. A maximized odds ratio of 1.425 (95% confidence interval: 0.6-3.2) and a positive likelihood ratio of 1.154 were found for individuals that scored a 17 or below on the FMS. Our results are contrary to previous research and indicate that the FMS is not useful for predicting musculoskeletal injury risk. [Athletic Training & Sports Health Care. 2014;6(4):161-169.]

Participation in sports comes with a variable risk for injury, with sports involving contact having a higher risk. According to the National Collegiate Athletic Association (NCAA) Injury Surveillance System, more than 50% of injuries in football occur to the lower extremity.1 The standard precautionary measure to determine any preexisting condition that may lead to injury is the preparticipation physical examination. The preparticipation physical examination includes obtaining information about the athlete to reduce the chances of severe injury and death by identifying predisposing physical factors, identifying disqualifying conditions, and recommending rehabilitative measures or counseling to direct the athlete to an appropriate sport.2 However, not all injuries can be predicted via preparticipation physical examinations. Preparticipation physical examinations may fall short in the identification and prevention of injuries caused by insufficient functional movement. A functional movement is the ability to produce and maintain a balance between mobility and stability along the kinetic chain while performing fundamental patterns with accuracy and efficiency.3 Identifying movement deviations can be critical in identifying an individual’s risk of injury. One method of identifying movement deviations is the Functional Movement Screen (FMS).

Little consideration is given to functional movement deficits, which may limit performance and predispose the individual to microtraumatic or macrotraumatic injury.4 The FMS is a battery of fundamental movement patterns that are assessed qualitatively by a trained examiner to identify the presence or absence of movement deviations or asymmetries.5 The FMS is designed to identify right- and left-side imbalances, as well as the inability to use mobility and stability through observing movement patterns.3 These imbalances and asymmetries can cause the utilization of compensatory movement patterns in the kinetic chain to achieve movement goals.4 The FMS has been found to be a reliable test,6-14 as well as an efficient test.15 The FMS has been proposed as an injury-screening tool,
and at this time there is limited evidence for its effectiveness. Based on the work of several studies, a score of 14 or below has been reported to be the point where injury rates differ the most among those who are injured and those who are uninjured over a competitive season.\textsuperscript{5,16,17} These studies indicate a 4- to 11-fold greater chance of injury if the score is below or equal to a 14, compared with a score of 15 or greater.\textsuperscript{5,16} These previous studies have some limitations in their interpretation due to methodological and statistical choices.

One study evaluated odds ratios instead of positive likelihood ratios. Likelihood ratios are often considered the best statistic for determining the usefulness of a diagnostic test or diagnostic accuracy.\textsuperscript{18,19} Likelihood ratios use sensitivity and specificity to evaluate the probability of injury occurrence, whereas odds ratios use exposures and outcomes to create a measure between the two.\textsuperscript{18} A positive likelihood ratio indicates the increase in odds favoring the condition given a positive test result, thus indicating the shift in probability favoring the condition when the test is positive.\textsuperscript{18} For example, there would be a shift in probability for injury occurrence (the condition) when an individual tests and has a score $\leq 14$ (positive test), if 14 is in fact the cutoff score used to determine differences in injury risk. Diagnostic tests measured on a continuous scale are frequently transformed into multilevel ordinal outcomes based on a cutoff score, where the likelihood ratio can be calculated for each level of the test.\textsuperscript{18} Because the FMS is a test scored from 0 to 21, we can calculate likelihood ratios for each score (or level) of the test to determine the optimal cutoff score. However, an odds ratio is a measure of association between an exposure and an outcome and represents the odds that an outcome will occur, given a particular exposure compared with the odds of the outcome occurring in the absence of that exposure.\textsuperscript{20} Odds ratios take into consideration exposures (FMS testing) and outcomes (those injured and uninjured throughout the season) to create a ratio of true outcomes to false outcomes. In the current study, the cutoff score was established via a receiver operator characteristic (ROC) curve. For a given cutoff score, the true comparison would be anyone who was injured to anyone who was not injured. If the odds ratio equates to 2, then it can be interpreted that there were 2 times more true outcomes to false outcomes.

One challenge with interpreting the data using FMS to predict injury is that the cutoff score is often determined based on previous research versus being calculated with the sample of interest. For example, studies have elected to apply the cutoff score established by the previous literature that studied National Football League (NFL) football players and apply it to Division III female athletes.\textsuperscript{5,16} A limitation with this methodological choice is that the samples are different; therefore, the cutoff score may not apply to the sample being examined where if an ROC curve analysis was run a cutoff score could have been established for the current sample of interest. Another study acknowledged that the cutoff score of 14 should be used with caution because it did not represent the athletic population that was targeted in the study.\textsuperscript{9} Those authors went on to suggest that further studies need to be conducted on other athletic and occupational groups before determining a substantiated cutoff value.\textsuperscript{6} Currently, only 1 research study has performed an ROC curve analysis to determine a cutoff score specific to the study sample.\textsuperscript{21} Therefore, determining FMS cutoff scores for diverse samples is needed to understand its utility.

Prospective evaluation of the FMS for injury prediction is limited and is limited to smaller samples within the sport of football. Large samples have been investigated using the military population and other service professionals.\textsuperscript{22-24} A larger sample size for a football population would allow for greater statistical power and, hopefully, a more definitive cutoff score to determine differing levels of injury risk than previous research. To determine the predictive value of the FMS, we tracked the injuries of NCAA Division I intercollegiate football players.

Our study is unique for several reasons. We justify a cutoff score by completing an ROC analysis. This will more accurately describe our sample by evaluating the sample’s FMS scores and injuries instead of applying a cutoff score that was established by different samples. No study, to our knowledge, has investigated specific injury types and the usefulness of the FMS to determine its predictive utility for the injury types. In addition, the injuries found in the sample will be stratified by upper extremity, lower extremity, overuse, noncontact, and a loss of time of $>10$ days. The purpose of the current study was to determine the usefulness of the FMS to predict injury in collegiate football players, as well as each substratification of the sample.
METHOD

Study Design
The study design was a prospective cohort study. The independent variable was groups (injured and noninjured), whereas the dependent variable was the composite FMS scores. Injury evaluations were performed by the investigators of the study, as well as by the staff athletic trainers responsible for the care of the football team. If necessary, some injuries were confirmed through diagnostic imaging and evaluation from the team’s orthopedic doctors. Injuries were collected for the sample as a whole and were further stratified for type of injury by upper extremity injury, lower extremity injury, overuse injury, noncontact injury, and injury resulting in a loss of >10 days. Upper extremity injury consisted of any pathology to the shoulder, upper arm, elbow, forearm, wrist, hand, and digits. Lower extremity injury consisted of any pathology to the hip, groin, thigh, knee, lower leg, ankle, foot, and digits. Overuse injury consisted of any tendinopathy, muscle spasm or tightness, and soreness that occurred to any part of the body. Noncontact injury consisted of any injury that was caused by a noncontact mechanism to any part of the body. Injury resulting in a loss of >10 days consisted of any injury to any part of the body. Recurrent injuries were not recorded, as the initial injury already placed the individual in the specific injury substratification.

Participants
The study included 144 NCAA Division I football players. No participants were dropped from the football team; therefore, there was no participant loss for this study. The participants’ average age was 18.9 (±1.3) years, average height was 73.7 (±2.7) inches, and average weight was 225.2 (±43.8) pounds. All participants were on the official team roster by the beginning of preseason and were medically cleared to participate. Participants were excluded if they had any orthopedic surgery within the past 3 months, an injury resulting in a time loss >1 week within the past 6 weeks, vestibular injury or illness, or signs or symptoms of a concussion or postconcussion syndrome. Participants signed informed consent forms, and the study received approval by the University of Kentucky Medical Institutional Review Board. All participants were tested prior to their designated weight lifting time on the day of testing, with no warm-up. To our knowledge, participants had no previous experience with the FMS and were given detailed instructions prior to each movement pattern in the FMS.

Instrumentation
The study used the FMS test kit (Functional Movement Systems, Danville, Virginia) to administer FMS testing. All testing was completed in the E.J. Nutter Athletic Training Room on the campus of the University of Kentucky. Sportsware 2011 Injury Surveillance Software (Computer Sports Medicine Inc, Stoughton, Massachusetts) and Microsoft Excel (Microsoft Corporation, Redmond, Washington) were used to track and record injuries. Microsoft Excel spreadsheets were also used to record participant descriptive data, FMS scores, and sort injury data. Statistical Package for the Social Sciences (SPSS) (IBM Corporation, Somers, New York) was used for statistical analysis.

Procedures
The FMS was administered at the beginning of the 2010 regular football season and the 2012 regular football season. No football player who was tested in the 2010 season was used during testing in the 2012 season. Testing procedures were administered by 2 certified athletic trainers (J.K.B., B.W.W.) who had 2 years and 1.5 years of experience, respectively, with the FMS prior to data collection. Neither assessor at the time of testing had formal instruction on the FMS or held FMS testing credentials. Both assessors were initially instructed by clinicians who were using the FMS in the workplace. Each assessor had performed more than 150 assessments each for the study and for clinical practice, with 25% of the assessment being completed prior to testing. Intertester reliability was not able to be assessed because 1 investigator took a new job; therefore, we have no interrater reliability estimates. However, results from previous FMS reliability investigations suggest that raters with similar experience with the FMS have good-to-excellent interrater reliability.7-14,22 Participants wore shorts and a t-shirt that would not restrict any movement during testing. A thorough explanation of the FMS was provided to each participant. The investigator explained each individual test of the FMS to each participant before completion of the functional movement.

Functional Movement Screen
The FMS is used to quantitatively assess the quality of functional movement patterns. The FMS con-
sists of 7 exercises, scored on an ordinal scale from 0 (lowest) to 3 (highest), and 3 clearing examinations, scored either positive for producing pain or negative for failing to produce pain. The highest score one can achieve on the FMS is a 21. Each participant was given 3 attempts at each exercise. If the tests were performed bilaterally, then each testing side was given 3 attempts. The mode of scores was taken for each test. For example, on any test, if the participant received a score of 2, 2, and 3 for their 3 attempts, respectively, then the score accepted for that test would be a 2. We chose to pick the mode of the scores rather than select the lowest score because we thought that it better described the individual’s functional ability. We thought the most consistent score, for example, the one that occurred two-thirds of the time, is more indicative of the participants’ representative score on the test, rather than the score that appears one-third of the time. Calculation of an average is not consistent with the scoring of the FMS because the FMS scoring scale uses whole integers. However unlikely, if a participant was to score a 1, 2, and 3 on an individual test, the score of 2 would be taken, as it was the average and median of the scores. However, because a score of 1, 2, and 3 never occurred in this study, taking the average was never necessary. To calculate an overall FMS score, all scores from the individual tests are totaled. In tests that were performed bilaterally, the lowest score of either side is used for recording. For example, on the hurdle step-over test, if the left leg received a score of 3 and the right leg received a score of 2, the lowest score was used for that test to total the FMS score. If tests that had clearing examinations, such as shoulder mobility, rotary stability, and trunk stability push up, produced pain in participants, the participant received a score of 0 on the associated test. For instance, if a participant scored a 3 on the trunk stability push up but had pain with the spinal extension clearing examination, the score of 3 was omitted and resulted in a score of 0. All testing procedures were done in accordance with the protocol described by Cook et al.4,25

### Injury Tracking

Injuries were tracked over 1 season (this included pre-season, as well as regular season practice and games, and any postseason or championship games or practices). Injuries were tracked only for practices and games. Injuries acquired through conditioning, lifting, or activities of daily living were not included. Injury evaluations were performed by the investigators of the study, as well as by the staff athletic trainers responsible for the care of the football team. If necessary, some injuries were confirmed through diagnostic imaging and evaluation from team orthopedic doctors. A breakdown of injured participants for the sample and each stratification of the sample can be found in Table 1. All recurrent injuries were omitted from Table 1 because the initial injury already placed that participant in the injured group. Software being used for injury reporting was Sportware 2011, in addition to injury surveillance using a Microsoft Excel spreadsheet. Injury was defined as a musculoskeletal injury that (1) occurred as a result of participation in an organized intercollegiate practice or competition, (2) requires medical attention by a certified athletic trainer or physician, and (3) results in restriction of participation or performance for 1 or more calendar days beyond the day of injury. This definition fits the injury requirements, as established according to the NCAA Injury Surveillance System.1

### Data Reduction

Data from all FMS tests were calculated from each individual test of the FMS to form a composite FMS score, using the clearance examinations to negate individual tests when necessary. The composite scores were the dependent variable. The FMS scores were compared with

<table>
<thead>
<tr>
<th>SAMPLE STRATIFICATION</th>
<th>INJURED</th>
<th>HEALTHY</th>
</tr>
</thead>
<tbody>
<tr>
<td>All injuries</td>
<td>93</td>
<td>51</td>
</tr>
<tr>
<td>Upper extremity injuries</td>
<td>34</td>
<td>110</td>
</tr>
<tr>
<td>Lower extremity injuries</td>
<td>75</td>
<td>69</td>
</tr>
<tr>
<td>Overuse injuries</td>
<td>56</td>
<td>88</td>
</tr>
<tr>
<td>Noncontact injuries</td>
<td>60</td>
<td>84</td>
</tr>
<tr>
<td>Injuries with &gt;10 day loss of time</td>
<td>29</td>
<td>115</td>
</tr>
</tbody>
</table>
the independent variable, groups for the entire sample, and each substratification of the sample.

Statistical Analysis
We used the ROC curve analysis\textsuperscript{26} to determine differences between the injured and uninjured groups for the entire sample and for each substratification of the sample. Individuals who were injured were defined as the state variable, compared with those who went uninjured throughout the season. The optimal point on the curve was determined when the true–positive rate (sensitivity) was maximized and the false–positive rate (1-specificity) was minimized, equating to the point with the highest positive likelihood ratio.\textsuperscript{26} This point was determined to be the cutoff score for the sample. A separate analysis was conducted for each substratification to obtain a specific cutoff score for that stratification of the sample. A \textit{P} value for the positive likelihood ratios was calculated using a nonparametric Mann–Whitney analysis. Odds ratios were calculated using a 2×2 contingency table as seen in Table 2, with the formula \((a/c) / (b/d)\). The odds ratio \textit{P} value was calculated using a Pearson’s uncorrected chi-square test. The odds ratios, \textit{P} values, and raw numbers used for the odds analysis can be found in Table 3. The odds ratio for the sample and for each substratification of the sample was calculated at the optimal cutoff score that was determined by the ROC curve analysis. The cutoff score is the cumulative FMS score found by ROC curve analysis to be the point of greatest difference between injured and uninjured individuals. Injured participants were included in potentially more than 1 injury stratification if they fit the injury description. For example, an individual with a noncontact, overuse, lower extremity injury would be included as injured for the whole sample, as well as 3 stratifications of the sample.

RESULTS
The average FMS score was 16.1 (±1.9). No significant differences were noted in the FMS scores between injured and uninjured participants for the entire sample (\textit{P} = .854). For the entire sample, a cutoff score was maximized at 17, with a true–positive rate and false–positive rate of .495 and .492, respectively. A positive likelihood ratio of 1.154 and an odds ratio of 1.425 (95% confidence interval [CI], 0.6-3.2) were found at the cutoff score of 17. Figure 1 depicts the ROC curve for the entire sample. Injuries were further stratified due to the large sample size for upper extremity injury, lower extremity injury, overuse injury, noncontact injury, and injuries causing a loss of time of \(>10\) days; the ROC curves for each of the stratifications can be found in Figures 2-6. No difference in FMS scores was found between the injured and uninjured participants for all sample stratifications. Table 4 shows the predictive values for the entire sample, as well as each substratification for the determined cutoff score.

DISCUSSION
Our study investigated a sample of NCAA Division I intercollegiate football players and considered 5 substratifications of injuries. This sport included contact and noncontact injuries. Our average FMS score was 16.1 and was closely associated with previous findings.\textsuperscript{4,6,21,27} However, our optimal cutoff score (17) was higher than previous studies.\textsuperscript{4,6} Only 1 other study used an ROC curve analysis,\textsuperscript{21} which also reported a cutoff score of 17; however, they found a larger positive
Kiesel et al\(^5\) reported a cutoff score of 14. The score of 14 was determined as the point where the sensitivity (true-positive rate) and 1-specificity (false-positive rate) were maximized by use of an ROC curve. Kiesel et al\(^5\) determined that players with a score \(\leq 14\) had an odds ratio of 11.67, thus indicating that they had an 11-fold greater chance of sustaining a serious injury compared with a player who scored above 14. The odds ratio actually indicates that at the cutoff of 14, the odds that an individual would actually have a true score would be 11-fold greater than a false one. It does not differentiate between those who had positive and negative outcomes. However, when likelihood ratios were used, we found that the FMS was no more useful to predict injury than the flip of a coin for any cutoff score for our entire sample or for any stratification of the sample. Regardless of how the sample was stratified, the area under the curve was <0 to 0.50, therefore resulting in no point with a strong positive likelihood ratio. For our entire sample, sensitivity (true-positive rate), 0.495 and 1-specificity (false-positive rate), 0.492, were maximized at a cutoff score of 17. For this cutoff score, we found a positive likelihood ratio of 1.54. The positive likelihood ratio represents the increase in odds favoring the condition given a positive test.\(^18\) For example, small-to-moderate findings would be defined as a positive likelihood ratio between 2 and 10, whereas larger findings would be defined as a positive likelihood ratio
greater than 10.18,28 A positive likelihood ratio between 1 and 2 indicates that the probability of increased risk is small and rarely important.18 A ratio of 1 mathematically indicates that there is approximately an equal amount of true-positives to false-positives, which is not indicative of a good diagnostic test. Our results indicated a low predictive ability for the FMS on intercollegiate football players.

Potential explanations for the differences between our findings and the results from Kiesel et al is the difference in the sample examined, as well as the injury definition used. Kiesel et al examined NFL football players, whereas we examined a sample of NCAA football players, who may not be considered elite athletes. Approximately 3% of collegiate football players make the NFL, which includes only the best athletes and most talented players.29 Therefore, a sample of collegiate football players similar to ours may include 4 potential NFL players if you apply the 3% statistic to our pool of participants. This indicates that the sample used by Kiesel et al and our sample of college athletes may differ more than is typically considered. Therefore, it is recommended that an ROC curve analysis should be done for each sample independently and using previously identified cutoff scores may not be generalizable.

More importantly, our injury definition was more inclusive of injuries than the definition used by Kiesel et al.3 Their definition was an injury of sufficient magnitude to result in the athlete being placed on injured reserve, a minimum of 3 weeks’ time loss, versus our definition that required the athlete to miss a minimum of 1 day of participation. The injuries compiled in our database included injuries involving significant time loss (>10 days), as well as injuries that were less severe and involved a loss of ≥1 day. The injuries causing a time loss of less than 10 days were then categorized into the upper extremity, lower extremity, overuse, or noncontact stratification. The injuries we captured were likely not as significant or catastrophic, but they still caused time loss, which we believe to be clinically meaningful. Our definition of injury may have contributed to our external validity because the definition is more inclusive of injuries causing less time loss. These injuries are not captured in Kiesel et al due to their restrictive injury definition. In collegiate football, on average, more than one-fourth of all injuries are considered serious, with serious injury being defined as a time loss of >10 days.1 Including injuries that are not considered serious and cause less time loss is one reason that the cutoff score in the current study was 17. Injuries causing less time loss may occur at such a high rate that they occur regardless of the individuals functional movement patterns. When we stratified the sample and looked specifically at the injuries that caused a time loss of >10 days, and we restricted our injury definition to be less inclusive of injuries, and
the cutoff score fell to 12 (Table 3). This may indicate that injury type has a significant effect on FMS cutoff scores, although it seems to make no difference on the usefulness of the test for injury prediction, as seen in our results.

The differences in scoring systems could have also altered the FMS scores when our results are compared with those of Kiesel et al. We chose to alter the scoring system so that the individual scores from each test were more representative of the participant’s performance for that individual test. We believe this was achieved by using the mode for each test, rather than using the score that was the lowest. Taking the lowest score skews the composite FMS score to be lower. Although our mean FMS score (16.1) for all participants was similar to the mean of FMS scores in Kiesel et al (16.9), our cutoff score was higher (17). If the scoring changes influenced our cutoff scores, we would expect them to inflate not only the cutoff score but the mean FMS score of the sample because we chose to take the 2 most consistent scores instead of just the lowest score per test. This procedure is different than that proposed by Cook et al, where the authors chose to always take the lowest score per test. However, our mean FMS score for the sample is similar to what has been reported in previous studies, indicating that the change of scoring is comparable to the original method, and our elevated cutoff score can be attributed to the sample and not to the methodological change in scoring.

One limitation of the current study was that 2 different investigators performed the testing; therefore, intertester reliability was not able to be recorded. However, it has been found by multiple studies that investigators with similar backgrounds and experience with the FMS have good-to-excellent reliability, meaning that investigators with equal background and experience rated participants similarly. Investigators with more extensive background and experience rate more consistently than those with less experience and background. The investigators of the current study had similar experience and backgrounds with the FMS, and it can therefore be assumed they have good-to-excellent reliability in terms of testing participants with the FMS.

Another limitation of the study is that both investigators did not hold certification for the FMS at the time of testing. However, both had significant training with the FMS from clinical instructors and had implemented FMS testing before initiating the study. Certification does not seem to be contributing factor because it has been reported that FMS certification does not lead to the most accurate or reliable screens.

We determined that the FMS is not useful in predicting injury when evaluative criteria is applied to the entire sample. This finding is contrary to previous literature that indicated predictive ability. In an effort to determine whether the predictive capability was improved when stratifying injuries, we stratified injuries into several categories: upper extremity injury, lower extremity injury, overuse injuries, noncontact injuries, and injuries causing a time loss of >10 days. We determined that the FMS was not any more predictive if the sample is stratified.

**IMPLICATIONS FOR CLINICAL PRACTICE**

We determined that the FMS test has limited predictive ability for musculoskeletal injury in NCAA Division I
football players. Regardless of how the injuries were broken down and stratified (extremity, overuse, contact, or time loss), there was little predictive utility of the FMS for any stratification of the sample. The FMS test was originally developed as a screening tool for functional asymmetries, guiding corrective strategies to eliminate functional deviations. The FMS has been demonstrated as a reliable test for rater6–9,11–13 and session7,9,11; therefore, it is a test that should be utilized within the athletic population to screen for functional asymmetries. However, the results of this study suggest that it should not be used to predict an individual’s risk of injury.

REFERENCES