**Appendix 1. Background, Measurement Properties, Testing Procedures and Data Reduction for Self-Reported and Functional Study Outcomes**

**Knee Injury and Osteoarthritis Outcome Score**

TheKOOS is a self-report measure designed to evaluate knee related symptoms and function in young active patients with knee injury and osteoarthritis (OA). It has been validated in different populations varying in age, disease duration and activity levels and it has been shown to have high test-retest reliability.19 The KOOS consists of 42 items in five subscales: pain, other symptoms, function in daily living, function in sport and recreation, and knee-related quality of life. Each item was scored on a 5-point likert scale ranging from no problems to extreme problems. Subscale scores were then summed, and the total sub-scale score transformed to a 0-100 scale (higher scores indicating better outcome).

**Intermittent and Constant Osteoarthritis Pain Questionnaire**

The Intermittent and Constant OA Pain Questionnaire is a self-report measure designed to evaluate constant and intermittent pain intensity including frequency and impact on mood, sleep and quality of life in persons with hip and knee OA.20,35 As many patients with knee OA report that their initial presentation of pain was intermittent, often during tasks such as climbing the stairs,36 the ICOAP, which includes a subscale for intermittent symptoms20 was seen as particularly relevant to the cohort under investigation. This self-report questionnaire consists of 11 items forming two subscales (5 items addressing constant pain and 6 items addressing intermittent pain). The ICOAP has good internal consistency, test retest reliability (ICC=0.85) and construct validity when compared to KOOS and Western Ontario and McMaster Universities Arthritis Index Scores.20 Each item was scored on a 5-point likert scale ranging from no pain to high (disability-severely limiting) pain. Sub-scale scores were then summed and the total sub-scale score transformed to a 1-100 scale (higher scores indicating poorer outcome).

**Strength, Triple-Hop and Y Balance Test**

Normalized knee extensor, flexor isometric strength (Nm/kg) was assessed using hand-held isometric dynamometry (Model 01163, Lafayette Instrument, USA). Participants performed one practice and three experimental trials (5 second maximal effort of pushing/pulling into an immovable strap) for all muscles, interspersed by a 15 second rest periods, with the peak value from the three trials recorded for analyses. Knee extension strength was measured using previously described methods37 (peak force ICC: intra-session 0.96-0.97). Specifically, with participants seated with hips and knees in 90o and 60o flexion respectively, the dynamometer was placed 5 cm proximal to the distal tip of the lateral malleoli on the anterior shin and secured to the leg with an immovable strap.38, 39 Knee flexion strength was tested in a prone position with the knee in 60 degrees flexion and the dynamometer placed 5cm proximal to the distal tip of the lateral malleoli on the posterior aspect of the ankle, and secured to the leg with an immovable strap. The peak isometric strength (N) scores were converted to torque values (Nm; force x distance between joint line and dynamometer position) and normalized to body weight (Nm/kg)*.* For the purposes of this investigation, the index leg refers to the injured leg of the injured participants and corresponding leg of the matched control.

Dynamic Balance was assessed with a triple-hop for distance and Y- Balance (YBT) tests. As previously described18 the triple-hop test is a performance-based measure that evaluates neuromuscular control, force generating capacity and leg confidence that has been validated and used in individuals with ACL injury.40 Longitudinally, it has been shown to be moderately correlated to the Global Rating Scale (r=0.44) and Lower Extremity Functional Scale (r=0.26),41 and has been hypothesized to be a useful tool to predict risk of future problems as a result of a previous knee injury. After a practice trial each participant performed 2 trials of 3 consecutive single-leg hops with the goal of jumping as far as possible.42 For a trial to be included, the landing had to be solid without excessive unbalanced movements or twisting of the foot. The maximum distance across trials was recorded for each leg and expressed as a percentage of leg length. The YBT is a combined measure of postural control, strength, range of motion, and proprioceptive abilities. Previous studies have reported excellent test-retest reliability (ICC 0.82 - 0.87),43, 44 an ability to detect balance deficits in young adults with chronic ankle instability,47 and predict lower limb injuries in high school basketball players.44 Participants stood on one leg in the centre of a grid, with hands on the pelvis, and reached with the non-stance leg as far as possible in anterior, posterolateral and posteromedial directions.44 A trial was discarded and repeated if the participant failed to maintain unilateral stance, lifted or moved the stance foot from the grid, touched down with the reach foot, or failed to return the reach foot to the starting position. After a practice trial on each leg, three trials on each leg were completed with 15-second rests between trials. The maximal reach distance at the point where the most distal part of the foot extended in all three directions to was measured, totalled and normalized for lower limb length (YBT score = maximal anterior distance + maximal posterolateral distance + maximal posteromedial distance - 3 times the distance from the ipsilateral anterior superior iliac spine to the tip of the pointed first toe or leg length).

Prior to strength, triple-hop and YBT data collection all examiners were given a copy of a written description of the individual tests and how there were to be scored. Each examiner practiced the test they were responsible for administrating under the guidance of an experienced examiner during a minimum of three 1-hour training sessions prior to testing study participants.

**MRI Sequences and MRI Osteoarthritis Knee Score**

The MRI sequences employed in this investigation included:

1. Sagittal proton density (PD) TR/TE 1500/10 ms, slice thickness 3.5 mm, field of view (FOV) 150x140 mm
2. Sagittal and coronal PD fat saturated (FS) with TR/TE 2660/28 ms slice thickness 3.5 mm, field of view (FOV) 150x140 mm
3. 3D gradient echo FIESTA sequence with TR/TE 10.5/4.2 ms, flip angle 55°, slice thickness 1.0 mm and isotropic voxels

The MOAKS is a semi-qualitative method for providing a comprehensive inventory of possible abnormalities related to knee OA. A detailed description of MOAKS scoring has been previously reported.11, 21 Briefly, each knee image was divided into 14 sub-regions (inclusive of the tibio-femoral and patella-femoral compartments) and scored on specific features related to OA including; osteophytes, synovitis/effusion and bone marrow, articular cartilage, meniscal and ligamentous lesions. The central and posterior femoral sub-regions, and anterior, central and posterior tibial sub-regions were used to assess for bone marrow and articular cartilage lesions in both the medial and lateral tibiofemoral compartments. With respect to the patellofemoral compartment the medial and lateral femoral trochlea and patella were assessed for OA features. Bone marrow lesions were scored based on the size relative to each sub-region (0 =none, 1 <33%, 2 between 33–66%, and 3 >66%) using the STIR and TSE sequences. Using all available sequences, articular cartilage damage was scored as the percentage of each sub-region showing partial or full-thickness loss (0=no damage, 1 if <10% of regional cartilage involved, 2 if 10-75%, 3 if >75%), and osteophytes were scored based on size (0-3 for none/small/medium/large). When combining sub-regions to report prevalence of an OA feature (i.e., medial central femur and medial posterior femur and medial femur), only the largest size of each OA feature was included for analysis. Meniscal tears, maceration, and signal changes were identified based on the type and location as outlined by Hunter et al 2011.11 The extent of extrusion (anterior or medial/lateral) was also quantified as 0 if <2 mm, 1 if 2-2.9 mm, up to 3 if >5 mm. Note that MOAKS is an inventory-type scoring system, which provides a ‘profile’ showing to what extent each of the above features of arthritis and joint damage are present. In this study of young athletes, most knees were normal or nearly normal.