Validating the '[blinded name] Protocol': Assessing the reliability of hip muscle strength measurements using a motorised dynamometer and electromyography

Abstract

Background:

Muscle weakness is common following injury in athletes, and in the presence of hip pathology. It will cause abnormal hip biomechanics and can predict future injury. However, objective measurement of hip muscle strength is difficult to perform accurately and reliably. Therefore it is challenging to determine when an athlete has returned to pre-injury levels of strength. In addition, there is currently no standardised method of obtaining measurements, which prevents the data being compared or shared between research centres.

Purpose:

The purpose of this study is to comprehensively assess the inter- and intra-observer reliability of our standardised muscle strength measurement protocol. In addition, we have published a set of normative data for hip muscle strength according to the protocol.

Study Design:

This is a descriptive laboratory study.

Level of Evidence:

Level 2, inception cohort study

Methods:

Sixteen healthy male volunteers (age = 28.3 ± 7.9 years) were recruited. Those with a previous history of hip injuries or disorders were excluded. These volunteers underwent strength testing according to the [blinded name] Protocol on four separate occasions, performed by two independent assessors. Maximal voluntary contractions, fatigue fluctuations and electromyography measurements were recorded. Intra- and inter-observer reliability was assessed using inter-class correlation coefficient (ICC).

Results:

Good to excellent correlation was seen for both intra- and inter-observer reliability across almost all hip movements for maximal contractions; ICC ranges 0.78-0.93 and 0.78-0.96 respectively. The standard error of the mean for all hip movements was also extremely low, 2-3%.

Conclusion:

The [blinded name] Protocol is a highly reliable method for objective measurement of hip muscle strength. We recommend future studies use this protocol, or the principles underpinning it, to enable data sharing and comparison across different studies.

Clinical Relevance:

This is a description and analysis of hip muscle strength measurement. If widely used, it will allow for accurate and objective strength assessment and closer monitoring of hip injuries and pathology.

What is known about the subject:

Muscle weakness is very common in injured athletes and hip pathology.

Measuring hip muscle strength accurately is difficult.

There is currently no consensus on the best method of measuring hip muscle strength

What this study adds to existing knowledge:

This is a comprehensive description and reliability assessment of a hip muscle strength measurement protocol. Our analysis shows high reliability and accuracy of the protocol. In addition, we have provided the normative dataset for our sample.

Keywords:

Hip/Pelvis/Thigh, Muscle injuries, Muscle physiology, [blinded name] Protocol, Muscle strength measurement, Reliability

Introduction

The hip joint is a highly congruent and stable ball-and-socket joint, and is the most effective lever in the body [30]. This joint is surrounded by the most powerful muscles in the body, capable of generating huge forces to propel the body forward during running and jumping. Due to the high demands on the hip joint, surrounding muscle injuries are extremely common. They constitute one third of all time lost in men's professional soccer [13]. Hip muscle weakness has also been shown to be predictive of future knee ligament injuries [19]. In addition, there are numerous pathological hip conditions where muscle weakness can be identified [23]. Poor muscular control can lead to altered hip biomechanics, abnormal gait, instability, pain and further injury.

Despite its importance, it remains difficult to document hip muscle strength accurately. There are a number of fundamental reasons for the difficulty in measuring muscle strength reliably, which can be divided into the following categories: device type, joint positions, measurements recorded [24]. In addition the specific methods of strength measurement varies widely between centres. The reason is the vast number of variables involved in obtaining the measurements [22]. For this very reason, there are no 'normal ranges' of muscle strength for planes of movement, which makes judging an individual as weak or strong very difficult. This lack of consensus prevents comparison of data between centres.

To standardise the measurement of hip muscle strength in future studies, our group devised a protocol based on a systematic review of the literature on the topic [24]. This set of guidelines is termed the '[blinded name] Protocol', based on the city in which it was conceived. It was published with the aim of enabling measurement of hip muscle strength accurately, while minimising measurement error. If widely adopted, it would allow comparison of data across different studies.

The [blinded name] Protocol

The basic principles of the protocol are listed below. Further detail and the underlying reasons for the selection of our parameters are included in Appendix A.

- Use a motorised dynamometer if practicable
- Use isometric tests, as they are more reliable
- Stabilise participants during muscle testing
- Body and hip positions must be standardised (Table 1)
- The testing protocol must be followed identically for all participants

Objectives

In this study, we tested the reliability of muscle strength measurement using the [blinded name] protocol principles. The **primary objective** was to test the inter and intra-observer reliability of the protocol. The **secondary objective** was to create a set of normative data according to this protocol.

Methods

Sixteen healthy male volunteers (age = 28.3 ± 7.9 years; height = 1.82 ± 0.07 m; mass = 81.8 ± 10.1 kg) were recruited from a young adult population according to the following criteria.

Inclusion criteria:

- Aged 18-45
- Male
- Be able and willing to consent to participation in study

Exclusion criteria:

- Previous diagnosis of hip disease
- Previous hip injuries, investigations, or pain

Previous studies assessing reliability of motorised dynamometers have had sample sizes of 14-22 [8,12,21,25,29]. It was not possible to calculate the power for this study *a priori* as most of the parameters were unknown. Therefore we aimed for 20 participants in this study to fall within the range of published literature. However, the COVID-19 pandemic coincided with the latter stages of recruitment and measurement, therefore we stopped recruiting after 16 participants. *Post-hoc* power analysis showed a sample size of 14 participants was sufficient to achieve 80% power based on the ICC figures for hip extension.

Testing was carried out at the Sport and Exercise Sciences Laboratory at our local university. Ethical approval was granted for the study. Participants attended the laboratory on four separate occasions no less than 3 days apart each to prevent muscle fatigue affecting the measurements. Two testing sessions were conducted by each of the two assessors. The order in which they were tested by the assessors was pseudorandomised, and each testing session was independent of the previous sessions. The testing protocol was identical for each visit.

At the start of each session, participants had six pairs of silver-chloride surface EMG electrodes attached at sites on their right on the tensor fascia latae (TFL), rectus femoris, bicep femoris, gluteus medius and gluteus maximus according to the Seniam guidelines [18]; and a further site on the adductor longus according to Claibourne et al. [10]. Skin was prepared by shaving the area followed by cleaning with alcohol wipes. Once sensors were attached, participants performed a five-minute warm-up on a stationary bicycle. The following muscles were of interest in each plane (Table 1).

Plane of hip motion	Muscles activation measured
ABduction	Gluteus medius and tensor fascia lata
ADduction	Adductors longus
Extension	Gluteus maximus, bicep femoris
Flexion	Rectus femoris and tensor fascia lata

Table 1 – EMG measurements for muscle during each plane of motion

The muscle testing procedure was carried out on a motorised dynamometer (HUMAC Norm, CSMi, Soughton, MA, USA). Participants were secured into the dynamometer chair using a combination of straps, in each case the pelvis and torso were secured as the proximal segment to the moving femur. The movements were tested in the same order for each test: ABduction, ADduction, extension, flexion, internal, and external rotation. Testing positions related to each movement are detailed in Appendix 1.

Once secured into the dynamometer chair, the protocol for each movement was the same: isometric maximal voluntary contraction (MVC) test followed by the isometric fatigue test. Prior to the MVC test the participant was given a 5-second familiarisation trial. This was followed by three 5-second maximal effort isometric contractions each separated by 10 seconds of rest. Participants then rested in position for 5 minutes. All participants received the same amount of verbal encouragement throughout. They then performed a trial of less than 10 seconds at 60% of the MVC, to familiarise themselves with the force required to reach the target torque. This involved reaching, but not holding at, the target torque. The target torque and current torque measurements were displayed digitally on the screen in view of the participant. Participants were then instructed to hold the torque value at the target torque for 40 seconds. They were then given 5 minutes of rest before moving on to the next movement, repeating the process.

During the first testing session, the MVC value for each movement was calculated and used to determine the target torque for the fatigue trials. This target was used for all of the remaining session. This was to avoid confounding the variability in the MVC values and the variability in the fatigue variables.

For the MVC and fatigue trials, both EMG and joint torque from the dynamometer were recorded. EMG was captured at 1000 Hz using a Biomonitor ME6000 portable EMG recording through MegaWin capture software (Mega Electronics Ltd, Kuopio, Finland). All data was exported and analysed in Matlab (2019b, MathWorks Inc., Natick, MA, USA). Dynamometer torque data was collected at 100Hz.

Data Analysis

MVC torque was defined as the maximum mean value over a 0.3-second interval across all three MVC trials [3]. This value was normalised to the body mass of the volunteer. Standard error of the mean was calculated for the sample group for each movement plane to assess the accuracy of the measurements. The torque fluctuations in the fatigue trial were calculated as the coefficient of variation of the data [7]. EMG median frequency (MDF) was calculated for the MVC and each of 3 time points for the first, middle and last 10 seconds of the 40 second capture. Only muscles appropriate to the action were analysed for the respective movement. The following variables were analysed: MVC torque, fatigue torque fluctuations, change in EMG MDF between the three timepoints of the fatigue trial.

Each variable was analysed for intra- and inter-observer reliability using the intra-class correlation coefficient – ICC (3, 2). The test parameters included two-way mixed effects with absolute agreement and single rater/measurement. The 95% confidence intervals were calculated for the ICC values. Based on Koo and Li [20], strength of correlation was considered “poor” (ICC < 0.5), "moderate" (0.5–0.75), “good” (0.75–0.9) and "excellent" (ICC > 0.9).

Results

Correlation

The overall ICC values showed predominantly good to excellent correlation for both intra- and inter-observer reliability of the MVC torque measurements.

Intra-observer reliability

Table 2 shows the intra-observer reliability (ICC) for both testers. These were calculated based on the MVC torque measurements and torque fluctuations. MVC results show good to excellent correlation for both testers (ICC 0.78-0.96), for the majority of the movement planes. Extension showed the highest and most consistent ICC values (0.92-0.93). Fatigue fluctuations ICC values showed lower correlation across the movement planes for both testers. There was a wide range of correlation values, from poor to excellent for different movement planes.

	ICC - Tester 1		**ICC – Tester 2**	
	MVC	Fatigue Fluctuations	MVC	Fatigue Fluctuations
ABduction	0.88	0.85	0.83	0.59
ADduction	0.80	0.63	0.82	0.43
Extension	**0.93**	0.75	**0.92**	0.65
Flexion	0.72	0.69	0.89	0.66
Internal Rotation	0.90	0.19	0.78	0.84
External Rotation	**0.93**	**0.96**	0.78	0.58

Table 2 – Intra-observer reliability of measurements taken by two testers: maximal voluntary contractions (MVC) and torque fluctuations.

Inter-observer reliability

Inter-observer reliability was performed using the MVC torque measurements (Table 3). Correlation between testers was good to excellent in all planes except ABduction, with ICC values ranging between 0.78-0.96. ABduction correlation was moderate (ICC 0.69). Once again, extension had the highest inter-observer reliability, with excellent correlation (ICC 0.96).

	ICC (95% CI)
ABduction	0.69 (0.22-0.87)
ADduction	0.78 (0.47-0.91)
Extension	0.96 (0.92-0.98)
Flexion	0.86 (0.70-0.93)
Internal Rotation	0.85 (0.66-0.93)
External Rotation	0.85 (0.69-0.93)

Table 3 – Inter-observer reliability of measurements

Normative data for muscle strength

Table 4 depicts the normative data for the MVC measurements for our sample, showing the relative strengths of different muscle groups. Extension was the strongest plane of movement, followed by ABduction; generating torque values which are over double those of almost all other planes. The standard error of mean is very low and comparable across all planes, ranging from 2-3%.

	Mean ± SD (Nm)	Standard Error of Mean (% of Mean Value)
ABduction	2.04 ± 0.35	0.04 (2%)
ADduction	1.40 ± 0.29	0.04 (3%)
Extension	4.05 ± 1.09	0.14 (3%)
Flexion	1.81 ± 0.26	0.03 (2%)
Internal Rotation	0.58 ± 0.16	0.02 (3%)
External Rotation	0.72 ± 0.15	0.02 (3%)

Table 4 – Maximal voluntary contraction (MVC) values and standard error of mean, which is expressed as a percentage of the mean torque values of MVC.

Electromyography

Figure 5 shows the trend in EMG MDF during the 40 second fatigue contractions, split into three timepoints: first, middle and last 10 seconds. All muscle groups show fatiguability over time, except biceps femoris during hip extension.

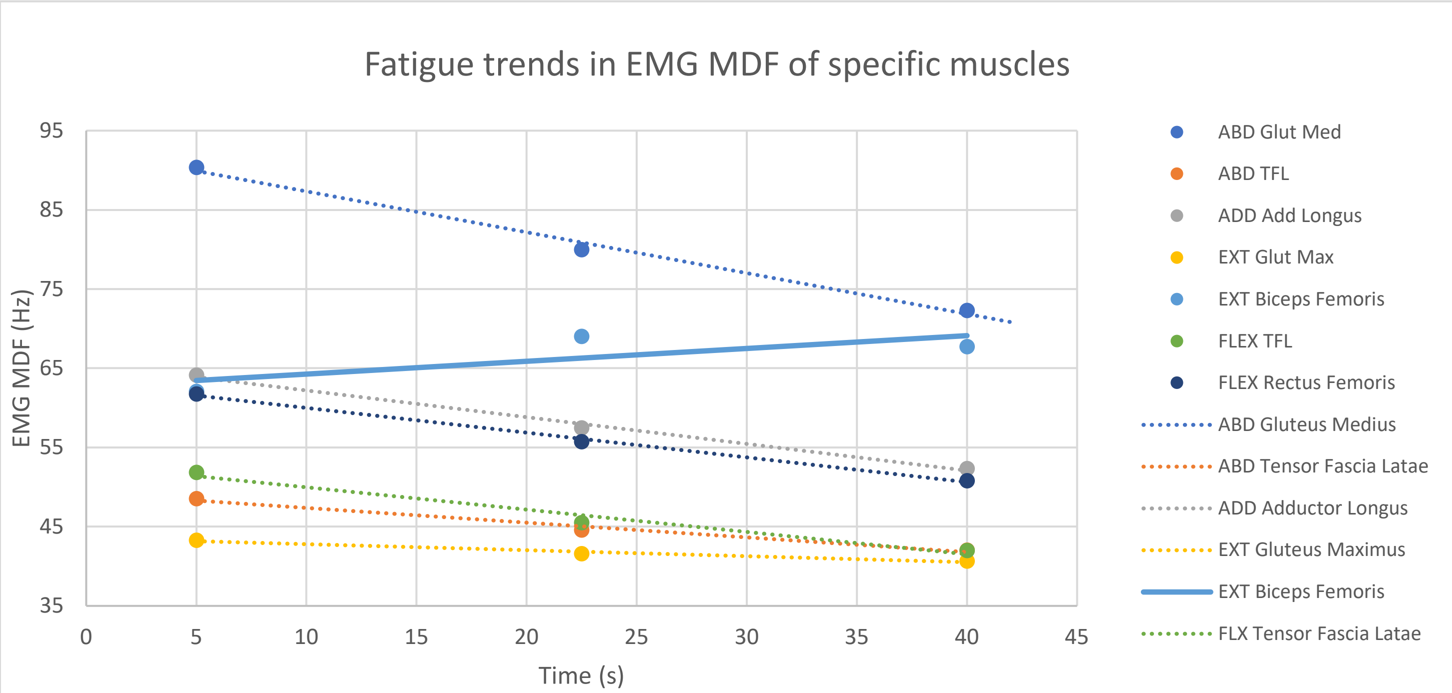

Figure 5 - Trend data for EMG median frequencies. The linear lines highlight the trend of EMG findings during the 40 seconds of fatigue strength testing. Reducing trends were seen in all muscles measured, except biceps femoris during hip extension.

ABD = Abduction; ADD = ADduction; EXT = Extension; FLEX = Flexion.

Discussion

Measuring hip muscle strength objectively and reproducibly is challenging. Hand held dynamometers are a cheap and convenient way of objectively measuring force exerted. These advantages make this a much easier tool to use for clinicians in the outpatient clinic. However, hand held devices only measure the counter-force required to stop the limb from moving. As a result, the measurements are prone to repeatability errors. This assessor-dependent measurement error has been documented in the published literature [6,31,33]. Therefore a consensus statement from the British Association of Sport and Exercise Sciences endorsed the use of motorised dynamometers where practicable [2].

The primary objective of this study was to test the inter and intra-observer reliability of the [blinded name] Protocol. The results show that the [blinded name] Protocol is a reliable method of measuring muscle strength, achieving good to excellent correlation within and between testers for MVC torque values. Among the different hip movements, extension was the most reliable plane of movement to measure, showing excellent correlation within and between testers (0.92-0.96). The MVC values showed higher reliability compared with fatigue fluctuations. Casartelli et al have found fatigue fluctuations to be a reliable measurements [8]. Despite this, however, the clinical application of fatigue fluctuation measurements is not entirely clear. Our protocol has also shown high intra- and inter-observer reliability in measuring hip extension in the supine position; this findings is not consistent among other studies [29]. This may reflect the hip position in which the movement was performed in the [blinded name] Protocol.

Intra-observer reliability testing showed the lowest reliability in hip flexion for MVC (0.72). This movement plane however still showed moderate correlation. The reason for slightly lower reliability in this case is not clear, especially as tester 2 showed a high ICC for the same movement (0.89). The MVC figures generally showed higher intra-observer reliability compared with fatigue fluctuations during the same planes of movement. While correlation was generally good to excellent for MVC, they were moderate to good for fatigue fluctuations, except ADduction and internal rotation. In particular, fatigue fluctuations in internal rotation showed poor correlation for tester 1, even though this correlation was good for tester 2. Our feedback from the volunteers was that internal rotation is not a comfortable movement to hold for 40 seconds. Therefore, our participants were likely using different strategies to hold the contraction for the fatiguability test, leading to less reliable measurements. This may also explain the relatively lower ICC values for the fatigue tests, compared to MVC. It is more difficult to sustain the same level of contraction for 40 seconds than to present peak contractions. Therefore this may increase the fluctuation and reduce the reliability of the fatiguability measurements.

Inter-observer reliability testing has shown correlation to be good to excellent for the majority of movement planes (0.78-0.96), with only one outlier in testing hip ABduction, showing moderate correlation between testers (0.69). The reason may be the body position for this test. ABduction and ADduction were measured in the lateral decubitus position. The volunteers were stabilised in this position using straps and firm padding. However, there is inherently less stability of the body in the lateral position over supine. Therefore when maximal contractions are performed, it may be easier to inadvertently lean either side of the true lateral position, thus creating some error between testers. When the confidence intervals of the ICC are examined more closely, they are both relatively wide. Whereas extension is once again the most accurate movement plane to measure, as it has the narrowest range of ICC (0.92-0.98).

A number of studies have compared inter- and intra-observer reliability of their measurement protocols using different devices and body positions. These have shown generally good to excellent reliability (ICC 0.71-0.95)[11,14,16,17]. Diamond et al also assessed the reliability of measurements in femoroacetabular impingement. This also showed good to excellent reliability (ICC 0.87-0.97) [12]. The ICC values for the [blinded name] Protocol have been comparable to these previous studies.

The EMG MDF values could be used to detect muscle fatigue during a prolonged contraction. These traces have shown a downward trend during the fatigue tests in all muscles tested, except one (Figure 5). The only outlier was biceps femoris (BF) during hip extension. Our theory to explain this is to consider gluteus maximus as the major extensor of the hip. Therefore, gluteus maximus is recruited early in the extension movement. BF then becomes recruited later in the movement and thus its fatigue response is delayed. The results support this theory as the EMG MDF for BF increases in the first half of the fatigue test, then becomes fatigued by the end of the test. Further research is required to understand the true cause of the pattern seen. Although we have not included ICC calculations for reliability of EMG measurements, two previous studies have shown potentially high reliability in a test-retest setting with EMG MDF (ICC ranging from 0.63-0.98) [26,27].

Overall, the [blinded name] Protocol has proven to be very accurate at providing MVC values. The standard error of the mean is 2-3% for all hip movement planes. This is an exceptionally low figure, especially when considering MVC findings in in pathological conditions. Casartelli et al compared these measurements in normal individuals and those diagnosed with femoroacetabular impingement in 42 participants [8]. An average difference of 16% between the two groups was detected, which is 5-8 times the measurement error in this study. Therefore, the [blinded name] Protocol would be able to detect a small change in MVC; such as that found in hip pathology.

The secondary objective was to create a set of normative data for men according to the [blinded name] Protocol. We have also included the normative values obtained, which may be used for comparison across different studies if the same measurement protocol is used. The two most powerful movements were found to be hip extension and ABduction. This finding is expected, as they are the two anti-gravity muscle groups and can generate large amounts of torque in bipedal motion. Casartelli et al also noted these two groups to be the most powerful in the hip [8]. However, they found extension to be weaker than ABduction. The reason for this may be due to the testing position of the muscle groups. They tested hip extension with the hip joint in 45 degrees of flexion. In our protocol extension was measured in 90 degrees of hip flexion, therefore the muscle fibres were more elongated in our study. We recommend hip extension to be measured in 90 degrees of flexion to ensure maximal voluntary contraction values are being recorded. The weakest movements recorded in our protocol were internal and external rotation, consistent with previously published data [8,12].

This normative data could be used as reference values to which pathological muscle weakness could be compared in the future. This would be useful for clinicians to compare hip strength in pathology or to monitor recover during rehabilitation from injury or surgery. Although we have use a motorised dynamometer, a hand held device may be used to obtain proxy measurements. However, it must be noted that the reliability of these figures will be lower, and therefore the values interpreted with caution. In addition, there is currently no consensus on grading of muscle weakness using MVC values. Currently, the MRC grading system is the only classification used to grade muscle weakness. However, it is generally accepted that a significant muscle weakness must be present to detect a clinical change from grade 5 to grade 4 power. Casartelli et al published a 11-28% reduction in power in femoroacetabular impingement syndrome patients compared to normal controls. [8] This equated to just under one standard deviation of the mean in their control sample population. Future research is required to develop an accurate and reproducible classification system for pathological muscle weakness using MVC values.

We believe future studies into muscle strength should utilise this protocol to measure hip muscle strength. This will not only generate accurate data, but it will also enable data to be combined from multiple sites and centres. This standardisation will enable future meta-analyses of the generated data, which is not possible due to the variations in measurement technique.

Conclusion

We have shown the [blinded name] Protocol to produce highly reliable measurements of hip muscle strength and recommend future studies adopt a similar approach using the principles outlined. This will enable comparisons across different studies. The protocol is very accurate, with a standard error of 2-3%, thus able to detect small differences in muscle strength.

Extension is the most reliable movement to measure and testing in the lateral position (for ABduction and ADduction) is less reliable. Both MVC and fatigue fluctuations can be used as measurements, though MVC is more reliable.

References

1. Almeida GPL, das Neves Rodrigues HL, de Freitas BW, de Paula Lima PO. Reliability and Validity of the Hip Stability Isometric Test (HipSIT): A New Method to Assess Hip Posterolateral Muscle Strength. *J Orthop Sports Phys Ther*. 2017;47(12):906-913. doi:10.2519/jospt.2017.7274

2. Baltzopoulos B, King M, Gleeson M, De Ste Croix M. The BASES expert statement on measurement of muscle strength with isokinetic dynamometry. , 31, pp.12-13. *Sport Exerc Sci*. 2012;(31):12-13.

3. Bernard J, Beldame J, Van Driessche S, et al. Does hip joint positioning affect maximal voluntary contraction in the gluteus maximus, gluteus medius, tensor fasciae latae and sartorius muscles? *Orthop Traumatol Surg Res*. 2017;103(7):999-1004. doi:10.1016/j.otsr.2017.07.009

4. Bilodeau M, Schindler-Ivens S, Williams DM, Chandran R, Sharma SS. EMG frequency content changes with increasing force and during fatigue in the quadriceps femoris muscle of men and women. *J Electromyogr Kinesiol*. 2003;13(1):83-92. doi:10.1016/S1050-6411(02)00050-0

5. Bloom N, Cornbleet SL. Hip Rotator Strength in Healthy Young Adults Measured in Hip Flexion and Extension by Using a Hand-held Dynamometer. *PM&R*. 2014;6(12):1137-1142. doi:10.1016/j.pmrj.2014.06.002

6. Bohannon RW. Intertester Reliability of Hand-Held Dynamometry: A Concise Summary of Published Research. *Percept Mot Skills*. 1999;88(3):899-902. doi:10.2466/pms.1999.88.3.899

7. Casartelli NC, Leunig M, Item-Glatthorn JF, Lepers R, Maffiuletti NA. Hip flexor muscle fatigue in patients with symptomatic femoroacetabular impingement. *Int Orthop*. 2012;36(5):967-973. doi:10.1007/s00264-011-1385-5

8. Casartelli NC, Maffiuletti NA, Item-Glatthorn JF, et al. Hip muscle weakness in patients with symptomatic femoroacetabular impingement. *Osteoarthritis Cartilage*. 2011;19(7):816-821. doi:10.1016/j.joca.2011.04.001

9. Charlton PC, Mentiplay BF, Grimaldi A, Pua Y-H, Clark RA. The reliability of a maximal isometric hip strength and simultaneous surface EMG screening protocol in elite, junior rugby league athletes. *J Sci Med Sport*. 2017;20(2):139-145. doi:10.1016/j.jsams.2016.06.008

10. Claiborne TL, Timmons MK, Pincivero DM. Test–retest reliability of cardinal plane isokinetic hip torque and EMG. *J Electromyogr Kinesiol*. 2009;19(5):e345-e352. doi:10.1016/j.jelekin.2008.07.005

11. Click Fenter P. Reliability of stabilised commercial dynamometers for measuring hip abduction strength: a pilot study. *Br J Sports Med*. 2003;37(4):331-334. doi:10.1136/bjsm.37.4.331

12. Diamond LE, Wrigley TV, Hinman RS, et al. Isometric and isokinetic hip strength and agonist/antagonist ratios in symptomatic femoroacetabular impingement. *J Sci Med Sport*. 2016;19(9):696-701. doi:10.1016/j.jsams.2015.10.002

13. Ekstrand J, Hägglund M, Waldén M. Epidemiology of muscle injuries in professional football (soccer). *Am J Sports Med*. 2011;39(6):1226-1232. doi:10.1177/0363546510395879

14. Ford-Smith CD, Wyman JF, Elswick RK, Fernandez T. Reliability of stationary dynamometer muscle strength testing in community-dwelling older adults. *Arch Phys Med Rehabil*. 2001;82(8):1128-1132. doi:10.1053/apmr.2001.24291

15. Frost KL, Bertocci GE, Wassinger CA, Munin MC, Burdett RG, Fitzgerald SG. Isometric performance following total hip arthroplasty and rehabilitation. *J Rehabil Res Dev*. 2006;43(4):435. doi:10.1682/JRRD.2005.06.0100

16. Gerodimos V, Karatrantou K, Paschalis V, et al. Reliability of concentric and eccentric strength of hip abductor and adductor muscles in young soccer players. *Biol Sport*. 2015;32(4):351-356. doi:10.5604/20831862.1189202

17. Habets B, Staal JB, Tijssen M, van Cingel R. Intrarater reliability of the Humac NORM isokinetic dynamometer for strength measurements of the knee and shoulder muscles. *BMC Res Notes*. 2018;11(1):15. doi:10.1186/s13104-018-3128-9

18. Hermens HJ, Freriks B, Disselhorst-Klug C, Rau G. Development of recommendations for SEMG sensors and sensor placement procedures. *J Electromyogr Kinesiol*. 2000;10(5):361-374. doi:10.1016/S1050-6411(00)00027-4

19. Khayambashi K, Ghoddosi N, Straub RK, Powers CM. Hip Muscle Strength Predicts Noncontact Anterior Cruciate Ligament Injury in Male and Female Athletes: A Prospective Study. *Am J Sports Med*. 2016;44(2):355-361. doi:10.1177/0363546515616237

20. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

21. Lu Y-M, Lin J-H, Hsiao S-F, Liu M-F, Chen S-M, Lue Y-J. The Relative and Absolute Reliability of Leg Muscle Strength Testing by a Handheld Dynamometer: *J Strength Cond Res*. 2011;25(4):1065-1071. doi:10.1519/JSC.0b013e3181d650a6

22. Maffiuletti NA. Assessment of Hip and Knee Muscle Function in Orthopaedic Practice and Research: *J Bone Jt Surg-Am Vol*. 2010;92(1):220-229. doi:10.2106/JBJS.I.00305

23. Marin-Peña O, Lund B, Ayeni OR, et al. Basic Concepts in Hip Arthroscopy. In: Kerkhoffs GMMJ, Haddad F, Hirschmann MT, Karlsson J, Seil R, eds. *ESSKA Instructional Course Lecture Book*. Springer Berlin Heidelberg; 2018:45-67. doi:10.1007/978-3-662-56127-0_4

24. Mayne E, Memarzadeh A, Raut P, Arora A, Khanduja V. Measuring hip muscle strength in patients with femoroacetabular impingement and other hip pathologies: A systematic review. *Bone Jt Res*. 2017;6(1):66-72. doi:10.1302/2046-3758.61.BJR-2016-0081

25. Meyer C, Corten K, Wesseling M, et al. Test-Retest Reliability of Innovated Strength Tests for Hip Muscles. Carrier D, ed. *PLoS ONE*. 2013;8(11):e81149. doi:10.1371/journal.pone.0081149

26. Minoshima Y, Nishimura Y, Tsuboi H, et al. Reliability of Power Spectral Analysis of Surface Electromyogram Recorded during Sustained Vastus Medialis Isometric Contraction in Assessment of Muscle Fatigability. *Open J Ther Rehabil*. 2017;05(02):43-52. doi:10.4236/ojtr.2017.52005

27. Mutchler JA, Weinhandl JT, Hoch MC, Van Lunen BL. Reliability and fatigue characteristics of a standing hip isometric endurance protocol. *J Electromyogr Kinesiol*. 2015;25(4):667-674. doi:10.1016/j.jelekin.2015.02.003

28. Roman-Liu D. The influence of confounding factors on the relationship between muscle contraction level and MF and MPF values of EMG signal: a review. *Int J Occup Saf Ergon*. 2016;22(1):77-91. doi:10.1080/10803548.2015.1116817

29. Seko T, Kumamoto T, Miura S, et al. Measuring seated hip extensor strength using a handheld dynamometer: an examination of the reliability and validity of the protocol. *J Phys Ther Sci*. 2015;27(7):2179-2182. doi:10.1589/jpts.27.2179

30. Standring S, ed. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. Forty-first edition. Elsevier Limited; 2016.

31. Thorborg K, Bandholm T, Hölmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(3):550-555. doi:10.1007/s00167-012-2115-2

32. Widler KS, Glatthorn JF, Bizzini M, et al. Assessment of Hip Abductor Muscle Strength. A Validity and Reliability Study: *J Bone Jt Surg*. 2009;91(11):2666-2672. doi:10.2106/JBJS.H.01119

33. Wikholm JB, Bohannon RW. Hand-held Dynamometer Measurements: Tester Strength Makes a Difference. *J Orthop Sports Phys Ther*. 1991;13(4):191-198. doi:10.2519/jospt.1991.13.4.191

Appendix A – Details of the [blinded name] Protocol

The [blinded name] Protocol

The basic principles of the Protocol are outlines in the main text of the article. However, the device type, body positions and measurement parameters obtained are discussed below.

Device

Motorised dynamometers have been established as the preferred method of objective measurement of muscle strength, and endorsed by the British Association of Sport and Exercise Sciences [2]. The reasons for this are discussed in the original [blinded name] Protocol article [24].

Body Position

It is crucial to have the participants in identical body positions to ensure muscle groups are being tested in their most efficient range. The forces generated can vary depending on the tested position of the hip joint [32]. In addition, while performing the high force movements such as hip extension, the whole body may move on the machine. Therefore, the torso and pelvis must be stabilised on the machine to prevent measurement error [11,32].

Different body and hip positions can affect the reliability of the measurements taken. For example maximum hip ABduction strength has been shown to be more reliably measured in the lateral position than supine [25,32]. Based on a review of the published literature, the following positions were chosen for the [blinded name] Protocol (Table 1) [1,5,8,12,15,21,25,29,32]. These had shown the highest and most reliable measurements for each plane of movement.

Plane of hip motion	Body position	Hip position	Position of force pads
ABduction	Lateral	Hip abducted to 15 degrees	Knee*
ADduction	Lateral	Hip abducted to 15 degrees	Knee
Extension	Supine	Hip flexed to 90 degrees	Knee
Flexion	Supine	Hip flexed to 45 degrees	Knee
Internal rotation	Supine	Seat inclined by 30 degrees, Knee flexed to 90 degrees	Ankle**
External rotation	Supine	Seat inclined by 30 degrees, Knee flexed to 90 degrees	Ankle

Table 1: Details of body position during testing of each movement.

* Immediately proximal to the superior pole of the patella

** Immediately proximal to the malleoli

Measurements

There are a number of parameters which can be measured to assess different aspects of muscle strength. Previous clinical studies have obtained torque measurements and fatigue fluctuations to assess strength and fatiguability of muscles [7,8,12,15,25]. Maximum voluntary contraction (MVC) is a logical parameter to measure muscle strength. This is the maximum torque generated during the movement.

Fatigue fluctuations have been used to measure muscle fatiguability [7]. This appears to be a reliable measurement, although its clinical application is not entirely clear. In addition, electromyography (EMG) has been used to analyse the quality of muscular contractions. Although EMG generates a lot of data from each movement, its reliability is highly variable [4]. Due to its non-invasive applicability, surface EMG is widely used for this purpose. However, It is a cutaneous measure, rather than a direct muscular measurement. Therefore trace quality and amplitude can vary hugely, even in the same subject during the same session [4,9,28]. Overall, the addition of EMG to torque measurements adds another dimension of muscle analysis. It would help identify the muscles which dominate or fatigue more easily during certain contractions.