

Reliability and Validity of Observational Risk Screening in Evaluating Dynamic Knee Valgus

Christina L. Ekegren, PT, MSc¹, William C. Miller, OT, PhD², Richard G. Celebrini, PT³, Janice J. Eng, PT/OT, PhD⁴, and Donna L. MacIntyre, PT, PhD⁵

¹Lecturer, Brunel University, School of Health Sciences & Social Care, Division of Physiotherapy, West London, UK

²Associate Professor, University of British Columbia, Faculty of Medicine, Department of Occupational Science & Occupational Therapy, Vancouver, BC, Canada

³PhD candidate, University of British Columbia, Faculty of Medicine, Graduate Program in Rehabilitation Sciences, Vancouver, BC, Canada

⁴Professor, University of British Columbia, Faculty of Medicine, Department of Physical Therapy, Vancouver, BC, Canada

⁵Associate Professor, University of British Columbia, Faculty of Medicine, Department of Physical Therapy, Vancouver, BC, Canada

Abstract

Study Design—Nonexperimental methodological study.

Objectives—To determine the interrater and intrarater reliability and validity of using observational risk screening guidelines to evaluate dynamic knee valgus.

Background—A deficiency in the neuromuscular control of the hip has been identified as a key risk factor for non-contact anterior cruciate ligament (ACL) injury in post pubescent females. This deficiency can manifest itself as a valgus knee alignment during tasks involving hip and knee flexion. There are currently no scientifically tested methods to screen for dynamic knee valgus in the clinic or on the field.

Methods—Three physiotherapists used observational risk screening guidelines to rate 40 adolescent female soccer players according to their risk of ACL injury. The rating was based on the amount of dynamic knee valgus observed on a drop jump landing. Ratings were evaluated for intrarater and interrater agreement using kappa coefficients. Sensitivity and specificity of ratings were evaluated by comparing observational ratings with measurements obtained using 3-dimensional (3D) motion analysis.

Results—Kappa coefficients for intrarater and interrater agreement ranged from 0.75 to 0.85, indicating that ratings were reasonably consistent over time and between physiotherapists.

Address correspondence to Christina L. Ekegren, School of Health Sciences & Social Care, Mary Seacole Building, Brunel University, Uxbridge, UB8 3PH, England. christina.ekegren@brunel.ac.uk.

The University of British Columbia Clinical Research Ethics Board approved the protocol for this study.

Sensitivity values were inadequate, ranging from 67–87%. This indicated that raters failed to detect up to a third of “truly high risk” individuals. Specificity values ranged from 60–72% which was considered adequate for the purposes of the screen.

Conclusion—Observational risk screening is a practical and cost-effective method of screening for ACL injury risk. Rater agreement and specificity were acceptable for this method but sensitivity was not. To detect a greater proportion of individuals at risk of ACL injury, coaches and clinicians should ensure that they include additional tests for other high risk characteristics in their screening protocols.

Keywords

female; knee; movement analysis; neuromuscular; soccer

INTRODUCTION

Of all athletic knee injuries an anterior cruciate ligament (ACL) rupture is the most devastating, resulting in the greatest time lost from sport.⁸ The ACL plays a vital role in the normal function and stability of the knee, and individuals wishing to return to sport after rupture are encouraged to consider reconstructive surgery.²⁹ Those that opt out of surgery are often forced to reduce their level of physical activity and their involvement in sport.⁶ This can have serious implications later in life, with reduced physical activity being associated with a higher incidence of obesity, morbidity, and mortality.^{3, 21, 32} Of most concern is that, regardless of management, an ACL injury can triple the risk of osteoarthritis (OA) developing by middle age.³⁵ This disease may have a significant impact on quality of life by further limiting functional and leisure activities.^{7, 15} The need for ACL injury prevention is clear.

To facilitate injury prevention, several authors have recommended that athletes be pre-screened for the presence of injury risk factors.^{4, 18, 26} There is strong evidence to suggest that, if detected, high risk characteristics can be improved with training^{17, 28}; and with accurate identification of at-risk athletes, these preventative strategies can be targeted towards those most in need.¹³ In a study by Myer et al,³⁰ female athletes were categorized as high risk or low risk based on the presence of certain biomechanical risk factors. The authors found that high risk athletes decreased the magnitude of their risk factors following training to a greater extent than low risk athletes. The authors concluded that the efficiency and cost-effectiveness of intervention programs could be optimized by delivering the intervention specifically to those individuals at risk.

A deficiency in the neuromuscular control of the hip has been identified as a key risk factor for non-contact ACL injury in women.^{19, 26} This deficiency will often manifest itself as a medial collapse of the knee (“dynamic knee valgus”) during tasks involving hip and knee flexion.¹⁹ Through video analysis, researchers have concluded that dynamic knee valgus plays a role in non-contact ACL injury mechanisms.^{16, 36} Numerous biomechanical studies have shown that even in a controlled laboratory setting, women have a tendency to land from a jump with more knee valgus than men.^{14, 36, 42} Because knee valgus is known to increase ACL loading,¹⁹ this movement pattern is thought to be partly related to the 4 to 6 times

higher incidence of non-contact ACL injury in females.^{1, 34} A recent prospective study of adolescent female soccer, basketball, and volleyball players found that increased knee valgus angles combined with increased knee valgus moments were predictive of a future ACL injury (predictive r^2 value = 0.88).¹⁹ It has therefore been recommended that an athlete's neuromuscular control be evaluated prior to their participation in sport using dynamic tasks such as a drop-jump landing.¹⁹

Ideally, the gold standard, 3-dimensional (3D) motion analysis, should be used to evaluate kinematics during dynamic tasks. However, it is widely acknowledged that this technique is too time consuming and costly to be used in mass screening programs.^{2, 18, 26, 37, 40} Consequently, several researchers have attempted to develop simpler methods of screening for high risk knee valgus angles.^{5, 24, 31, 40} Some of these researchers have used 2-dimensional (2D) video analysis techniques to measure the amount of valgus at the knee.^{26, 31, 37, 40} It is believed that the frontal plane knee motion observed on video is representative of the dynamic knee valgus that can be measured using 3D techniques.²⁶

Because 2D measurements of knee valgus do not show a constant relationship with 3D measurements,^{26, 40} some authors have argued that they are not sensitive enough to use in ACL risk screening.³¹ However, there is evidence to suggest that 3D knee valgus is an important component of 2D valgus²⁶; and that the hip and knee rotations that contribute to the attainment of 3D knee valgus, such as hip internal rotation,¹³ also contribute to the appearance of frontal plane knee valgus.²⁶ This suggests that there is an association between 2D and 3D measures, which we feel makes 2D analysis of knee valgus worthy of inclusion in preliminary athletic screening. It must be noted however that 2D analysis may exaggerate the amount of valgus measured in 3D.²⁶

To promote screening in the field or clinical setting, screening tools must be quick and easy to use and should enable the coach or clinician to provide immediate feedback to the athlete. Thus, in an effort to further simplify the screening process, 2 studies have used observational analysis of knee valgus. Chmielewski et al⁵ studied the reliability of using observational rating guidelines to assess the amount of frontal plane trunk and lower limb motion exhibited during a unilateral squat and a lateral step-down task; while Krosshaug et al²⁴ investigated the ability of raters to estimate subjects' hip, knee, and ankle joint angles during running and cutting trials at predetermined points on a video. They also examined whether raters could visually detect a change in joint angle between 2 time points by asking them to classify the change in knee angle as valgus, neutral, or varus.

Interestingly, these researchers found that their observational risk screening methods lacked adequate reliability⁵ and validity.²⁴ We felt that this was primarily due to their rating guidelines being too vague and subjective and that, with clearer and simpler observational screening guidelines, reliability and validity might be improved. The purpose of this study was to examine the agreement and validity of using a set of novel observational risk screening guidelines to evaluate the amount of dynamic knee valgus exhibited on a drop-jump landing. We hypothesized that both intrarater and interrater agreement would be "substantial" ($\kappa > 0.61$)²⁵ and that observational risk screening would detect high risk individuals with a high level of sensitivity (> 80%) and moderate specificity (> 50%).

METHODS

Participants

Forty female participants (mean \pm SD age, 15 ± 1 years; height, 165 ± 6 cm; body mass, 60.0 ± 8.5 kg; body mass index, 21.9 ± 2.3 kg/m²) were recruited using a convenience sampling approach from local soccer teams. The sample size was based on a goodness-of-fit formula provided by Donner and Eliasziw,⁹ factoring for 80% power and 95% confidence. This formula was developed to construct inferences for the kappa statistic when the trait of interest is measured on a dichotomous scale. Participants were included if they 1) were aged 13 – 17 and 2) played soccer at a competitive level (provincial or gold club level). Females of this age range were selectively recruited to target a population with a high risk of ACL injury.^{1, 34} Participants were excluded if they had 1) experienced a back/lower limb injury requiring at least 30 days off from full training and matches in the past or requiring at least 10 days off in the 6 weeks prior to testing, or 2) any medical problems preventing participation in testing.

Data collection

After obtaining assent from participants and informed consent from their parents/guardians, participants were scheduled for data collection. Participants provided their demographic details and injury history and had their height (cm) and body mass (kg) measured using a height rod and mechanical balance scale (Health-O-Meter, Continental Scale Corporation, Bridgeview, Ill). Once attired in a standardized pair of tight-fitting shorts and their own running shoes, participants were taught how to perform the drop jump task. They were instructed to drop down onto a force plate embedded in the ground (Bertec, Columbus, Ohio, USA) from a 31cm box and immediately perform a maximum vertical jump (FIGURE 1). They were to keep their arms in the “stop position” (shoulders abducted 45° and elbows flexed 90°) to reduce momentum from arm swing. To minimize learning effects, 3 practice trials of the drop jump task were allowed. Following this, 9 consecutive drop-jump trials were conducted. A 1:5 work/rest ratio (10-second rest between trials) was implemented to reduce fatigue.²⁰ The protocol for this study was approved by the University of British Columbia Clinical Research Ethics Board and the rights of all participants were protected.

Instrumentation

A digital camcorder (Canon ZR800A, Canon, Lake Success, NY) was used to capture footage for viewing by raters. Video was recorded at 60Hz. The camera was set up on a tripod 150cm off the ground and 330cm forwards of the jumping box and was framed below the shoulder, ensuring participant anonymity. Each participant’s testing session was downloaded, edited, and then burnt onto compact discs (CDs) using Microsoft® Windows® Movie Maker (Version 5.1, Microsoft Corporation).

Observational ratings were validated by comparing them with the results from 3D motion analysis, captured simultaneously with the video footage. Prior to performing the drop-jumps, each participant had 12 infrared emitting diodes, 10mm in diameter, placed on the pelvis, thigh, leg, and foot of their dominant lower limb as described by Jian et al²² and Eng and Winter¹¹ (FIGURE 2). The markers were secured using double-sided tape. Limb

dominance was determined by asking the participant with which limb they would prefer to kick a ball. All participants selected the right lower extremity as their dominant side. Three-dimensional kinematics were measured using an Optotrak© motion analysis system consisting of 2 camera bars containing 3 cameras each (Optotrak 3020, NDI, Waterloo, Ontario). Data were sampled at 120Hz. The knee joint center was calculated using measurements (cm) of the dominant limb and foot.^{22, 41} Standardized joint coordinate systems for each segment (thigh, leg, and foot) were defined using digitized landmarks and 3 non collinear markers were used to track each segment.

Data Processing

Custom software written in Matlab (Version 14, The Mathworks Inc., Natick, MA) was used to calculate 3D kinematic rotation angles of the knee (cardan sequence: extension/adduction/internal rotation), with the key variable of interest being the knee valgus angle. The data were filtered using a Butterworth filter (4th order, zero-lag, low pass cut-off at 10Hz). Knee extension, adduction, and internal rotation angles were defined as positive values and were considered to be equal to zero based on the knee alignment values measured during an initial 4-second stationary recording of each participant standing with feet hip-width apart. Force plate data, sampled at 600Hz, were used to isolate landing phase kinematics. The landing phase was defined as the period from foot contact to toe off and was manually selected during Matlab processing from a graph of the vertical ground reaction force against time.

Trials were excluded during processing if markers were missing at the start or end of the landing phase or if there was greater than 10mm of motion between markers on each limb segment. The kinematics of all remaining trials were time normalized to 100% of the landing phase and overlaid on a single graph. Atypical trials were excluded using visual inspection. A total of 78% of trials remained for analysis and from these, 3 trials per participant were randomly selected. Only these trials appeared on the video shown to raters.

Rating Protocol

From a pool of 15 potential raters with at least 5 years of experience in private practice and a high level of sports and orthopedic expertise, 3 female physiotherapists volunteered (mean \pm SD years of clinical experience, 12 ± 3 years). They were mailed a 20-minute training CD prior to the first rating session which included background information about ACL injury risk, detailed rating instructions, and a practice rating session. The first rating session was attended by all 3 physiotherapists and the lead author. The video footage was projected onto a 2m \times 2m screen in a darkened room at our research laboratory. Ten minutes were spent reviewing the rating instructions and practicing using footage from a larger study. Raters were permitted to share their decisions and any disagreements were discussed until a consensus could be reached. The rating session commenced once all physiotherapists felt confident with the instructions.

The 40 participants were shown in random order. Raters were asked to give an overall rating of “high risk” or “low risk” to each of the participants based on their 3 trials. In addition, the following rules were set: 1) raters were to complete their assessment in the 15 seconds between each participant; 2) they were permitted a single viewing only and the footage was

never paused; 3) they were to refrain from sharing their ratings or making any comments; 4) they were to focus only on the dominant lower limb; and 5) they were to focus only on the first landing from the box.

To evaluate intrarater agreement, raters reassessed the same footage 2 weeks later. This time, raters were sent the footage on CD to view on their own computers. The order of participants was unchanged from the previous session. Rating sessions were scheduled 2 weeks apart to reduce the likelihood of raters remembering their initial assessments.³⁹

The guidelines for risk screening were developed by the lead author, an orthopaedic physiotherapist, in consultation with an experienced sports physiotherapist (RC). To date, no literature exists describing the visual appearance of a high risk valgus angle. Therefore, a set of observational risk screening guidelines were created based on our current understanding of ACL injury risk factors and on the normal spectrum of biomechanics seen in clinical practice. The guidelines were as follows: “If the patella moves inwards and ends up medial to the first toe, rate the individual as high risk” or “if the patella lands in line with the first toe rate the individual as low risk” (FIGURE 3). The guidelines were designed to capture as many potentially high risk athletes as possible. Therefore, if only one of the 3 trials was deemed high risk, raters were to assign an overall high risk rating to the athlete.

Data Analysis - Rater Agreement

We used percentage of agreement and the standard kappa (κ) coefficient for intrarater agreement (Analyse-it Software Ltd, Method Evaluation Edition, Leeds, UK) and the multirater kappa coefficient for interrater agreement (software © 2004 Jason King, PhD; available at <http://www.ccit.bcm.tmc.edu/jking/homepage/kappa.xls> or see APPENDIX 1 for formula).¹²

Data Analysis - Validity

Pilot data from a previous reliability study, where subjects were tested twice, 1 week apart, revealed that the amount of valgus motion during ground contact (maximum - minimum knee valgus) could be measured with better reliability than the peak (maximum) knee valgus angle. Therefore, to assess validity, we compared physiotherapists' ratings with each participant's mean knee valgus motion, measured using the gold standard, 3D motion analysis. The trials used to calculate the mean were the same 3 trials shown to the raters.

To develop our own gold standard cut point between “truly high risk” and “truly low risk” groups, the lead author reviewed the footage and gave each participant an expert rating of high risk or low risk. To ensure that the risk screening guidelines were applied precisely she was permitted to pause, decelerate, and rewind the video footage as necessary. This process was undertaken twice, 1 month apart. If there was a difference in judgement for a subject between time 1 and 2, the footage was reviewed a third time to come to a final decision. A receiver operating characteristic (ROC) curve was then constructed, linking the expert ratings with each participant's mean knee valgus motion value. The point on the curve with the highest sensitivity and specificity (usually the point closest to the upper left hand corner) was chosen as the final cut-off point.³⁹ Participants with a mean knee valgus motion value above this cut-off point were designated “truly high risk” and those with a value below the

cut-off were designated “truly low risk”. [Note: “high risk” and “low risk” refer to the 2D observational rating and “truly high risk” and “truly low risk” refer to the 3D motion analysis rating.] Sensitivity, specificity, positive predictive values (PV+), and negative predictive values (PV-) were calculated (Analyse-it Software) by comparing physiotherapists’ ratings with these true ratings (see APPENDIX 2 for formula).³³

While a perfect screening test would have a sensitivity of 100% and a specificity of 100%, in reality most screening tests are imprecise; sensitivity being sacrificed for specificity or vice versa. For our purposes, sensitivity was given priority over specificity, with the desired hypothesized level of sensitivity being set at 80% and that of specificity being set at 50%. Attaining a high level of sensitivity ensures that individuals with high risk characteristics are not falsely labeled low risk. It is important that these high risk individuals are detected as they may be more likely to experience an injury. Specificity was deemed to be less important, the main drawback of poor specificity being that some low risk individuals might be falsely labeled high risk and receive injury prevention training that, while not harmful, is unnecessary.

RESULTS

Rater Agreement

Intrarater Agreement—All 3 raters achieved “substantial” intrarater agreement, with κ values exceeding the hypothesized target (TABLE 1). Rater 1 achieved 90% agreement between time 1 and time 2 ratings and a κ of 0.80 (95%CI = 0.65–1.00), rater 2 had 92.5% agreement and κ of 0.85 (0.72–1.00), and rater three had 87.5% agreement and κ of 0.75 (0.58–1.00).

Interrater Agreement—At both time points, interrater agreement exceeded the hypothesized target (TABLE 2). The multirater kappa value was 0.80 (95%CI = 0.62–0.98) at time 1 and 0.77 (0.59–0.95) at time 2.

Validity

A cut-off point of 10.83° of knee valgus motion was chosen because out of all points in the upper left hand corner of the ROC curve (FIGURE 4), this point had the highest sensitivity (87%). Using results from 3D motion analysis, 15 participants (37.5%) were deemed to be “truly high risk” and 25 (62.5%) were deemed to be “truly low risk” in relation to the cut-off point. Sensitivity targets were met only twice - by rater 3 at time 1 and rater 1 at time 2 (TABLE 3). All raters exceeded hypothesized specificity targets.

DISCUSSION

The purpose of this study was to validate a set of observational risk screening guidelines designed to assess risk for ACL injury. The guidelines were designed for coaches and trainers to use with their athletes in the field, enabling them to perform quick pre-season screenings to determine readiness to play and also to direct injury prevention strategies. The guidelines were also aimed at clinicians who might wish to use a task such as the drop jump to assess their patients’ neuromuscular hip control in a more dynamic manner.

We built our simple approach on the foundational work of other groups^{5, 10, 23, 38} which was a key factor in achieving high rater agreement. Specifically, we developed explicit rating criteria incorporating easily discernible anatomical landmarks (the patella and the first toe) based on recommendations by Chmielewski et al,⁵ who had poor results with more ambiguous rating criteria such as “good”, “fair”, and “poor”. Our decision to use rating categories rather than asking raters to estimate exact range of motion values was based on the results of Knudson et al²³ who reported poor intrarater reliability with the alternative. Finally, our decision to limit the number of response categories to 2 was based on the assertion of Chmielewski et al⁵ that the use of multiple scoring response categories had led to less than desirable agreement in their study.

While our approach to screening might be viewed as over-simplistic, its straightforwardness is aimed at optimizing screening compliance amongst clinicians and coaches. We do acknowledge however, that screening should not focus *solely* on knee valgus as a risk factor. Knee valgus is not involved in all ACL injuries and not all individuals at risk of ACL injury exhibit this characteristic.²⁴ Other commonly observed ACL injury mechanism components include a knee that is close to full extension³⁶; and a center of mass that is not aligned vertically over the planted foot.⁴³ Additional female ACL risk factors that have been identified include limb dominance, which refers to having significant side-to-side differences in lower limb strength and neuromuscular control, and quadriceps dominance, which refers to a mismatch in strength and timing between the quadriceps and the hamstrings.^{16, 19} Meeuwisse’s model of injury etiology,²⁷ often cited in relation to ACL injury, asserts that a variety of intrinsic risk factors will predispose an athlete to injury and it is the complex interaction between intrinsic and extrinsic risk factors that makes an athlete susceptible to injury. Thus, it must be reinforced that these guidelines should be used in conjunction with other screening components as part of a comprehensive package. Providing there is adequate reliability and validity, additional tests for components such as lower limb imbalance or proximal control could also be included.

It is important that prior to undertaking risk screening, coaches and clinicians first receive adequate rater training. Eastlack et al¹⁰ identified that a lack of standardization of rater training had led to poor rater reliability in their study. We therefore provided the same in-depth training to all of our raters, giving them ample opportunity to practice and to ask questions prior to starting. We also sought to improve consistency and interrater reliability by selecting a homogeneous population of raters. All 3 raters were physiotherapists with at least 5 years of experience in private practice and a high level of sports and orthopedic expertise. While this may limit the generalizability of our findings to other health professionals with different levels of experience, previous studies,^{23, 38} suggest that experience may not greatly affect rating skill. Other factors, such as dynamic visual acuity, experience in performing the task being rated, the amount of specific training received and the perceptual style of the raters, may actually be more influential.²³

Despite the importance of validity in risk screening, few previous studies have investigated this metric. In our study, sensitivity was inadequate - with up to a third of “truly high risk” individuals failing to be detected by raters. In risk screening, if there is a failure to identify individuals at risk of injury, they could miss out on vital injury prevention initiatives and go

on to experience an injury. As a result, there might ultimately be little reduction in injury rates. Had the raters labelled more individuals as high risk, sensitivity may have been higher. Thus, in future, raters should be encouraged to err on the side of caution and only assign a low risk rating where there is absolutely no doubt of that status. It must be acknowledged however that observational analysis is unlikely to ever attain the sensitivity of more precise 3D methods, highlighting the importance of including other high risk characteristics in athletic screening protocols.

Another factor affecting sensitivity was the prevalence of “truly high risk” and “truly low risk” individuals in the sample, which, in this study, was entirely dependent on the cut-off point chosen. Had the cut-off point been set higher, e.g. 12°, there would have been fewer “truly high risk” individuals in the sample which may have improved sensitivity. However, this would have forced us to sacrifice our specificity values. Low specificity means that a large proportion of low risk individuals have been mistaken as high risk. Not only does this result in some individuals receiving unnecessary injury prevention training, it must also be considered that, in a competitive team setting, there may be an undesirable negative stigma attached to a high risk label. Furthermore, wrongfully identifying individuals as high risk may make them fearful of injury and reluctant to participate in sport.

In deciding whether a risk screening program is financially viable, positive predictive values can be a useful guide. Our positive predictive values indicated that 35–48% of individuals with a high risk rating may not actually have been at risk, meaning that up to almost half of the sample would have received unnecessary injury prevention training -potentially an inefficient use of resources. Negative predictive values estimate the likelihood that an individual with a low risk rating is “truly low risk”. In risk screening, it is important to achieve a high negative predictive value as this means that fewer high risk individuals have been falsely labelled low risk. Our negative predictive values ranged from 79–90% which means that 10–21% of participants rated low risk were actually high risk. This is an unacceptably high proportion, suggesting once again that our method lacked the sensitivity to detect all high risk individuals.

An additional factor that may have influenced validity was the method used to assess 3D knee valgus. The decision to use valgus excursion (maximum-minimum) was based on the much better reliability with which excursion could be measured compared to the peak angle. Using this measure meant that, if, as an example, a participant landed in 12° of varus and moved to neutral during the drop jump, they would be categorized as high risk (12° of motion). However, the rater would observe that the knee was over the first toe and categorize the participant as low risk.

On examining our data, we found that this problem had not occurred. The data showed that the highest degree of knee varus recorded in our sample was 2.6°, with only 5 out of our 40 participants landing in any amount of varus at all. This suggests that in our sample it was uncommon for an individual to land with a varus knee alignment and then move to a neutral position. We also examined the data to see whether any individuals had recorded minimal valgus excursion angles and were therefore categorized low risk but had high peak valgus angles, therefore appearing high risk to raters. As an example, an individual might land in

15° of valgus, move to a peak valgus of 20° and end up with a minimal excursion value of 5°. We did not find this to be the case. Seven out of the 10 individuals with the highest peak valgus also had the highest amount of excursion, suggesting that their high peak angles were the result of considerable valgus motion occurring during the landing phase.

Finally, we recalculated sensitivity, specificity and predictive values using the less reliable peak valgus measure as the gold standard. We found that the sensitivity and predictive values actually got worse, suggesting that what the raters had observed was more closely associated with the excursion value. This finding is interesting, considering that the raters' instructions were more related to a finite knee position than they were to knee motion. It may be, as suggested by Knudson,²³ that the human eye is better at detecting sweeps of body motion than at evaluating discrete positions at key points in the movement. These findings may also be related to the poor reliability of our peak valgus measure. We acknowledge, however, that these findings may be unique to our sample and that there would certainly be merit in re-evaluating the participants incorporating the concept of knee excursion in our instructions to raters.

A number of limitations may have affected our study findings. Using our methods it was difficult to choose an ideal cut-off point as our ROC curve was quite irregular, changing direction at lower cut point levels. The abnormal shape resulted from the poor equivalency between observational ratings and the results from our gold standard, 3D motion analysis. It seems that the knee angles captured with 3D motion analysis were not always evident with respect to frontal plane motion. Laboratory-based measurement error, caused by excessive surface marker motion or inaccurate marker placement, may have been partly responsible for this. However, the more likely explanation is that 3D measures cannot be used interchangeably with 2D measures. This is consistent with previous reports.

McLean et al²⁶ reported only a moderate association ($r^2 = 0.25-0.36$) between 2D and 3D knee valgus angles and Krosshaug et al²⁴ reported poor agreement ($\kappa = 0.19$) between knee valgus estimates and 3D values. Willson and Davis⁴⁰ found that the frontal plane projection angle (FPPA) of knee valgus only contributed 23–30% of the variance of 3D kinematics and that the FPPA was not significantly correlated with 3D knee abduction. In accordance with our findings, McLean et al's²⁶ 3D angles were consistently smaller than the corresponding angles measured from video. This may have resulted from the knee flexion and hip internal rotation that produces knee valgus being more obvious on video in the frontal plane than in the orientation of the knee joint axis.

Better sensitivity might have been achieved had our raters been able to observe the participants live (in 3 dimensions) rather than on video (in 2 dimensions). This would have ensured a comparison of more similar methods of analysis (i.e. 3D with 3D) and would have better replicated what would be done in the field. The decision to use video rather than live screening was a pragmatic one as each participant's laboratory testing session took at least 2 hours to complete. Also, if we had used live screening our rater agreement may have suffered owing to difficulties ensuring a consistent viewpoint for all 3 raters. Further investigation into the reliability and validity of live screening methods is required. It may also be useful to investigate these guidelines without the athletes wearing infrared emitting

diodes as these would not be present in the field or clinical setting. We do not believe that the markers assisted the raters to make their judgement to any large extent as they were not placed on the patella or the first toe. However, they may have been an added distraction.

Our results may also have been influenced by the different manner in which the 2 rating sessions were run. At time 1, the rating session was held at the research laboratory, while at time 2, raters were mailed the footage on CD to watch on their own home computers. Viewing conditions may have been suboptimal in this setting and raters did not receive the same in-depth training as they had at time 1. Also, while the raters did receive a written reminder of the rules and guidelines, they did not receive the same practice time that they had received at time 1. These factors may explain the slightly lower kappa and sensitivity values at the second evaluation session. In addition, because the raters were working without supervision we had to trust that they were following all of the rating rules, such as not pausing or rewinding the footage. However, because there was no systematic improvement in their ratings between the two evaluation sessions, we felt it was likely that they had adhered to the study protocol.

Another limitation of our study was that we only examined the dominant lower limb. This decision was based on evidence suggesting that the dominant limb displays more high risk neuromuscular characteristics than the non-dominant limb.^{13, 16} Interestingly though, both the dominant and the non-dominant limb appear to be equally at risk of ACL injury, with the dominant limb being over-utilized and therefore put under more strain and the non-dominant limb being under-utilized and thereby made weaker and less resilient.^{16, 19} Therefore, future research requiring raters to examine both lower limbs is warranted. Also, while our study sample was selectively recruited to target the population most at risk of ACL injury,^{1, 34} future testing should be conducted with a larger sample of randomly-selected athletes, including male soccer players, children and adults. This would enable us to assess the usefulness of the screening guidelines in other populations who may also be at risk of ACL injuries, thereby further validating the screen. Furthermore, it is clear that prospective evidence is needed as to what the actual biomechanical cut point is between individuals who are and those who are not at risk of ACL injury. Only then will it be possible to examine the *true* validity of clinical risk screening guidelines.

CONCLUSION

Raters were able to reliably detect those athletes with a midline collapse of the knee in the frontal plane but they missed certain high risk athletes that had been detected using 3D methods. We have contributed to the literature by testing a standardized field screening method but it must be acknowledged that observational risk screening is unlikely to achieve the sensitivity of more precise methods, such as 3D motion analysis. Coaches and clinicians should ensure that, to detect a greater proportion of individuals at risk of ACL injury, additional tests for other high risk characteristics should also be included in their screening protocols.

Acknowledgments

This work was supported by an operating grant from the Vancouver Foundation's British Columbia Medical Service Foundation (BCMSF). Further support for this study was given in a BCMSF Summer Scholarship (CLE) and career scientist awards from the Canadian Institute of Health Research (CIHR) to WCM (CIHR-72159) and JJE and the Michael Smith Foundation for Health Research (JJE).

The authors acknowledge physiotherapists Nadine Nembhard, Marilou Lamy and Dana Ranahan for their involvement in this study. In addition, the authors acknowledge Tom Depew and JD Johnston for their assistance with software programming and data processing. Finally the authors acknowledge Bill Trenaman and Larry Moro from the Port Moody Soccer Club and Markus Reinkens from the BC Soccer Association for their assistance with subject recruitment.

References

1. Arendt E, Dick R. Knee injury patterns among men and women in collegiate basketball and soccer. NCAA data and review of literature. *Am J Sports Med.* 1995; 23:694–701. [PubMed: 8600737]
2. Barber-Westin SD, Noyes FR, Galloway M. Jump-land characteristics and muscle strength development in young athletes: A gender comparison of 1140 athletes 9 to 17 years of age. *Am J Sports Med.* 2006; 34:375–384. [PubMed: 16282578]
3. Blair SN, Brodney S. Effects of physical inactivity and obesity on morbidity and mortality: Current evidence and research issues. *Med Sci Sports Exerc.* 1999; 31:S646–62. [PubMed: 10593541]
4. Bonci CM. Assessment and evaluation of predisposing factors to anterior cruciate ligament injury. *J Athl Train.* 1999; 34:155–164. [PubMed: 16558559]
5. Chmielewski TL, Hodges MJ, Horodyski M, Bishop MD, Conrad BP, Tillman SM. 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; 37:122–129. [PubMed: 17416127]
6. Daniel DM, Stone ML, Dobson BE, Fithian DC, Rossman DJ, Kaufman KR. Fate of the ACL-injured patient. A prospective outcome study. *Am J Sports Med.* 1994; 22:632–644. [PubMed: 7810787]
7. Dekker J, Tola P, Aufdemkampe G, Winckers M. Negative affect, pain and disability in osteoarthritis patients: The mediating role of muscle weakness. *Behav Res Ther.* 1993; 31:203–206. [PubMed: 8442746]
8. Dick R, Putukian M, Agel J, Evans TA, Marshall SW. Descriptive epidemiology of collegiate women's soccer injuries: National collegiate athletic association injury surveillance system, 1988–1989 through 2002–2003. *J Athl Train.* 2007; 42:278–285. [PubMed: 17710177]
9. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: Confidence interval construction, significance testing and sample size estimation. *Statistics in Medicine.* 1992; 11:1511–1519. [PubMed: 1410963]
10. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped observational gait-analysis assessments. *Phys Ther.* 1991; 71:465–472. [PubMed: 2034709]
11. Eng JJ, Winter DA. Kinetic analysis of the lower limbs during walking: What information can be gained from a three-dimensional model? *J Biomech.* 1995; 28:753–758. [PubMed: 7601875]
12. Fleiss, JL. *Statistical Methods for Rates and Proportions.* 2. New York: John Wiley and Sons, Inc; 1981.
13. Ford KR, Myer GD, Hewett TE. Valgus knee motion during landing in high school female and male basketball players. *Med Sci Sports Exerc.* 2003; 35:1745–1750. [PubMed: 14523314]
14. Ford KR, Myer GD, Smith RL, Vianello RM, Seiwert SL, Hewett TE. A comparison of dynamic coronal plane excursion between matched male and female athletes when performing single leg landings. *Clin Biomech (Bristol, Avon).* 2006; 21:33–40.
15. Hartwick M, Meeuwisse W, Vandertuin J, Maitland M. Knee pain in the ACL-deficient osteoarthritic knee and its relationship to quality of life. *Physiother Res Int.* 2003; 8:83–92. [PubMed: 12879730]

16. Hewett TE, Myer GD, Ford KR. Decrease in neuromuscular control about the knee with maturation in female athletes. *J Bone Joint Surg Am*. 2004; 86-A:1601–1608. [PubMed: 15292405]
17. Hewett TE, Myer GD, Ford KR. Reducing knee and anterior cruciate ligament injuries among female athletes: A systematic review of neuromuscular training interventions. *J Knee Surg*. 2005; 18:82–88. [PubMed: 15742602]
18. Hewett TE, Myer GD, Ford KR, Slauterbeck JR. Preparticipation physical examination using a box drop vertical jump test in young athletes: The effects of puberty and sex. *Clin J Sport Med*. 2006; 16:298–304. [PubMed: 16858212]
19. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: A prospective study. *Am J Sports Med*. 2005; 33:492–501. [PubMed: 15722287]
20. Jacobs C, Mattacola C. Sex differences in eccentric hip-abductor strength and knee-joint kinematics when landing from a jump. *Journal of Sport Rehabilitation*. 2005; 14:346–355.
21. Jebb SA, Moore MS. Contribution of a sedentary lifestyle and inactivity to the etiology of overweight and obesity: Current evidence and research issues. *Med Sci Sports Exerc*. 1999; 31:S534–41. [PubMed: 10593524]
22. Jian Y, Dinter DA, Ishac MG, Gilchrist L. Trajectory of the body COG and COP during initiation and termination of gait. *Gait Posture*. 1993; 1:9–22.
23. Knudson D. Validity and reliability of visual ratings of the vertical jump. *Percept Mot Skills*. 1999; 89:642–648. [PubMed: 10597600]
24. Krosshaug T, Nakamae A, Boden B, et al. Estimating 3D joint kinematics from video sequences of running and cutting maneuvers--assessing the accuracy of simple visual inspection. *Gait Posture*. 2007; 26:378–385. [PubMed: 17141503]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
26. McLean SG, Walker K, Ford KR, Myer GD, Hewett TE, van den Bogert AJ. Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *Br J Sports Med*. 2005; 39:355–362. [PubMed: 15911607]
27. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A dynamic model of etiology in sport injury: The recursive nature of risk and causation. *Clin J Sport Med*. 2007; 17:215–219. [PubMed: 17513916]
28. Mizner RL, Kawaguchi JK, Chmielewski TL. Muscle strength in the lower extremity does not predict postinjury improvements in the landing patterns of female athletes. *J Orthop Sports Phys Ther*. 2008; 38:353–361. [PubMed: 18515963]
29. Moksnes H, Snyder-Mackler L, Risberg MA. Individuals with an anterior cruciate ligament-deficient knee classified as noncopers may be candidates for nonsurgical rehabilitation. *J Orthop Sports Phys Ther*. 2008; 38:586–595. [PubMed: 18979658]
30. Myer GD, Ford KR, Brent JL, Hewett TE. Differential neuromuscular training effects on ACL injury risk factors in “high-risk” versus “low-risk” athletes. *BMC Musculoskelet Disord*. 2007; 8:39. [PubMed: 17488502]
31. Noyes FR, Barber-Westin SD, Fleckenstein C, Walsh C, West J. The drop-jump screening test: Difference in lower limb control by gender and effect of neuromuscular training in female athletes. *Am J Sports Med*. 2005; 33:197–207. [PubMed: 15701605]
32. Patrick K, Norman GJ, Calfas KJ, et al. Diet, physical activity, and sedentary behaviors as risk factors for overweight in adolescence. *Arch Pediatr Adolesc Med*. 2004; 158:385–390. [PubMed: 15066880]
33. Portney, LG., Watkins, MP. *Foundations of Clinical Research: Applications to Practice*. 2. New Jersey: Prentice Hall Health; 2000.
34. Powell JW, Barber-Foss KD. Sex-related injury patterns among selected high school sports. *Am J Sports Med*. 2000; 28:385–391. [PubMed: 10843133]
35. Roos H, Lindberg H, Gardsell P, Lohmander LS, Wingstrand H. The prevalence of gonarthrosis and its relation to meniscectomy in former soccer players. *Am J Sports Med*. 1994; 22:219–222. [PubMed: 8198190]

36. Sell TC, Ferris CM, Abt JP, et al. The effect of direction and reaction on the neuromuscular and biomechanical characteristics of the knee during tasks that simulate the noncontact anterior cruciate ligament injury mechanism. *Am J Sports Med.* 2006; 34:43–54. [PubMed: 16210581]
37. Sigward SM, Ota S, Powers CM. Predictors of frontal plane knee excursion during a drop land in young female soccer players. *J Orthop Sports Phys Ther.* 2008; 38:661–667. [PubMed: 18978451]
38. Somers DL, Hanson JA, Kedzierski CM, Nestor KL, Quinlivan KY. The influence of experience on the reliability of goniometric and visual measurement of forefoot position. *J Orthop Sports Phys Ther.* 1997; 25:192–202. [PubMed: 9048325]
39. Streiner, DL., Norman, GR. *A Practical Guide to their Development and use.* 3. New York: Oxford University Press Inc; 2003. Health Measurement Scales.
40. Willson JD, Davis IS. Utility of the frontal plane projection angle in females with patellofemoral pain. *J Orthop Sports Phys Ther.* 2008; 38:606–615. [PubMed: 18827327]
41. Yeadon MR, Morlock M. The appropriate use of regression equations for the estimation of segmental inertia parameters. *J Biomech.* 1989; 22:683–689. [PubMed: 2808449]
42. Yu B, Lin CF, Garrett WE. Lower extremity biomechanics during the landing of a stop-jump task. *Clin Biomech (Bristol, Avon).* 2006; 21:297–305.
43. Zazulak BT, Hewett TE, Reeves NP, Goldberg B, Cholewicki J. Deficits in neuromuscular control of the trunk predict knee injury risk: A prospective biomechanical-epidemiologic study. *Am J Sports Med.* 2007; 35:1123–1130. [PubMed: 17468378]

Appendix A: Calculations for Multirater Kappa¹²

$$\hat{k} = \frac{\bar{p} - \bar{p}_e}{1 - \bar{p}_e}$$

Where \bar{p} is the mean agreement for all participants ($n = 40$), agreement is

$$\sum_{j=1}^R \frac{n_{ij}(n_{ij}-1)}{K(K-1)},$$

$$\bar{p}_e = \sum_{j=1}^R \hat{p}_i^2$$

n_{ij} is the number of raters who classified participant i in category j , K is the total number of raters (3), R is the number of decision categories (2), and \hat{p}_i is the proportion of all classifications that fall within each decision category.

Appendix B: Calculations for Sensitivity, Specificity, Positive Predicted Value (PV+), and Negative Predicted Value (PV-) ³³

$$\text{Sensitivity, } \frac{a}{a+c}$$

$$\text{Specificity, } \frac{d}{b+d}$$

$$\text{PV+}, \frac{a}{a+b}$$

$$\text{PV-}, \frac{d}{c+d}$$

Observational Ratings	Truly High Risk (10.83° Valgus)	Truly Low Risk (10.83° Valgus)	Total
High Risk	<i>a</i>	<i>b</i>	<i>a + b</i>
Low Risk	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>n</i>

KEY POINTS

Findings

Physiotherapists achieved “substantial” rater agreement when evaluating dynamic knee valgus using observational risk screening guidelines. However their ratings lacked adequate sensitivity when compared with the results from 3D motion analysis.

Implication

Inadequate sensitivity resulted in up to a third of “truly high risk” individuals failing to be detected by raters. The consequence for these individuals is that they might miss out on future injury prevention initiatives and go on to experience an injury.

Caution

Observational risk screening is unlikely to achieve the sensitivity of more precise methods, such as 3D motion analysis. Also, it is as yet unknown exactly how much dynamic knee valgus may predispose an individual to future injury. Until this is known, the validity of screening guidelines such as this will remain unclear.



Figure 1.
Drop-jump task. To perform a drop-jump, the participant drops off a box onto the ground and then performs a maximum vertical jump.

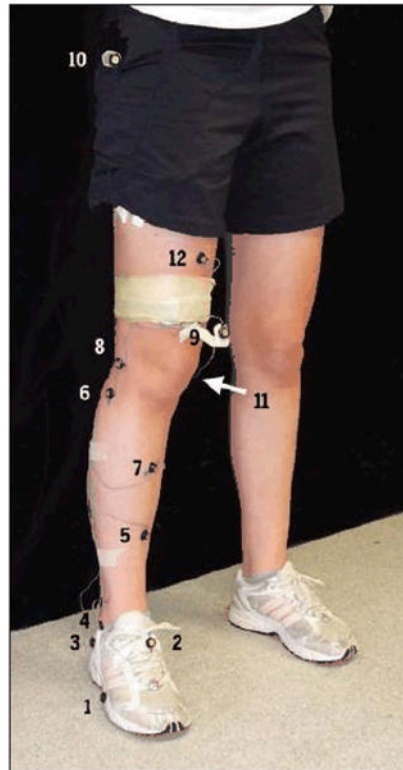


Figure 2.

Location of infrared emitting diodes. (1) Head of fifth metatarsal; (2) dorsal foot (midpoint of metatarsals and ankle); (3) lateral heel; (4) lateral malleolus; (5) midshank (anterior aspect of tibia, midpoint of ankle and knee); (6) head of fibula; (7) middle of tibia; (8) lateral femoral condyle; (9) lower thigh; (10) greater trochanter; (11) medial femoral condyle; (12) middle of femur



Figure 3. Low-risk and high-risk landing. Figure on left shows a high-risk participant where the patella has moved inward and ended up medial to the first toe. Figure on right shows a low-risk participant where the patella has remained in line with the first toe.

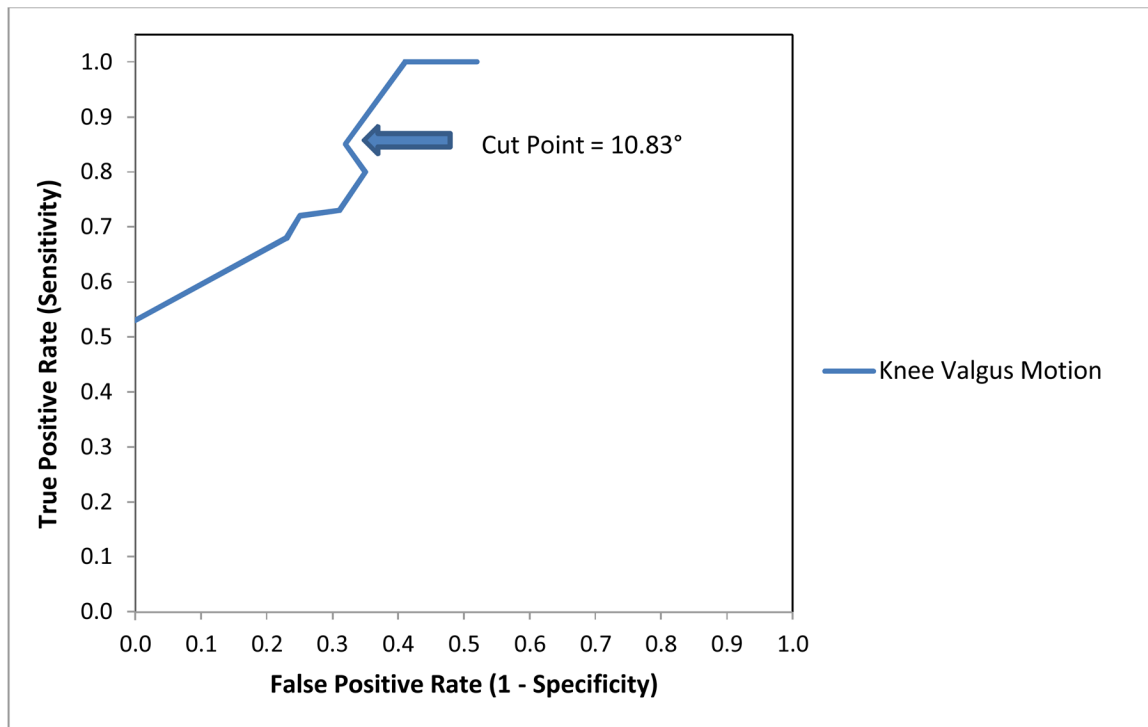


Figure 4. Receiver Operating Characteristic (ROC) curve links the expert ratings with the results from 3-dimensional motion analysis. The point in the upper left hand corner with the highest sensitivity is the cut point and is equal to 10.83° knee valgus motion.

Table 1

Intrarater Agreement of Observational Ratings of Knee Valgus Angle.

	Percentage of Agreement	Kappa Value	95% Confidence Intervals	
			Lower	Upper
Rater 1	90.0%	0.80	0.65	1.00
Rater 2	92.5%	0.85	0.72	1.00
Rater 3	87.5%	0.75	0.58	1.00

Table 2

Interrater Agreement of Observational Ratings of Knee Valgus Angle

	Kappa Values		95% Confidence Intervals	
	Time 1	Time 2	Time 1	Time 2
Multirater kappa	0.80	0.77	0.62 to 0.98	0.59 to 0.95

Table 3

Validity of Observational Ratings when Compared with Results From 3-Dimensional Motion Analysis *

	Sensitivity	Specificity	PV +	PV -
Time 1				
Rater 1	73	72	61	82
Rater 2	73	60	52	79
Rater 3	87	72	65	90
Time 2				
Rater 1	87	64	59	89
Rater 2	67	68	56	77
Rater 3	67	72	59	78

Abbreviations: PV+, positive predicted value; PV-, negative predicted value.

* Values expressed as a percent (%)